A Defense of Moral Motivation Externalism

By Rafael Martins

M.A. Philosophy - University of Kansas, 2014
M.A. Philosophy -Universidade Federal do Rio de Janeiro, 2009
B.A. Philosophy - Universidade Federal de Ouro Preto, 2007
B.A. History – Universidade Federal de Ouro Preto, 2005

Submitted to the graduate degree program in Philosophy and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chair: Dale Dorsey, Professor

_____

Brad Cokelet, Associate Professor

_____

Ben Eggleston, Professor

_____

Jason Raibley, Associate Professor

_____

Brittnee Carter, Assistant Professor

Date Defended: July 13th, 2021

The dissertation committee for Rafael Martins certifies that this is the approved version of the following dissertation:

A Defense of Moral Motivation Externalism

Chair: Dale Dorsey

Date Approved: July 13th, 2021

Abstract

First, I argue that moral reasons are not *necessarily* overriding. Then I show that since moral reasons are not necessarily overriding, moral judgements are not necessarily motivating. The corollary is that motivation cannot be an a priori criterion of determining which beliefs are or are not genuinely moral. I then defend moral judgment externalism (motivation depends on external factors to the moral judgment, such as favored conditions, and, for this reason, the connection between moral judgment and motivation is contingent) from classic and contemporary objections. Since my view about practical reason is committed to the existence of external reasons, I begin by respond to long-standing problems for reasons externalism: the alienation and explanatory constraints. After that, I tackle a specific opponent in the semantic field: the moral twin earth problem. I respond to the problem by evaluating some answers to the open-question argument, and more specifically the moral realist proposal which argues that moral terms should be defined synthetically like terms used in science. In theory, then, moral terms, just like names used to designate natural kinds, are causally regulated by extensions in the world and behave like rigid designators. However, the moral twin earth argument seeks to show that the moral realist semantic is an implausible explanation of the determination of the reference of moral terms because it cannot accommodate internalist semantic intuitions. Even when the realist seeks to establish the meaning of the moral terms using the causal method, he falls into an uncomfortable position because his moral semantic theory does not accommodate both the objective and the motivational essence of morality. Finally, I develop a theory for mitigating the problems raised by moral twin earth by resorting to the notion of *referential intentions*. My hope is to develop some argumentative updates that would improve the ability of moral realism's semantic theory to determine reference across different hypothetical linguistic communities.

Acknowledgements

The road up to the conclusion of my doctorate was long, winding, and extremely challenging. Painful at times. However, many people have contributed to make my ride beautiful and pleasant. I am incredibly lucky for having been dearly helped by many people, as well as having the opportunity to share amazing and unforgettable moments with them. I am wholeheartedly thankful to: Moacyr Martins, Olivia Cesta Martins, Antonieta Cesta, Marcia Denise Martins, Daniela Martins Cotting and Thiago Cotting, Bauru family, Liza Yuzhakova, Clark Roberts, Chris Bodamer, Jay Werner, Tasia Zouzas, Chris Bunn, Theresa Schmidt, Larry Northrop. The University Press of Kansas and Conrad Roberts, International Support Services and Charles Olcese, Hodgie Bricke, Mario Nogueira, Maria Clara Dias, Maria Aglaé Vilardo, Colin Crawford, Desiderio Murcho, Derick Darby, Ann Cudd, Dale Dorsey, Richard DeGeorge, Erin Frykholm, Don Marquis, Sarah Robins, Marcy Lascano, Jason Raibley, Brad Cokelet, Ben Eggleston, John Bricke, John Symons, Brittnee Carter, Eric Atwood, Alex Barajas, Molly Mahoney, Cindy Hodges, KU Philosophy Department faculty, staff, and graduate students, Fulbright, Capes, and Gaiola de Ouro. Thank you so much.

To my dear loving mother
Marcia Denise Martins

Table of Contents

Chapter 1 Setting the Scene

1. The Debate on Moral Motivation (Modalism vs. Anti-modalism)

Much of the plausibility of metaethical anti-realism, such as non-cognitivism and error theory, rests on an honorable tradition according to which *moral* judgments are *necessarily* motivational.[1] Moral judgment internalism (MJI), as the simple thesis is known, says that when a person makes a moral judgment, either she is necessarily motivated to act according to this judgment, or she does not actually hold this judgment. The view seems very intuitive. If I claim that consuming counterfeit products[2] is morally wrong, and yet wear fake Nikes, either I do not *really* believe doing so is morally wrong, or something has gone wrong, and hence, the asymmetry between my claim and my action requires explanation.

There are innumerable theories about the motivational character of morality. But they divide in two families, whose ultimate difference reduces to their views about the alethic or modal relations of necessity and contingency between different relata[3] involved in moral judgments, depending on theory (Platonic, Aristotelian, Humean, Kantian, etc.) and method (conceptual, semantic, causal, psychological, etc.). The family of internalisms establishes necessary links in either mereological, causal, or semantic terms between moral judgment (or reason) and

---

[1] Fredrik Björklund, Gunnar Björnsson, John Eriksson, Ragnar Francen Olinder and Caj Strandberg. "Recent Work on Motivational Internalism" in *Analysis* vol. 72 No.1, January 2012.

[2] Studies show that profits from counterfeit goods sales finance organized crime, such as human trafficking and terrorism. Cf. United Nations Office on Drugs and Crime, *The Illicit Trafficking of Counterfeit Goods and Transnational Organized Crime* https://www.unodc.org/e4j/en/organized-crime/module-3/key-issues/counterfeit-products-trafficking.html

[3] Relata in this context are usually defined in either Humean or Anti-Humean terms. A Humean defines them as belief and desire, being a matter of debate whether a moral judgment is one or the other or a mix of both (besires). I will briefly explain these nuances in the following paragraphs. Anti-Humeans, as far as my survey goes, can be Kantian, like Thomas Nagel, *The Possibility of Altruism* and Christine Korsgaard, *The Sources of Normativity*, or Aristotelian, like McDowell and Kristjánsson

motivation.[4] In contrast, moral judgment externalism (MJE), be it in the context of conceptual identity, moral psychology, or practical reason, advocates that the connection between moral judgment and motivation is contingent.[5]

Given the different conceptions of moral motivation and the variety of methods to investigate it, there may be some confusion about *where exactly* lie the necessities and contingencies involved in morality, and consequently about the nature of moral judgments. [6] So, allow me to clear that out. One form of this debate is established in terms of the *modality* of the connection between moral judgment and motivation. Roughly speaking, this is a binary view, i.e. either moral judgements are necessarily motivational or not, period. This formulation does not worry (much) about whether the method explaining MJI is conceptual, semantic, or causal. All it wants is to get down to raw modalities between relevant relata. But some people find intuitive that moral motivation, like other motivations, is a matter of degree, and find this binarism too narrow a conceptual space.[7]

---

[4] Tresan, Jon. Metaethical Internalism: Another Neglected Distinction, *The Journal of Ethics*, 2009 and Fletcher and Ridge, *Having It Both Ways* 2014 Tresan, J, chapter 5 *Diachronic Hybrid Moral Realism*. James Dreier, Dispositions and Fetishes: Externalist Models of Moral Motivation, *Philosophy and Phenomenological Research*, Vol. 61, No. 3 (Nov., 2000), pp. 619-638

[5] Shafer-Landau, *Moral Realism: a defence*, part III chaps 5 and 6. and "A defense of Motivational Externalism", *Philosophical Studies*, 2000 and David Brink, "Moral Motivation" *Ethics* 107 (1997): 4-32. David Brink, *Moral Realism and The Foundation of Ethics,* CUP, 1989. Zangwill, Nick. "Motivational Externalism: Formulation, Methodology, Rationality, and Indifference." In Gunnar Björnsson, Fredrik Björklund, Caj Strandberg, John Eriksson, Ragnar Francén Olinder (editors). *Motivational Internalism.* Oxford University Press, 2015; Sigrun Svavarsdottir, "Moral Cognitivism and Motivation", *The Philosophical Review*, V. 108, n.2 1999.

[6] In fact, never has it been noticed that the internal/external vocabulary was first employed by W. D. Falk to mean exactly that this metaphor is confusing. Falk is a modalist, since he is concerned with whether deontic commands are necessarily motivational. To be precise, Falk used "external" in the sense that deontic commands from a Deity or a Government, or even one's parents, for that matter, impact an agent as external to his set of prudential interests. Which is one of the reasons why, Falk reminds, we say "internalize" a norm, command, principle, etc., usually through habituation. One of the reasons the vocabulary is confusing is that the command could arise *inside* the standpoint of prudential interests and yet *not* be necessarily motivational. "Ought" and Motivation", *Proceedings of the Aristotelian Society*, New Series, Vol. 48 (1947 - 1948), pp. 111-138 R.J. Wallace also complains the terms internal/external are "multiply ambiguous" p. 356. "How to Argue About Practical Reason" *Mind*, Vol. 99, No. 395 (Jul., 1990), pp. 355-385

[7] Zangwill: "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. Intuitively, we want some things more than others, and we believe some things to a greater degree than others. (We are more confident of some claims than others.) Our

A second form of the debate explicitly defended by Nick Zangwill[8], and underlying various metaethical hybridisms, such as Michael Ridge's ecumenical expressivism[9], is to approach the debate in terms of the *essence* or *constitution*[10] of moral judgements and avoid modal talk.[11] The nuances here are thin. The essentialist can say that it is part of the essence of a moral judgment (*qua* belief) to cause motivation (*qua* desire), but that motivation is mereologically external to the moral judgment in itself. A constitutionalist prefers to say that moral judgments are *constituted by* characteristics from both belief and desire, such as *besire* theorists.[12] Both will say that motivation is *internal* to moral judgment. In this manner, the debate turns on what kind of pro-attitude a moral judgment is. One of the apparent gains of this view is that it allows one to say that moral judgments do *not necessarily* motivate, even though motivation is part of the essence or constitution of genuine moral judgments, hence internal to it. Moreover, since they avoid modal talk, they can also say that moral judgments do not necessarily issue *overriding*[13] motivation, which allows them to explain a number of cases where people seem to make genuine moral judgments and yet are not motivated.

---

mental world is not black and white. This is often overlooked in the motivation debate. We must reformulate [MJI], and its modal consequences, so as to take account of this. We need an account of how strength of motivation is built into moral beliefs of different degrees." "Indifference Argument", *Philosophical Studies* 2008, 138: 91-124pg. 95

[8] Zangwill, *Indifference Argument* and *Besires and The Motivation Debate, Externalist Moral Motivation.* To be clear, Zangwill argues against modalism, because he advocates, conceptually and empirically, that moral motivation is a matter of degree. However, he is an externalist because he fully embraces the Humean view that beliefs are utterly inert and hence desires are *not* part of the essence nor constitution of moral judgments.

[9] James Dreier, "Ecumenical Expressivism, Finessing Frege", *Ethics* 116 (January 2006): 302–336

[10] Besire theory is an example

[11] Zangwill, Nick. "Motivational Externalism: Formulation, Methodology, Rationality, and Indifference." In Gunnar Björnsson, Fredrik Björklund, Caj Strandberg, John Eriksson, Ragnar Francén Olinder (editors). *Motivational Internalism.* Oxford University Press, 2015; Fredrik Björklund, Gunnar Björnsson, John Eriksson, Ragnar Francen Olinder and Caj Strandberg. "Recent Work on Motivational Internalism" in *Analysis* vol. 72 No.1, January 2012.

[12] Nick Zangwill, *Besires and The Motivation Debate;* Altham, J. E. J. (1986). The legacy of emotivism. In G. Macdonald & C. Wright (Eds.), Fact, science and morality: Essays on A.J. Ayer's language, truth and logic (pp. 275–288). Oxford: Basil Blackwell.; Bradley, Richard. "Desire as Belief Revisited." Analysis 69 (2009): 31-37. Lewis, David. "Desire as Belief." Mind 97 (1988): 323-32. Lewis, David. "Desire as Belief II." Mind 105 (1996): 303-313. Price, Huw. "Defending Desire-as-Belief." Mind 98 (1989): 119-27.

[13] Iakovos Vasiliou, "Introduction" in *Moral Motivation: A History*, OUP, 2016

I disagree about the apparent anti-modalist advantage. First, the internal/external vocabulary can be a confusing metaphor, since saying that motivation is internal to moral judgment is uninformative until it is replaced by either modality or dependence. Now, not much is gained by dependence talk, because the next most natural question is whether a moral judgment, to be genuine, is necessarily or contingently dependent on motivation, be motivation connected with or part of moral judgment. The metaphysical question about the number of mental states (e.g. belief and desire or besire) involved in moral judgment is equally not advantageous for the anti-modalist, since by replacing connection talk for mereological talk may well over or under postulate wholes and parts.

Modalists can be neutral about the number of distinguishable mental states which compose a moral judgment. To some extent, establishing the modality between motivation and moral judgment is independent of the psychological composition of a moral judgment. One can be practically inert in relation to a plethora of mental states and motivation may be something different altogether. Motivational concerns about moral action affect both cognitivists and non-cognitivists alike. Moreover, modalists do not deny that the mental states or processes involved in moral judgement may have parts, they are just concerned whether motivation is a *necessary* part or essence of a genuine moral judgment, or not. In case one claims that motivational status of mental states or reasons for action and the like is a matter of degree, like Zangwill, or goes dispositional like Slote and Prinz, the modalist may be satisfied with the contingency of moral motivation accepted by those views.

Lastly, though many externalists talk about the contingent relation between belief and desire, the modalist is not committed to overpostulating entities, since even if one discards either belief or desire as the ultimate nature of moral judgments, reducing it to a single mental state or

pro-attitude, someone worried about modalities will still ask whether moral judgments are *necessarily* motivating in themselves. So, merging belief's and desire's attitudinal characteristics or discarding either does not guarantee necessary motivation. Again, the modalist will just ask whether moral judgements, in order to be considered as such, are *necessarily* motivational or not. Therefore, in what follows I maintain the debate in terms of the modal connection between moral judgment and motivation, be them two distinguishable wholes, primitive or otherwise, or the latter merely a part of the former. The modalist conception of the debate on moral motivation cuts across all proposals in metaethical metaphysics and epistemology.

1.1. Sentimentalism and Rationalism about Motivation and Practical Reason

I have mentioned above that two salient contentions about moral motivation refer to whether moral judgment and motivation are one single entity (e.g. besires[14], beliefs, intentions, etc.) or two distinct ones (such as a belief and a desire; belief and will/intention) and the degree of motivation (which can be put in terms of both doxastic and desiderative degree[15]). Since these contentions regard the nature of moral judgements, they lead us to the debate between sentimentalism and rationalism[16] about moral motivation and practical reason. At their extremes, these camps fundamentally diverge about the primacy[17] of beliefs and desires in order to fully

[14] A besire can be a mental state that is both a belief and a desire; a desire that accounts for the objectivity of our moral judgments (a special type of desire), a belief that accounts for the motivation of our moral judgments (a special type of belief), and lastly besire can be a mental state that accounts for the motivation of our moral judgments and that accounts for the objectivity of our moral judgments, but is not a belief or a desire.

[15] Zangwill for instance advocates a degree conception for both belief and desire. Op. cit.

[16] I find this nomenclature more precise give the history of the debate, but given the popularity of the so-called Humean Theory of Motivation (Smith, Pettit, and many others), this debate goes by other names such as Humeanism vs. Anti-Humeanism or response-dependent/subjectivism vs. objectivist conceptions of value and practical reason, such as Kantian (rational agency) and Aristotelian (perfectionism). David Brink, "The Significance of Desire", *Oxford Studies in Metaethics,* 3 (2008): 5-46.

[17] In case necessary, we can use the terminology: Strong/Hard Rationalism about moral motivation: Motivation is exclusively constituted by beliefs (or reason). Weak/Soft Rationalism: Motivation is constituted by both a belief and a desire, with the belief representing the major source of motivation. Weak/Soft Sentimentalism: Motivation is constituted both by a belief and a desire, with the desire representing the major source of motivation. Strong/Hard Sentimentalism: Motivation is exclusively constituted by desires. I am taking this way of drawing the scenario from

describe the nature of moral judgments and reasons, to explain[18] how they cause action, to justify their authority, establish their resonance, list functional implications, determine semantic content, etc.

The simple belief-desire thesis has become the paradigmatic vocabulary for talking about motivation, but some clarifications are in order. First, when we use term "desire" we often *equate* it to motivation. But we can also conceptually detach motivation from desire. For example, take severe depression. Humeans explain the depressed's inaction in terms of lacking desires, weakened desires, and other anomalies that preempt desire, since the depressed seem to retain their evaluative beliefs, but lack the desire to comply.[19] However, we can say, plausibly, that the depressed *really* desires to get up, pursue life projects, and experience blissful social activities, but they just can't. And it does not matter how much they desire a different life; for them, motivation simply does not reduce to desire.[20] Part of what makes the depressed frustrated is that he really desires to act according to what he knows to be good for him, and even for others, but whole-body motivation seems to take more than desire. That said, skeptics about desire can either say that even though desire is part of motivation, motivation does not reduce to desire[21], or they can simply jettison

---

Kristján Kristjánsson *Virtuous Emotions*, OUP, 2019. And Daniel Eggers "Moral Motivation in Early 18th Century Moral Rationalism", *European Journal of Philosophy*, 2019. The same conceptual distinctions can be made about belief and desire, see Edwards, Gary, "The Many Faces of Besire Theory." Thesis, Georgia State University, 2011. https://scholarworks.gsu.edu/philosophy_theses/91

[18] For instance, Bernard Williams says "if something can be a reason for action, then it could be someone's reason for acting on a particular occasion, and it would then figure in an explanation of that action." *Moral Luck,* p. 106-7

[19] Michael Stocker pg. 744 (1979) "Desiring the Bad: An Essay in Moral Psychology," *Journal of Philosophy* 76(12): 738-53., Alfred Mele, pg. 733-4 (1996) "Internalist Moral Cognitivism and Listlessness," *Ethics* 106: 727-53., Michael Smith: "the effect of… depression is to remove [one's] desire." 1994, p. 135

[20] Stevem Swartzer, "Humean Externalism and the Argument from Depression", *Journal of Ethics and Social Philosophy*, v. 9, n. 2, 2015.

[21] Robert Audi, "Intending", "Intention, Cognitive Commitment, and Planning", "Action, Intention and Reason". Joshua May, "Because I Believe It's The Right Thing To Do"
Ethical Theory and Moral Practice, Vol. 16, No. 4, (2013), pp. 791-808

desires in favor of cognitive conceptions of judgment[22], intention or volition to determine value and explain action.[23]

Given the simple belief-desire[24] thesis, the main difference between the two camps has been whether the nature of moral judgment, motivation and action requires the presence of desire.[25] But since some have questioned the motivational efficacy of desire, Pettit's description of the feud is helpful:

> "What divides Humeans and anti-Humeans, by all accounts, is a difference of view about the potency of reason in motivating behaviour. The issue between them is not whether motivating reasons always involve desires but whether they always involve the presence of non-cognitive states, states which reason on its own is incapable of producing. If the thesis about desires is relevant to that issue, that is only because it is assumed that desires are non-cognitive states of this kind."[26]

So, in what follows I will be using desire in that sense, but aware that desire and motivation are not (always) conceptually the same thing, which allows the use of a larger philosophical background beyond the belief-desire picture. For example, putting a reductionist view of motivation to desire (conceptually) aside for moment, Aristotle, Hume, and Kant, all have received both internalist and externalist interpretations. While McDowell[27] and Caj Strandberg[28] read internalism in an Aristotelian view of motivation and reason, Kristján Kristjánsson[29] and Brad

---

[22] Dale Dorsey, "Subjectivism Without Desire", *The Philosophical Review*, Vol. 121, No. 3 (2012), pp. 407-442; Steven Swartzer op cit and "A Challenge to Humean Externalism" *Philosophical Studies,* 2018

[23] Nagel, *Possibility of Altruism*; Korsgaard, *Skepticism About Practical Reason*, *Sources of Normativity*, McDowell op. cit., Dancy, Platts, Audi op.cit.

[24] Though many call the simple belief-desire thesis "Humean Theory of Motivation", the distinction between belief and desire pre-dates far back David Hume.

[25] Smith, Pettit, HTM

[26] Philip Pettit, "Humeans, Anti-Humeans, and Motivation" *Mind*, New Series, Vol. 96, No. 384 (Oct., 1987), pp. 530-533

[27] Op. Cit. Caj Strandberg,(2000). Aristotle's internalism in the Nicomachean ethics. *Journal of Value Inquiry*, 34(1), 71–87. Kristjánsson explicitly is a cognitivist and hybrid externalist. McDowell flirts with externalism under some interpretations but seems to ultimately endorse some sort of (Aristotelian) internalism about the nature of moral *judgment* (when expressed by the "fully virtuous", not just by the continent). All anachronisms (regarding belief/desire) about Aristotle aside, of course, reason why both philosophers claim to be merely exploring Aristotelian alternatives to good old 20th century Humeanism about reasons and motivation.

[28] Various, check later

[29] "Aristotelian Motivational Externalism, *Philosophical Studies*, 2013 and *Virtuous Emotions,* OUP, 2018.

Inwood[30] defend an externalist interpretation. Hume has been seen as an internalist, because of dispositionalism and other response/desiderative/subjectivist/reason-is-inert conceptions of value, while others think the Humean belief-desire theory favors externalism, because belief and desire are two distinct existences, whose connection is contingent. Kant is seen as an externalist about reasons, because moral reasons, the orthodox[31] Kantian says, are determined without desire, and sometimes even *in spite of* desires, but Kant certainly thinks that once one *knows* what morality requires, one *is* motivated accordingly, or one is irrational[32].

Hence, internalism and externalism about motivation, reasons, and value, apply equally to sentimentalism and rationalism, because no matter what the epistemic nature of the moral judgment is, belief, desire, etc., it remain plausible to ask whether one is *necessarily* motivated to act accordingly and/or has ultimate reason for it. This question, sometimes put as "why be moral?", is ubiquitous because it expresses a plausible, at least initially, conceptual distinction between *pro tanto* reasons (S-reasons) and all-things-considered reasons[33]. As long as this gap remains, the modalist concern about motivation remains. Lastly, insofar as the question about motivation remains attractive for sentimentalists and rationalists alike, I find no convincing reason to abandon my focus on the ultimate modality between moral judgment and motivation, regardless of schools of thought.

1.2. Belief, Desire, Directions of Fit

It is not that easy to distinguish between beliefs and desires. For the purposes of this dissertation, I will just briefly make some comments about this distinction. Smith defends the

---

[30] Brad Inwood, *Ethics and Human Action in Early Stoicism*, OUP, 1985
[31] Robert Louden, *Kant's Impure Ethics*, OUP, 2000. Allen Wood, various
[32] *Venia* anachronisms.
[33] David Brink, "Moral Conflict and Its Structure", *The Philosophical Review*, Vol. 103, No. 2 (Apr., 1994), pp. 215-247

distinction in terms that if an agent has a belief that *p* and a desire that *p* at the moment of a perception that ¬*p*, the belief that *p* tends to disappear whereas the desire that *p* may well persist.[34] Copp & Sobel[35] point out, Smith distinguishes between the agent's *previous mental state* (which can be both a belief that *p* and a desire that *p*) and a *new introduced state* (the perception with content that ¬*p*). Hence, what turns a previous mental state into a belief or desire is its tendency to remain or disappear in face of the new introduced mental state.

But Copp and Sobel try to show that there are two possible objections to be mentioned with regard to this point. First, we must ask: what does the new introduced state consist of? If, on one hand, this new mental state is a belief, the explanation is presupposing the very difference that it intends to explain, but if, on the other, the introduced mental state is a desire, there is nothing more trivial than an agent's having conflictive desires, regardless of the introduced state.

Additionally, whatever the nature or content of the introduced mental state is, there is no rational need for the previously supported belief or desire to dissipate. A perception with content ¬*p* is not always a good reason for us to instantaneously change our believe that *p,* or our desire for *p.* From horse betting to weather forecast, examples abound. Exactly because many of our beliefs are of inductive or inferential nature (and the desires they instrumentally serve), the fitting between such beliefs (and their correspondent desires) and the experience from which they stem is not always instantaneous. Our beliefs and desires are, to some extent, revisable, but to that end, we often need successive experiences and time for reflection. For example, few disillusionments are as deep as *coming to believe* one better give up a life's dream project.

The caution with which we make our doxastic and desiderative revisions shows that it is not always rational to abandon a belief or a desire from a first, or even a number of, contrary

[34] Smith 1994, p. 115
[35] "Against Direction of Fit Accounts of Belief and Desire" *Analysis* 2001

impressions. So, the way by which people accept or reject their beliefs and desires does not sit comfortably within Smith's model. Caveat made, in what follows I will rely on the belief-desire conceptual distinction of mental states, but aware of its limitations.

1.3. Naturalism, Subjectivism, Objectivism

In the context of metaethics, naturalism[36] investigates what moral properties are, what is the epistemic function of moral intuitions, and seeks to elucidate the metaphysical status of moral phenomenology. This method tries to see philosophy as continuous to science and not solely made of *a priori* reasoning whose substantial truths are immune to empirical tests. Some positions hold that a philosophical explanation of moral practice should be apt to empirical investigation, and some go as far as to propose that moral concepts bear some explanatory weight in other empirical or social sciences.[37] On the other hand, proponents of non-naturalism[38] and anti-reductionists believe that *not all reality* is susceptible to scientific inquiry, that is, there are facts that are not captured by laws of nature or by externalist essentialism, such as strong physicalism, for instance.

Moral anti-realism feeds off the intuition that from a naturalistic point of view, there is nothing in concrete reality, *external* to the human mind and heart that qualifies as genuinely moral. Some examples are in order. Says Hume:

---

[36] Naturalism has no clear definition as to its theoretical status. The term can be used to refer to a method of inquiry or to designate a substantial philosophical thesis as such (Papineau, 2015). For instance, Gibbard and Railton argue that a satisfactory explanation of the nature and function of morality must be supported by the results obtained by evolutionary biology, or the best current theory about how we have become what we are (Gibbard 1990, Railton, 1986). I am interested in *semantic* naturalism, where an analysis of any concept must designate concrete properties, objects, or events. For example, the concept of "moral goodness" can be analyzed in terms referring to properties discovered by empirical investigations, such as those performed by psychology and physiology (such as the maximization property of well-being, if we admit that well-being is a complex psychophysical state). Moreover, substantial naturalism proposes that the interpretation of philosophical concepts *bears some explanatory and causal weight in the empirical and social sciences* (Railton, 1993: 315). This latter perspective has also been discussed between Harman (1977, 1986) and Sturgeon (1986).

[37] Railton, 1993, 315; Sayre-McCord, Sturgeon. *Moral Explanations* (for example, historical explanations heavily employ moral terms)

[38] Shafer-Landau, 2003

"Take any action allow'd to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, of real existence, which we call *vice*. In whichever you take it, you find only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflection into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action."[39]

Arguing against absolute values, Wittgenstein mentions a similar idea, with a Moorean gist, as one of the views that morality cannot deliver objective values:

"If for instance in our world-book we read the description of a murder with its all details physical and psychological, the mere description of these facts will contain nothing which we could call an *ethical* proposition. The murder will be on exactly the same level as any other event, for instance the falling of a stone. Certainly the reading of this description might cause us pain or rage or any other emotion, or we might read about the pain or rage caused by this murder in other people when they heard of it, but there will simply be facts, facts, and facts but no Ethics."[40]

Hence, the anti-realist concludes that moral properties have no concrete existence outside the human mind and heart. For the same reason, the non-cognitivist argues that moral judgments are only the expression of emotions. Nevertheless, realists will argue from different starting points to the conclusion that moral judgements reflect moral facts. For example, many realists insist that mostly everyone would hold similar pro-attitudes in face of more or less paradigmatic cases. Take murder. Who could deny that it is *true* that killing an innocent human being is *prima facie*[41] morally wrong? The realist will then argue that it is a perception of moral *facts* that grounds everyone's holding similar beliefs (and perhaps similar desires) about murder. In this manner, murder would be *objectively* wrong. The judgment is objective not just because of its being a consensus, but in the sense that it *purports to report facts in virtue of which it is true or false (or merely elicited)*. According to this view there are facts about the killing of human beings in virtue of which it is

[39] Hume *Treatise*. 3.1.1, p. 301 (his italics)
[40] Wittgenstein. "Lecture on Ethics". *The Philosophical Review*, vol. 74, n.1, 1965. (his italics) p. 6-7
[41] More on *prima facie* moral judgments, David Ross, *The Right and Good*, OUP, 2002. To be more precise, when I qualify judgments and reasons as "prima facie", "pro tanto", or "initial", I only mean to make sure that the judgment or reason in question may or may not become an all-things-considered, and that they have enough doxastic or desiderative force to make them at least initially inescapable from the agent's consideration.

true that killing an innocent human being is morally wrong. Hence, for a realist, an appraiser's or an agent's non-cognitive pro-attitude, *in situ* or otherwise, may have many functions (e.g. foster social trust), but not the one of determining the legitimacy of sincerely held moral judgments.

So, morality seems to be both objective and subjective. But these two characteristics of morality seem to be in conflict. One can clearly acquiesce to the normative authority of the wrongness of murder while not feeling compelled to spare the life of an innocent human being[42], and at the same time, one can have feelings that have no normative authority whatsoever.[43] According to this distinction, moral properties, because they are considered necessarily motivating, would be subjective, as they are emotional responses to events and objects. Yet, observation of moral practice suggests that moral agents frequently seem motivated to behave according to standards, principles, or values, that seem to be objective and whose truth value or normative significance is in a way independent of the mind that entertains them. This asymmetry has been referred by W. D. Hudson as the *central problem*[44]*,* Michael Smith as the *moral problem*[45] and by Frank Jackson as the *location problem*[46].

This problem has also influenced theories about the meanings of moral terms. For example, objectivists have not had an easy way to explain reference fixing in naturalist vocabulary, due to the motivational disposition taken to be necessarily part of the meaning of moral terms.[47] But, virtually no one would deny that motivations may vary drastically. So, internalism would render the meaning of moral terms hostage to each person's motivations. That would lead to moral

---

[42] Zangwill's mercenary case, various op. cit.
[43] Brink, *Significance*
[44] *The Is-Ought Question*: *A Collection of Papers on the Central Problem in Moral Philosophy.* Macmillan, 1969.
[45] Smith. *The moral problem*. Blackwell, 1994.
[46] Jackson. *From metaphysics to ethics*: *a defence of conceptual analysis*, OUP, 1998.
[47] Mackie writes: "And I do not think it is going too far to say that this assumption [about the distinction between normativity and objectivity] has been incorporated in the basic, conventional, meanings of moral terms. Any analysis of the meanings of moral terms which omits this claim to objective, intrinsic, prescriptivity is to that extent incomplete" Mackie, (1977, p. 35).

relativism if people would name things "morally good" *only* when they hold the relevant motivation towards those things. In order to stop this variance, subjectivists resort to idealization. It stops the variance because criteria of idealization, coherence, departure, parsimony, etc. confer normative authority to subjective states.[48] But then subjectivism cannot account for the normative authority of subjective states without losing its naturalist appeal, since any account requires some idealization.[49] That is an worrisome hit when naturalism is one of subjectivism's most important motivations. Subjectivists in turn argue that objectivism misses the link between value and people's motivations since it seems to name values independent of how people really feel about them. In this manner, the accommodation of the motivational[50] character attributed to moral judgments remains a field of contention and a determining metaethical factor about the legitimacy of moral judgments.

1.4 Motivation and Rationality

MJI claims that an individual who genuinely believes to be morally required to perform a certain action is necessarily motivated to (or, necessarily has the desire to) act accordingly, or he is irrational. But why is the unmotivated irrational? Well, there are two senses of rationality here. Reasons are *facts* that explain and justify actions and pro-attitudes such as beliefs and desires. So, if reasons are facts, to be irrational is to incorrectly track the facts, or fail to recognize them in some relevant way, or fail to respect them in some way. For example, that the sum of the internal angles of a tringle is 180 degrees is a mathematical fact. If I believe the sum equals to any other

[48] Brandt, Lewis, Railton, Smith

[49] David Brink, The Significance of Desire. *Oxford Studies in Metaethics* 3:5-45 (2008); Dale Dorsey, Subjectivism without Desire. *Nous* 51:1 (2017) and Idealization and the Heart of Subjectivism. *The Philosophical Review*, Vol. 121, No. 3 (2012)

[50] Hare, Ethics: problems and answers, 1997.: "Let us now ask in a preliminary way if moral judgments (for example, those expressed with 'must') are acts of speech prescriptive or descriptive. The answer is that they are both things, but that the distinction needs to be carefully preserved, otherwise we will not be able to understand the different characteristics of 'must' sentences that binds them to both genders" (pp. 31-32).

number, I am being mathematically irrational, because I am not correctly tracking mathematical facts. If I am building a bridge and fail to recognize this mathematical fact, I am even more significantly mathematically irrational. Additionally, it seems obvious that an engineer certainly is motivated to respect this mathematical fact. Now suppose it is a fact that killing an innocent human being is wrong. Was I to believe otherwise, I would be morally irrational, since I am not correctly tracking moral facts.[51] If I have someone's life in my hands and fail to recognize this moral fact, I am even more drastically morally irrational. Like the engineer, it seems obvious that a surgeon certainly is motivated to respect this moral fact.

Diametrically opposed to MJI, MJE states that *motivation in itself* depends on external factors to the moral judgment, such as favored conditions[52], and, for this reason, the connection between moral judgment and motivation is contingent.[53] The *indifference argument* has been considered the master argument against MJI.[54] The indifference argument can be made in a priori and a posteriori terms. But, basically, the argument capitalizes on the possibility, conceptual and empirical, that people can sincerely make a moral judgment and yet be motivationally indifferent to that, and most importantly, remain rational in the relevant sense.

But there have been some doubts[55] about the dialectic strength of this argument depending on how it is illustrated. The morally indifferent has been exemplified by Knaves (Hume), Fools

[51] For example, something like the "available to us" in "Rage, passion, depression, distraction, grief, physical or mental illness: all these things could cause us to act irrationally, that is, to fail to be motivationally responsive to the rational considerations available to us. Korsgaard, 1986, p. 13

[52] For example: proximity, imagination, eloquence, repetition, temper, nature of the object, context, vivacity, risk, etc. Hume *Treatise*, 2.3.6.

[53] Brink 1989, p. 75-79,

[54] Zangwill, *Philosophical Studies* (2008) p. 92

[55] Zangwill 2008, section 3, Kristjánsson, 2012 pg. 431: Reasons for [these] doubts can be skepticism about thought experiments as well as about empirical experiments, based on arguments such as departure from commonsense, conceptual underdetermination (e.g. moral vs. prudential value), etc.

(Hobbes), amoralists (Brink), mercenaries[56] (Zangwill), cynics (Svavarsdottir), cannibals (Hare), depressed (Stocker, Mele, Smith), etc. I am not inclined to use those examples, because I think it is dialectically stronger to defend motivational externalism about morality based on the dualism (or even pluralism[57] if one is prepared to go that far) of practical reason[58]. It shall strike as more compelling if we can show that good and strong-willed persons may care about things other than morality. After all, there is more to life than morality. Notwithstanding the power of this intuition, the move requires that we not only that, but that the other options are sufficiently rationally justified. And knaves and mercenaries don't seem to pass this test.

One of the motivations for externalism is the intuition that we can genuinely believe in having a moral reason and yet be motivated by a *reason* other than moral, say a prudential. For externalists, normative rationality encompasses morality, but does not reduce to it, since there are reasons of other normative nature, such as prudential, aesthetic, professional, political, etc. Hence, an individual who believes to be morally required to perform an action, i.e. genuinely recognizes a moral reason, and nevertheless, does not feel motivated to act accordingly, may be rational and linguistically competent, since he may be motivated by other more stringent reason(s). The stringency of normative reasons is determined by a number of subjective and objective factors. But because a moral reason is not always *necessarily* an all-things-considered reason, which in turn, indeed may or may not necessarily carry motivational force, the demands of practical reason

[56] Zangwill: "a mercenary I once met on vacation exuded moral indifference. He was in control, reflective and articulate. Everything he said convinced me that he was perfectly aware that his vocation was genuinely morally wrong, not merely what people conventionally call 'wrong'. He fully understood the wrongness of his vocation. But he was not very concerned about that. He was more concerned with his immediate interests and concerns, that is, colloquially, looking after number one. There was no moral cognitive lack. He made that quite clear. Indeed he insisted on it. The mercenary was unusually indifferent to the demands of morality; but he shared moral beliefs with the rest of us, and with his former self. He insisted on that." 2008, p. 102

[57] Michael Stocker, *Plural and Conflicting Values*, OUP, 1990

[58] Foot, 1978, Crisp, Brink, Dorsey

are not necessarily symmetrical to the demands of morality. Conclusion, not being morally motivated is not irrational. Here is another way to put this:

> "According to [Humean motivational internalism], one who has a normative belief…is thereby motivated to do something. This is so often said as for it to seem almost rude to ask: motivate the believer to do *what*? In some cases, the answer all but suggest itself. If I believe I ought to mow my lawn, then what I am motivated to do is to mow my lawn. If I believe that I ought to pay my taxes, then what I am motivated to do is to pay my taxes…But what of my other normative beliefs? Such as, for example, my belief that Hitler is a villain? Or worse yet for these purposes, my belief that you ought to mow your lawn? Here is not at all clear what I am to be thought to be motivated to do in virtue of having the belief…It may well be supposed that if I believe that Hitler is a villain, then if my town council were to declare a referendum on whether to put up a statue to him, I would be motivated to vote against doing so. Again, if I believe you ought to mow your lawn, then if I came to believe that I was situated as you are, I would be motivated to mow my own lawn. But as things stand now, I do have those beliefs about Hitler and you, but I am not now motivated to do anything at all by them." (Judith Jarvis Thomson, "The Legacy of Principia" in Horgan and Timmons, *Metaethics After Moore*, OUP, 2006, pg. 241-2)

Now something must be said about the distinction between morality and prudence. The separation between a person's value judgments and her motivations can be traced back in the history of Western thought as far Homer's *Odyssey.*[59] But it assumes philosophical contours with Socratic intellectualism. Socrates denies the very existence of akrasia[60]. According to Socrates, if Penelope does not act according to her avowedly knowledge of what the good is, call it a belief ϕ, then she does not actually hold belief ϕ. The Socratic explanation for her not doing ϕ is that actually, possibly unknown to her, Penelope does not hold belief ϕ, perhaps she even holds another belief ψ.[61]

---

[59] Vasiliou, Introduction in Vasiliou, Iakovos (ed.) *Moral Motivation: A History.* OUP, 2016

[60] Insert, pg for "no one errs willingly"

[61] See *Meno* 77b6–78c2 and *Protagoras* 352b1–358d4. Cf. Brickhouse, T.C. and Smith, N.D. "Socratic moral psychology". CUP, 2010. Nehemas, Alexander. *Socratic Intellectualism.* In Cleary, John (ed.), Proceedings of the Boston Area Colloquium in Ancient Philosophy, vol. 2. University Press of America, 1987. Penner, Terry and Christopher Rowe. "The desire for the good: Is the *Meno* inconsistent with the *Gorgias*?" *Phronesis* 39: 1994. p. 1–25. 1990. "Plato and Davidson: Parts of the soul and weakness of will." *Canadian Journal of Philosophy* 16 (Suppl.): 1990, p. 35–72. Santas, Gerasimos. The Socratic Paradoxes. *The Philosophical Review*. Vol. 73, No. 2, 1964, pp. 147-164.

Intuitive as it sounds, many find this Socratic intellectualism too strong. Notwithstanding, to provide an alternative account to the phenomenology of rational action is no easy task[62]. I am inclined to acquiesce to Socratic intellectualism were we to think that if you are *not* motivated, *perhaps necessarily* motivated to act according to *an all-things-considered* belief about what you ought to do at time t, then you probably do not believe ϕ. "Believe" is ambiguous here, because it seems more precise to say that you may still believe ϕ*,* while holding other normative beliefs like belief ψ, or that belief ϕ is *not yet* an all-things-considered normative belief.

There is yet a second ambiguity in need of light. Socratic intellectualism is never clear about whether the normative belief at work is moral or prudential; let alone political, professional aesthetic, and so forth. Why couldn't it be the case that I usually act quite in accordance with my prudential beliefs, or political, for that matter, and yet fail dramatically to act according to my moral beliefs? As much as this distinction is not clear today, it was never clear in Greek words. But though the appeal to the good of the city, as opposed to the good of a private citizen, may not always issue clear-cut categorization at times, it is sufficient for now. Aristotle says:

> For even if the good is the same for a city as for an individual, still the good of the city is apparently a greater and more complete good to acquire and preserve. For while it is satisfactory to acquire and preserve the good even for an individual, it is finer and more divine to acquire and preserve it for a people and for cities.[63]

I find this passage extremely important from the point of view of anti-moral rationalism. Regardless of what Aristotle himself would respond, had I had my life and the future of the city at stake, these words leave no doubt about what ought to be done. But anyone will wonder; Isn't that asking too much of me? There being conditions of my future survival elsewhere, I sure would consider reasons to preserve my own good. Hobbes, for instance, says:

---

[62] Davidson, various. Brian Weatherson, *Normative Externalism*, OUP, 2019

[63] *Nicomachean Ethics,* Book I, Chapter 2, 1094b. I'm taking this quote as illustrative. There is an interesting debate about motivation vis-à-vis virtue development in Aristotle. Kristjánsson op. cit

……he questions whether injustice [break political covenants]…may not sometimes stand with that reason which dictateth to every man his own good;…And I say it is not against reason…when a man doth a thing which…tendeth to his own destruction, yet such events do not make it [injustice] reasonably or wisely done.[64]

Just to be clear, Hobbes is saying that when requirements of justice tend to my own destruction, it is never against reason to consider acting otherwise, but since justice, argues Hobbes, usually has prudential gains[65], injustice is often unreasonable. Anyway, I am using these two pieces of textual evidence to emphasize that a distinction between moral goodness and prudential goodness can be draw by appeal to the asymmetry between the good of the city and the good of the individual. For example, when I wonder about what I ought to do concerning a new bus stop in front of my house, at a city transit meeting, it is equally rational for me to think not just about how the new stop will benefit the city, but as well as how it will impact my tranquility. I may believe that a new bus stop is the right thing for the city and not be motivated to accept it, given its impact in my private life. Do I not believe that is good for the city? Of course I do! I just think my life will go unbearable if I need to put up with the ordeal will be brought to the front of my house. Hence, I have a reason, at least an *initial/prima facie/pro tanto* reason, to oppose the new bus stop. Moreover, that reason will not disappear *even if* I turn out to, all-things-considered, acquiesce to the new bus stop. My discomfort with the noise in front of my house will remain, even if I believe, all-things-considered, that that was the right thing to do. By all-things-considered belief ϕ, I mean a belief that I have an all-things-considered reason to ϕ. That said, I do not see why I could not still hold other beliefs about having other reasons, say belief β.

I will argue that moral reasons/requirements are not *necessarily* overriding, *viz*. a moral reason is not necessarily, by nature, a decisive reason against all other realms of practical reason/requirements. Hence, I believe it is possible to show that *since moral reasons are not*

[64] Hobbes. *Leviathan*, chap XV, § 4-5. This is the voice of Hobbes's Fool (Brink, *Moral Realism*, 1989)
[65] In other words, morality matches, so to speak, with prudence. Perhaps prudentially optimal.

necessarily overriding, moral judgements are not necessarily motivating. The corollary, seems to me, is that motivation cannot be an a priori criterion of determining which beliefs are or are not genuinely moral. So, methodologically, I will establish a view about the place of morality in overall practical reason and use its conceptual implications to reject MJI. I remain indifferent to internalism about personal goodness, and I am equally indifferent to internalism about all-things-considered reasons/requirements, i.e. that all-things-considered reasons/requirements are necessarily motivational. All I am saying is that *pro tanto* reasons, and moral reasons are born as such, are not necessarily overriding, and therefore, genuine endorsement of a moral reason does not necessarily require motivation. Moral reasons may be *inescapable* in the sense that they cannot be rationally ignored*, viz.* no relevant practical reasoning shall elect a reason as all-things-considered, before moral reasons *necessarily be considered*. But not actually electing a moral reason as all-things-considered is not (necessarily) a rational *failure*, i.e. is not irrational.

1.5 Motivation and Semantic

It seems trivial to say that people, while holding genuine moral convictions, may not be *actually* or *in situ* motivated to act accordingly. But I believe that idealized, conditional, or restricted versions of internalism do not correctly explain why an individual who ultimately acts as to satisfy reasons, other than moral, is yet not irrational. For internalists the absence of the connection is always a problem to be explained rather than a truth to be elucidated. Since I disagree with internalists, I will argue that moral judgment and motivation have a conceptual relationship of contingent nature. Therefore, I plan to conclude, as part of this project, that MJI cannot be, strictly, an *a priori* constraint on what is to count, semantically, as a moral judgment[66]. This

[66] I am defending the same as Shafer-Landau: "we need to be *argued to* [author italics] metaethical theories that incorporate MJI." p.270 Viz. from rudimentary to complex versions of MJI, internalism is very often merely assumed or presupposed to be in a way or another a necessary part of the nature of morality.

project, hence, has a specific opponent in the semantic field: Timmons and Horgan's moral twin earth argument. The subjective motivational character and the objective authority of morality thus confer to moral terms a *referential tension* because they give the impression that moral terms denote a single extension, which, nevertheless, presents two apparently inter-repelling properties: objectivity and subjectivity. So, we do not find in classic, and neo-classic[67], metaethics a semantic theory of moral terms, which satisfactorily accommodates those two properties of morality. So, I am going to evaluate two influential arguments, the open question argument, and the twin moral earth, raised against naturalist moral realism.

I am going to evaluate some answers to the open-question argument, and more specifically the moral realist proposal which argues that moral terms should be defined synthetically like terms used in science. This view draws on the semantic externalism of Kripke and Putnam. In theory, then, moral terms, just like names used to designate natural kinds, are causally regulated by extensions in the world and behave like rigid designators, that is, they designate the same entities in all possible worlds in which such words are deployed.

However, the moral twin earth argument seeks to show that the moral realist semantic is an implausible explanation of the determination of the reference of moral terms because it cannot accommodate semantic intuitions. Even when the realist seeks to establish the meaning of the moral terms using the causal method, he falls into an uncomfortable position because his moral semantic theory does not accommodate both the objective and the motivational essence of morality. Like Moore, Timmons and Horgan admit that the determination of the reference, in the moral case, is necessarily dependent on the desiderative dispositions of the observer, and therefore, designate different extensions across different possible worlds.

---

[67] I call neo-classic metaethics the prolific decades of 80s and 90s. Darwall, Gibbard, and Railton, *Moral Discourse and Practice* 1997, Toward *Fin de Siecle* Ethics: Some Trends.

I am going to argue in favor of the thesis that naturalistic moral realism can overcome moral twin earth argument by resorting to the notion of *referential intentions*.[68] My hope is to develop some argumentative updates that would improve the ability of moral realism's semantic theory to determine reference across different hypothetical linguistic communities.

Chapter 2 The Motivation Argument for Moral Non-Cognitivism/Anti-Realism

MJI helps to support the Motivation Argument, the illustrious argument against varieties of moral realism, moral cognitivism, or moral objectivism. Next, I will introduce the argument and discuss it. Then I argue against MJI, from the point of view of practical reason. Lastly, I address Smith's objection to MJE.

2.1. The Motivation Argument (MA)[69]

The *locus classicus* of the motivation argument is found in Hume's Treatise:

> "[M]orality…'tis suppos'd to **influence** our passions and actions, and to go beyond the calm and indolent judgments of the understanding. And this is confirm'd by common experience, which informs us, that **men are often govern'd** by their duties, and are deter'd from some actions by the opinion of injustice, and impell'd to others by that of obligation. Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv'd from reason; and that because reason, …, can never have any such influence. Morals **excite** passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason…An active principle can never be founded on an inactive; and if reason be inactive in itself, it must remain so in all its shapes and appearances.[70]

In short:

---

[68] Brink, Copp, Putnam

[69] This argument goes by various names. It is called "motivation argument" in Rachel Cohon. "Is Hume a noncognitivist in the motivation argument?" *Philosophical Studies*, 85, 1997; Nick Zangwill, "Besires and the Motivation Debate", *Theoria,* 2008, and *Motivational Externalism: Formulation, Methodology, Rationality, and Indifference* in Bjornsson et.al. *Moral Motivation* OUP, 2015; Michael Smith. "The Motivation Argument for Non-Cognitivism. In Charles Pigden. *Hume on Motivation and Virtue*, Palgrave 2009. It is called "internalist argument" in. Carl Strandberg, "The pragmatics of Moral Motivation", *Journal of Ethics,* 2011; William Tolhurst. "Moral Experience and the Internalist Argument Against Moral Realism" *American Philosophical Quarterly*, Vol. 32, No. 2 (Apr., 1995). "Antirealist argument" or yet "internalist antirealist argument" in Brink, Moral Realism, pgs. 37 and 42; Lastly it is called "influence argument" in Frank Snare, *Morals, Motivation, and Convention.* CUP, 1991.

[70] Hume, Treatise, 3.1.1 pg. 294

1. "Morals…have an influence on the actions and affections."
2. "Reason alone …can never have such influence."
3. "An active principle can never be founded on an inactive"[71]
4. "It follows that [morals] cannot be deriv'd from reason."

A few initial comments. First, Hume does not use modal language along the passage. I think that the contemporary modalist version of MJI reads necessity into phrases such as "morality influences/has an influence on passions" and "morality excites passions" as well as in "men are *often* governed by their duties". Second, this passage has been used to justify another famous one where Hume, some interpretations[72] say, declares that no "ought" can be inferred from an "is", without violating semantic and inferential rules.[73] Third, the motivation argument is one of the major influences of all sorts of noncognitivism which tend to be internalist. But as I have explained earlier, the modalist debate on motivation cuts across from cognitivism to non-cognitivism, since in both cases there will be the question of whether moral judgments are *necessarily* motivational. And in the modalist sense, the Humean theory of motivation has also been used by externalists as we will see when discussing Michael Smith's approach to the motivation argument.

First question is how Hume supports the Motivation Argument. Let us start by premise 2. Hume argues[74] that reason is inert, *viz.* that it has no influence on the passions and actions, or that reason itself cannot produce actions nor affections. In order to prove premise 2, Hume provides what Rachel Cohon calls the Representation Argument (RA)[75]:

1. "Reason is the discovery of truth or falsehood."
2. "Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact."

---

[71] Conservativeness of logic, See Pigden, pg.
[72] Chapter 2 this dissertation
[73] Pigden. It will be opportune to say some things about the no-ought-from-is passage after discussing the motivation argument. I will tackle that discussion in Chapter 2
[74] T 2.3.3; 3.1.1; and 3.3.1.
[75] T 3.1.1; Cohon, 1996 p. 252

3. "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."
4. "…our passions, volitions and actions…[are] original facts and realities, complete in themselves, and implying no reference to other passions, volitions, and actions."
5. Thus "…[our passions, volitions, and actions] are not susceptible of any such agreement or disagreement."
6. "'Tis impossible therefore, they can be pronounced either true or false, and be either contrary or conformable to reason."

Hume transitions from RA to MA's premise 2 in the following way. Reason can only approve something when it discovers its representational accuracy; and disapprove something when it discovers its representational failure Since desires and actions do not represent reality, be it demonstrative or empirical, they do not correspond to anything. If they don't correspond to anything, they can't be true or false or as in Hume's words, "reasonable or unreasonable". According to Cohon, Hume is thinking that if reason were to produce an action by its own, it would do so in the same way in which it produces true conclusions: by approving it, i.e., recognizing it as true or inferring it from premises. But reason cannot do these things with actions and passions – it cannot *prove* actions and passions true. It can only claim that it is either true or false that an action actually took place or inform us that an action would lead to certain consequences or yet that it has been produced by certain causes. But it cannot say that actions and passions themselves are true or false. Therefore, producing actions and passions is not the business of reason.

Now, in disagreement with the noncognitivist interpretation, I side with Cohon, who argues that since Hume is not talking about moral *judgments*, RA and MA together do not constitute a defense of noncognitivism, which is a thesis about judgments. Rather, Hume uses RA and MA to argue against (Clarkian) moral rationalism. More specifically, Hume defends a negative thesis about moral properties: they are not identical to reasonableness and unreasonableness.[76]

[76] Cohon, 253

To make the point, it pays to note that once Hume believes to have proved that actions do not have the property of being reasonable or unreasonable, he repeats MA in slightly different ways whose wording in bold supports Cohon's proposal[77]:

Rendition B
1. "**Morals** excite passions, and produce or prevent actions."
2. "Reason of itself is utterly impotent in this particular."
3. The rules of morality, therefore, are not conclusions of our reason."

Rendition C
1. "…reason can never ***immediately*** prevent or produce any action by contradicting or approving of it,"
2. "**Moral good and evil**…are found to have that influence"
3. "Therefore, reason "cannot be the source of the distinction betwixt moral good and evil."

Rendition D
1. "The **merit and demerit of actions** frequently contradict and sometimes control our natural propensities."
2. "But reason has no such influence"
3. "Moral distinctions, therefore, are not the offspring of reason."

The point being made is that actions can have property F, but cannot have property G, therefore F and G are distinct properties. Substantially speaking, moral properties, being properties of actions, are not identical to epistemic properties, which are properties of beliefs. Goodness is not reasonableness and badness is not unreasonableness. Note that Hume is not saying that moral *judgments* such as the judgment that "benevolence is laudable," cannot be true or false, reasonable or unreasonable. He is talking about the action's property of being laudable. I'd assume that once we are clear about the nature of moral qualities, we may be in condition to *conclude* whether it is true or false that an action instantiates it. Only then we have the question whether moral judgments have descriptive properties and thus can be true or false.

Now, why is MA taken as support for non-cognitivism? Once Hume has established that rational processes are inert, he adds that the results of demonstrative and causal reasonings, namely

[77] I assume that premise 3 is implicit.

beliefs, cannot *alone/immediately* produce/motivate actions. Here, the noncognitivist steps in to say that if rational judgments do not motivate action, while moral judgments do, then moral judgments cannot be rational judgments. If moral judgments are not beliefs, then they must be something else which *necessarily* possesses this motivational quality. Passions/desires seem to be the most suited to assume such responsibility. But since desires lack representative function, noncognitivism seems to be the best option for the nature of moral judgments. At this point, something like MJI began to be taken as criterion of membership for moral judgments. Judgments without that motivational quality are not considered *moral* judgments.

There are some inaccuracies in the noncognitivist interpretation of MA's premise 2. Even though beliefs are not inherently motivational as a matter of conceptual truth or as a matter of its primary function, Hume argues that beliefs can materially cause passions and hence actions. More specifically, in 1.3.10, *Of The Influences of Belief*, Hume distinguishes ideas that influence the will by giving rise to passions, which ultimately lead to actions: "Nature has implanted in the human mind a perception of good and evil, or in other words, of pain and pleasure, as the chief spring and moving principle of all its actions"[78] Then in 2.3.9, *Of the Direct Passions*, he describes how ideas about prospective pains and pleasures excite passions and influence the will:

> "The mind by an *original* instinct tends to unite itself with the good, and to avoid the evil, tho' they be conceiv'd merely in idea, and be consider'd as to exist in any future period of time. (…) [And then], "desire arises from good consider'd simply, and aversion is deriv'd from evil. The will exerts itself, when either the good or the absence of the evil may be attain'd by any action of the mind or body." [79]

Note that, as MA's rendition C suggests, ideas do influence the will, they just don't do it alone or *immediately*:

> "Did impressions alone influence the will, we shou'd every moment of our lives be subject to the greatest calamities; because, tho' we foresaw their approach, we shou'd not be provided by nature with any principle of action,

[78] T 1.3.10, pg. 81
[79] T 2.3.9. pg. 280

which might impel us to avoid them. On the other hand, did every idea influence our actions, our condition wou'd not be much mended. For such is the unsteadiness and activity of thought, that the images of everything, especially of goods and evils, are always wandering in the mind; and were it mov'd by every idle conception of this kind, it wou'd never enjoy a moment's peace and tranquility. Nature has, therefore, chosen a medium, and has neither bestow'd on every idea of good and evil the power of actuating the will, nor yet has entirely excluded them from this influence."[80]

So, if Hume has established that beliefs in the future of existence of a pleasure or pain influences the will and ultimately produces action, then we can at least say that reason has some role in moral judgment. It was an early noncognitivist mistake to think that since beliefs are mental entities with truth value, all they can produce (employed as premises of arguments) is a conclusion (another belief), and hence can't influence action. Then, they went a step further in claiming that moral judgments are completely devoid of cognitive qualities. Now, a third and more specific mistake is to assume a necessary connection between moral judgment and motivation, which leads us to MA's premise 1.

As I understand Hume's view, the following seems to be closest to it. We have seen that it is a material inference that excites passion. I can infer x about other person's sentiments as well as about my own sentiments. That inference, in turn, may or may not elicit a desire. But my desiring being elicited or not has nothing to with the legitimacy of my believing in such inference. We often observe people sincerely holding moral beliefs which often do not have sufficient or decisive motivational force to result in action. And importantly, that happens not because of weakness of will, but because the influence of one's moral belief provides only one motivational vector for a rational agent, where the final action is the result of a competition between many other vectors, provided by other normative domains such as political, legal, prudential, aesthetic, economic, and so forth. My beliefs relative to those domains have their own contingent motivational influences. So, I believe that something like MJE would be more congenial to the Humean view.

[80] T. 1.3.10, Pg. 82

If the motivation argument is not a defense of noncognitivism, then what is it? Let us see what it would be for reason of itself/alone/immediately to motivate action. According to Hume's reading of Clarkian rationalism[81], our capacity of reason shall recognize the "eternal fitnesses and unfitnesses (…[which] impose an obligation, not only on human creatures but on the Deity himself"[82]) of a range of possible actions and this very recognition excites passion and produces action. In this manner, we can say it is the very belief in an action's fitness which leads to action. Now two Humean reminders, first the beliefs he assigns practical influence are beliefs about prospective pleasure and pain, not beliefs about actions' reasonableness or the lack thereof. Second, as we have seen in the Representation Argument, actions (and desires) can't be rendered reasonable or unreasonable, because there are not representational entities.[83]

It's not only that the connection can't be shown to be universally and necessarily binding; it is also that one of its relata doesn't exist. *It is in this Clarkian sense that reason is impotent.* There is no rational process capable of perceiving this very feature of actions, because this feature does not exist. This is a negative thesis involved in MA. Now what is MA's positive corollary? We note that in all renditions of MA, Hume is talking about the possibility of actions having a certain property. Note that Hume says "moral good and evil". He is talking about the very actual goodness of actions, a property the perception of which would bind all rational creatures alike.

So, to my eyes, in the MA Hume is talking directly about actual moral goodness and badness of actions and that our perception of these properties, eventually beliefs, have motivating

---

[81] Clarke held that morally good acts are reasonable in the same sense in which mathematical theorems and simple truths of physics are reasonable, and the moral goodness and badness of actions consists in their reasonableness and unreasonableness.

[82] T, 456

T. In order…to prove, that the measures of right and wrong are eternal laws, obligatory on every rational mind, 'tis not sufficient to shew the relations upon which they are founded: We must also point out the connexion betwixt the relation and the will; we must prove that…in every well-disposed mind, it must take place and have its influence. Also, Clark pp 199

[83] T 415

power, while the reasonableness of actions does not. MA concludes that actual moral goodness and badness *themselves* as properties of actions are not conclusions of reason, not that judgments are not the offspring of reason. It is rather an argument that moral good and evil are not identical to, or instances of, reasonableness and unreasonableness. Consequently, it is not an argument for noncognitivism, the clam that moral judgments do not represent the world. Nor does it entail noncognitivism, because all Hume has said so far, moral goodness and perfidy might be objective properties of people's passions and actions. The Direct Argument shows us that whatever these moral properties are, they could not consist in the passions' or actions' inherent reasonableness, since that is a property that no passion or action can have. The MA makes a similar point: the moral goodness or badness of an action has a feature – motivational influence – which the reasonableness or unreasonableness of an action cannot have, because there is no such property of actions to have it. Therefore, goodness is not reasonableness – not the offspring of reason. In detail, because the reasonableness and unreasonableness of actions are nonsensical properties, we cannot recognize them; therefore, the recognition of them could not be what moral judgment consists in; and so, it cannot be the recognition of the reasonableness and unreasonableness of a prospective action that moves us to perform it.

Sure, moral properties, if they exist, shall enter the mind via perception. If our perception of it accurately track reality and we are certain of it, we may say we *believe* that action Phi has property m. So, a moral judgment, whatever they are, may or may not be motivationally efficient. But again, that is another question. In my view, whether a moral judgment ultimately motivates is contingent on its own stringency as well as on the stringency (rational weight) of other normative domains (such as prudence). But the MA is not about whether moral judgments are descriptive. It only says that moral judgments cannot be attributions of reasonableness to actions, because this

very idea is incoherent. So far as the MA is concerned, moral judgments could be descriptions of other properties of actions and passions. Hence, I am sympathetic to Cohon's final rendition of MA:

1. The reasonableness of passions and actions cannot move us to have or perform them.
2. The moral goodness and evil of passions or actions (often) do move us to have and perform them.
3. Therefore, the goodness of an action is not its reasonableness, nor is its evil its unreasonableness.

2.2. Membership Criterion: A preliminary objection to MJI

There is another problem with the noncognitivist interpretation which I have hinted before. Noncognitivism accepts judgments as *moral* judgments only when they are necessarily motivational. Let us not forget we are also talking about what morality is. Noncognitivists credit the distinction of ethics to its special motivational. According to this view, unlike other normative spheres, morality is the only one whose judgments necessarily motivate the agent who endorses them to act accordingly. If we believe that donating our unused clothes to the homeless is good, then, necessarily we are motivated to do it, or something else is wrong with us. We may be lying to ourselves to feel better, or we may have some mental disorder that draws our will to live away. But under normal conditions of rationality, people are usually more strongly motivated to act in accordance with their moral convictions than according to other normative convictions.

However, we may be simultaneously motivated to do a number of things. To note this, imagine that I am a pharmaceutical researcher and an avid jazz supporter. I learn that tonight there will be a jazz performance near my home. I believe that I must go to this show. Of course, I'm motivated to go to the show. But I am also convinced that I must finish my science project. The night is cold and so I decide to stay home. But now imagine that on the way to your job interview, you come across a child drowning in a lake next to you. Sure, you are motivated for your interview,

very motivated indeed. Notwithstanding, the corresponding motivation to your conviction that it is your duty to save the child is much stronger. You miss the interview and jump in the water. Note how someone may be motivated by various normative judgment. I can feel equally motivated by my prudential judgments, as well as in accordance with my aesthetic judgments, and so on.

If we can plausibly claim that an agent may be simultaneously motivated by different reason-judgments, and that the competition between the reasons is fierce, then it becomes less plausible to claim that motivation my moral judgments is always superior. One important refinement of this view is to say that morality has a greater rational force than all other normative domains. In other words, it is the view that moral reasons are always motivationally superior to all others. I will discuss this rendition of moral rationalism in the next section.

Now, let us look at MA's premise 1: What is Hume's argument/point for the view that morality [necessarily] influences actions? Note how what once were observational impressions about moral behavior furtively turned into a modal claim about the very concept of morality, and what genuinely counts as a moral judgment. Pick any judgment, if it is not necessarily motivational, it can be many things, but *moral* judgments. And along the same paragraphs, Hume provides a direct argument against moral rationalism:

1. "Actions may be laudable or blameable…"
2. "…but they cannot be reasonable or unreasonable…"
3. "Laudable or blameable, therefore, are not the same with reasonable or unreasonable"

2.3. **Introduction to Conceptual Moral Rationalism**

In this project I think that the relevant metaethical phenomenon to be explained is not *lack of motivation,* but that of *conflicting motivations*. This assumption led me to think that moral rationalism is a good starting point for analyzing conflicts of motivations. According to moral rationalism, if an agent has a moral requirement to ϕ, then he has a decisive or overriding reason

to ϕ. In other words, that what the agent has most reason to do is exactly that what the agent has a moral requirement to do. This rationalist claim can also be stated as the view that principles of morality are necessarily symmetrical to the principles of practical reason. Thus, when we ask ourselves what we ought to do, from an all-things-considered standpoint, the answer is that we ought to conform to the commands of morality and not doing so implies practical irrationality.

Michael Smith defends a version of moral rationalism on conceptual grounds. For this particular view, moral requirements are always decisive or overriding practical reasons as a matter of conceptual truth. Smith defends conceptual moral rationalism (CMR) based on: (a) the truth of the practicality requirement[84] (PR) and (b) the objectivity of moral judgments.[85]

PR is the thesis that "if an agent judges that it is right for her to ϕ in circumstances C, then either she is motivated to ϕ in C or she is practically irrational."[86] In support of it, Smith shares his intuition that if someone utters a moral judgment that ϕ is right or recognizes that she is required to ϕ, but is not really motivated to act accordingly, then she is not being sincere[87] about her moral utterances or moral standings. So, for instance, if I say that I believe that corruption is wrong and I find myself required not to be corrupt, but even so I do not have any motivation to act accordingly, then either I am lying, or I am irrational. But this is not an argument on which Smith relies for his defense of PR. He actually believes that the best defenses of PR are either to argue that "being suitably motivated is a condition of the mastery of moral terms"[88] or that "it is a striking fact about moral motivation that a change in motivation follows reliably in the wake of a change in moral

[84] Smith, 1994. pp. 68-76
[85] Idem, pp. 77-91
[86] Idem, p. 61
[87] Idem, p. 07
[88] Idem, p. 71

judgement."[89] I will only consider the first argument, since it is the one directly connect with the issues in moral language I have tackled along this project.

The mastery argument is given in terms of an analogy. It compares a person who expresses a moral judgement and a second who expresses a color judgement. However, the latter is blind from birth. Smith argues that we are prepared to say that she does not possess *full* mastery of color terms. Although she is able to employ color terms for practical purposes (e.g. She may have memorized that her house is green, so she can use the information as a reference in case, say, she gets lost and needs to ask someone's help), she lacks the suitable visual experiences that are essential contents of color concepts mastered by sighted people. In the case of moral judgements, if the agent does not have the appropriate motivation that is essential to the sincere expression of these moral judgements, then she does not have *full* mastery of moral terms. This analogy is not very compelling because it is controversial whether the causal contact that holds between the observer and something external to her that results in color perceptions and color language is metaphysically the same as the case when we talk of the motivational property of moral judgements and its relation with the agent. But I propose to consider for a moment that there may be some compelling argument in favor of understanding moral concepts in terms of secondary qualities and that it renders Smith's analogy plausible. Thus, PR would be true so far.

Although I agree that the amoralist objection to PR, namely, someone who acknowledges the existence of moral considerations or demands and remains unmoved[90] establishes a significant conceptual case against the practicality requirement, I am initially willing to accept PR. There is a widespread intuition that moral reasons or moral requirements have a privileged position or advantage in relation to other realms of reasons or practical requirements, as for instance aesthetic

[89] Idem, p. 71
[90] Brink, 1997

reasons, professional reasons, or prudential reasons. Though in disagreement, Foot notes that "it is supposed that moral considerations necessarily give reasons for acting to any man."[91] And there are uncountable ways in which moral judgments are claimed to have some special force in our overall practical considerations about what to do ranging from sentimentalists to rationalists.

If we accept PR for now, we can move on to the question: "Should we accept that there is also a conceptual connection between facts about the rightness of actions and facts about what we have reason to do?"[92] As this question was proposed by Smith seems to me unproblematic. It is true that that what one judges to be the right thing to do necessarily gives one at least some reason for action. However, it may well be the case that what one judges to be the *rationally* right thing to do is not the same thing as the *moral* thing to do. In trying to give Smith the best interpretation possible, this question can surely be better phrased to avoid equivocation between what we take to be the "right" thing to do and the "moral" thing to do. Thus, I think it would be more philosophically interesting if we substitute "rightness" for "morally required". Also, it is important to mention that what is in question is whether, given the truth of PR, one has *decisive* reason for action. So, I rephrase the question as: should we accept that there is a necessary connection between facts about what one is morally required to do and facts about what one has *decisive* reason to do? Those who, like Smith, defend CMR will say "yes". In my view the correct answer is "no".

For an anti-rationalist position if such a connection exists, it is a contingent one. But basically, anti-rationalism is the view that it is not the case that an agent's decisive or all-things-considered reason to act is *always* a moral reason. An agent may not act justified by a moral reason and still be perfectly rational. For this view, the act of judging about what we have most reason to

[91] Foot, 1972, p. 309
[92] Smith, 1994, p. 77

do is a process that involves weighing different normative factors, among which morality may have a privileged starting point, but it is certainly not the only factor.[93] Broadly speaking, anti-rationalists[94] have argued that there are rational justifications for why non-moral reasons may occasionally override moral requirements. Brink offers a good way of framing the problem in dispute between rationalists and anti-rationalists:

1. Moral requirements – i.e. other regarding obligations – apply to agents independently of their aims or interests.
2. Moral requirements necessarily provide agents with overriding reasons for action.
3. Reasons for action are ultimately dependent on the aims or interests of the agent who has them.
4. There is no necessary connection between other-regarding action and any aim or interest of the agent.[95]

Moral anti-rationalists deny 2. But the rationalists try to avoid 4 by arguing that 2 is a conceptual truth. But how can that be, if we acknowledge that there is an inevitable asymmetry between self-regarding requirements and genuine other-regarding requirements in decision-making? If we have requirements, other than moral, that are at least sufficiently rationally justified, like for instance, reasons related to the pursuit of our individual personal[96] perfection, then the connection between that what we have most reason to do and that what we have *moral* reason to

[93] It is important to mention that there is disagreement among anti-rationalists on whether morality does have some sort of special or privileged starting point in comparison to other realms of practical reason. Some may well deny that and say that morality is on a par with them and it is equally rational to act either motivated by moral requirements or by, say, prudential requirements. Others prefer to say that even though there is always necessarily some sufficient reason to fulfill moral requirements, it is not irrational not to conform to them in certain circumstances.
[94] Foot, 1972; Wolf, 1982; Brink, 1986; Dorsey, 2012
[95] Brink, 1992. P. 22
[96] It seems to me that libertarianism can be an example of a political theory that assigns high normative stringency to self-regarding reasons. For instance, it has been argued that based on the right of full self-ownership, every individual is free to pursue his own welfare and acquire as much property as he sees fit. However, the set of libertarian theories also run against the same asymmetry once it is constrained by (different interpretations of) the Lockean proviso. This proviso issues an other-regarding reason as it says that individuals are indeed free to acquire things, *but* insofar they leave "enough and as good" to others so that others also can pursue their own welfare and/or personal perfection. Cf. Locke, 2010, p. 288.

do is *contingent*, because it is plausible that moral reasons be overridden by non-moral reasons. A conceptual truth is based on a modal claim of necessity, so if the connection defended by rationalists is actually contingent, it cannot justify a conceptual truth about what we have most reason to do. It follows that CMR is false.

I acknowledge that my current theoretical decisions invite questions about how reasons in general, moral and all non-moral ones can have their strength measured and compared, so that it yields the sort of decision-making I am fostering here. Moral anti-rationalism indeed is a large theory and thus a long-term project. Notwithstanding, elaborating an objection to Smith's main argument in favor of CMR seems to me the right entrance to a future larger contribution to moral anti-rationalism. Hence, in the next section I will present and discuss Smith's main argument in favor of CMR.

2.3. Against Smith's Expectation Argument for Conceptual Moral Rationalism

The thesis that Smith is interested in defending is that "if agents are *morally* required to ϕ in circumstances C, then there is a requirement of rationality or reason for all agents in C". In other others, he defends that premise 2 of the puzzle presented above is a conceptual truth. Smith's argument that I want to focus goes as follows:

1. The only thing we can legitimately expect of rational agents as such is that they do what they are rationally required to do.
2. Moral requirements apply to rational agents as such.
3. It is a conceptual truth that if rational agents are morally required to act in a certain way, then we expect them to act in that way.
4. Being rational as such must suffice to ground our expectation that rational agents will do what they are morally required to do.
5. Moral requirements are themselves categorical requirements of reason.[97]

[97] Smith, 1994, p. 85

This argument tries to show that every time an agent asks herself what she ought to do from a rational point of view, we should expect that she will do what she is morally required to do. So, it seems very plausible to interpret Smith's argument as defending that moral reasons are decisively overriding reasons for action. Premise 3 is clearly doing much of the work in this argument. So, the pressing question is: "Does *expect* mean that we expect that agents *should* or *will* do that what is moral?" In my view, the answer is, contrary to Smith, that we actually only expect that people *should* act according to the demands of morality.

In order to understand my position, we can start by picking some principle whose substance is widely agreed to be moral, such as being impartial or telling the truth. It seems to me that just like act-consequentialism faces the demandingness objection[98] in relation to versions of the principle of impartiality, we can assume that the deontological moral principle that we should always tell the truth, no matter its consequences, is certainly as stringent and demanding as the moral principle of impartiality, as both impose high standards of self-sacrifice. Notwithstanding, I recognize that the principle of truth may not be as morally ecumenical as the principle of impartiality, since act-consequentialism could perhaps do without the former. Still, just like the principle of impartiality, the principle of telling the truth also seems to impose great sacrifices of prudential interests. Now, the question can be read as: should we expect that agents, when fully rational, will always do what they are morally required to do, even when it imposes harsh self-sacrifices on them? I do not think so.

The reason why I hold this view is that Smith's argument conflicts with the common phenomenology of trust and distrust. Assuming that CMR tells us that people will always do what they have moral reason to do, if they acknowledge the advice from their fully rational counterparts,

[98] Basically, the idea that a life fully lived strictly in accordance with the norms of morality impose self-sacrifices that extrapolate the limits of a prudential rationality.

it would be foolish if we ever distrust any rational person that she will do what is moral or, in other words, that she will conform to the relevant moral reason in question. Moral rationalism misguides the one who trusts, because it makes him believe that the trusted party will necessarily fulfill his expectations.

Trust is based on an assessment of the interests of the trusted. And it would be too risky to believe that the trusted, if rational, will necessarily have as the all-things-considered justified action the one that fulfills the moral requirement of always telling the truth or being impartial. In my view, moral rationalism cannot account for the actual instability of relations of trust when both parties are rationally justified. In other words, the one who trusts is rationally justified in thinking that there is a chance, perhaps even a high chance, that the trusted may not be telling the truth or being impartial in a given circumstance. Observation of everyday social relationships provides plenty of evidence that people indeed lie, break promises, or act expedient. So, although we are often willing to trust someone, we do not want to trust *blindly*. On the other side, the trusted party could perhaps be justified in *not* telling the truth if that option imposes too high a self-sacrifice. Therefore, if we agree that both parties are sufficiently rationally justified, then moral rationalism, as based on the argument from expectation cannot account for the phenomenon of distrust in sufficiently rational agents.

Think for instance of a Sunday Pre-Owned Car Selling Festival, a meeting where private owners get together to sell and buy all sorts of used cars. Now I am at the Festival looking for a car for myself. I have just landed a new job in Kansas City, but I am still a graduate student at KU, and living in Lawrence remains cost-wise optimal, so I need a cheap and reliable private transportation. Consider also, for the sake of the argument, that I know that it is an empirical truth that used cars have or will certainly have some minor problems, sometimes major problems. So,

we know ahead that these cars will rarely be in perfect conditions. Here I am, looking at an apparently fair shape car that fulfills all my criteria. However, the car is beyond 140.000 miles already. It may be the case that the car has a compromising major problem and the owner is just trying to get rid of it. Or, for all I know at this moment, it may still be running pretty great. I ask the owner about the car and he answers: "The car is in perfect conditions." I have no expertise in car mechanics and all I have is the seller's word. Should I trust that the seller is saying the truth?

The first question that is reasonable to expect from Smith is whether the seller is rational. In the previous section I showed that we lack good evidence to think that PR is not true. So, as the buyer, I am allowed to expect that his rational self gives him at least some sufficient reason to be telling the truth. But when I engage in a process of decision-making on whether I should buy the car, would it not be ingenuous to neglect the obvious fact that it is in the self-interest of the seller to sell the car? It seems rational for me, as the buyer, to consider that maybe the car really has some compromising problem which will cost me a lot of money, and especially whether the seller is lying to me because he is in a financial hardship and therefore has a strong reason not to tell the truth about the car's conditions as that would make it impossible for him to get rid of the car for a decent price. Thus, it would be at least sufficiently rational for the seller not to tell the truth, if doing it will impose significant burdens that he cannot cope with.

It may be argued that the seller is not rational or that he is simply acting irrationally because he is overlooking what his fully rational and informed counterpart prescribes him to do. For, Smith indeed assumes that agents are irrational by their own lights if they fail to have the desire (in line with the moral requirement in question, namely, telling the truth) that their fully informed rational counterparts would have them to desire.[99] My response is that we can still hold that it is irrational

[99] Sobel, 2001, p. 481

not to conform our behavior or desire as our ideal fully informed counterparts would advise us, but there is no reason to think that considerations of self-interest have no shelter in the fully informed counterpart of the seller, or of any rational agent. Therefore, I see no compelling reason to believe that the seller is *really* irrational, if he does not tell me the truth about the car.

Still, not satisfied, one could raise the objection that being in a hardship presses the agent to act in not-morally warranted prudential manner. However, that can be fixed by qualifying the seller as a professional car seller. Now we are at an official pre-owned car dealer. The car has a compromising problem, but the seller really needs to fulfill a minimum number of monthly sales. As it happens, not only has he not sold many cars this month, but, also, last month he fell far short from his minimum goal. The shop has been struggling to pay employees. "We need to keep the ship afloat", he ruminates. Should not the seller have at least a *pro tanto* professional reason to lie to me? In both sellers' cases, I am inclined to agree with Foot that:

> "The fact is that the man who rejects morality because he sees no reason to obey its rules can be convicted of villainy but not of inconsistency. Nor will his action necessarily be irrational. Irrational actions are those in which a man in some way defeats his own purposes, doing what is calculated to be disadvantageous or to frustrate his ends. Immorality does not necessarily involve such thing"[100]

Hence, if we agree that the involved parties are not irrational, when I am back to my problem with buying a cheap car, I see no reason for why I should expect that the seller will necessarily tell me the truth. It seems possible to imagine cases in which the seller would be rationally justified by some non-moral reason to not tell me the truth. Therefore, there is no reason for me to blindly and ingenuously trust his word. Had the expectation argument for CMR convinced me, it would have distorted the common phenomenology of trust and distrust that I and everyone else around are familiar with. That happens because in case Smith was right, I would

[100] Foot, 1972. p. 310

have firmly believed that the result of the seller's rational decision would have been necessarily in favor of the moral norm in question, namely, telling the truth. But if Foot is right, from the fact that the seller is rational, i.e. that he usually conforms to what he sees as rationally adequate, it does not follow that his all-things-considered reason will be a *moral* reason.

As I have been trying to show it is possible to agree on two conditions that Smith also accepts: (a) that the agent is rational because he does acknowledge to the prescriptions of her fully informed counterpart, and (b) that both involved parts are aware of what is the moral principle involved in the circumstance. We could even agree with Smith that if agents are sufficiently rational, they will come up with the same judgments about what they are morally required to do.[101] Notwithstanding, CMR leaves very little room for rational distrust, because it makes us think that if people are rational, they will necessarily take their moral reasons as all-things-considered overriding. In a world where people can reasonably be considered rational, we would not be justified in having doubts about what are people's all-things-considered reasons at a given case.

But as I tried to show, we have reasons to believe that people may not be trustworthy, even when they are rational, not because they morally evil, but because they have sufficiently justified reasons other than moral. Furthermore, trusting would not be at issue, if there were no possibility of being betrayed by rational agents. CMR cannot provide a satisfactory account for these cases because its argument from expectation fails to provide a solid basis for it. If my position here defended is plausible, the next step for future investigation is to provide an account of how moral reasons and non-moral reasons are weighted up against each other and how can we measure their differences in justificatory strength.

2.4. The argument from fetishism

[101] Smith, 1994, p. 86-7

But Smith has a rejoinder against MJE in form of a dilemma to the externalist. Smith claims that it is a striking fact about good and strong-willed people that their moral motivation reliably tracks the changes in their moral judgments. Then he defies the externalist to explain this putative fact. Call it the *tracking condition*. The internalist explains the reliability by way of a conceptual connection between moral judgement and motivation. The very concept of moral belief is such that it necessarily elicits a corresponding desire. But the externalist denies such conceptual connection. So, what is the externalist explanation? Smith points out to two options: non-derivative and derivative desires to do what one believes to be morally right. Smith calls non-derivative desires as *de re* desires and derivative as *de dicto* desires. The reductio then is that the externalist explanation using a *de re* desire does not meet the tracking condition, while the externalist deployment of a *de dicto* desire amounts to moral fetishism. Since these are the only available options, externalism is false.

A *de re* desire for something is a want for this thing's intrinsic features or base-level facts. A *de dicto* desire for the same thing is a want in virtue of the normative inscription under which this thing is classified. Suppose Kyra loves jazz. She desires to listen to high quality jazz and picks Kind of Blue. Her *de re* desire to listen to the album comes from its base-level facts such as melody, rhythm, tempo, tone, etc. which make it one of the most acclaimed jazz albums. On the other hand, supposed Kyra wants to listen to jazz and is a sucker for lists of top acclaimed albums. She learns about Kind of Blue and picks it because it is on the list. Now Kyra has a *de dicto* desire to listen to the album since the desire is a function of the album being classified as one of the most acclaimed, not of its base-level facts. Kyra's *de dicto* desire is derivative given that it derives from

an original desire to listen to whatever is on the list of most acclaimed jazz albums.[102] Now, a jazz connoisseur would frown upon Kyra's *de dicto* desire since that seems the *wrong motivation*. For a jazz connoisseur, Kyra is not a genuine jazz lover, but merely has a fetish for lists of most acclaimed albums.

Now back to Kyra's desire to do what is morally right. On the first horn of the externalist dilemma, *de re* moral desires cannot ground the tracking condition, because once moral beliefs are not necessarily motivational, a change in moral beliefs has no necessary implications for motivation. Kyra may believe in the rightness of famine relief and yet have no correspondent *de re* desires. Hence the last externalist resort is to appeal to a *de dicto* desire. To see how that unfolds take the following change in moral judgment:

> ***Kyra***: Kyra is a good and strong-willed person and so she usually desires to do whatever turns out to be right. She used to believe that donating to famine relief is a responsibility of governments and big companies. Hence, she was not motivated to donate. Today she has been convinced that every individual citizen is also proportionally required to donate for famine relief. Since she is a good and strong-willed agent, she is motivated to give her fair share.

Suppose the externalist appeals to *de dicto* desires to explain the tracking condition. One of the signs that Kyra is a good moral agent is that her desires have changed in accordance with her moral beliefs. But her desire to relieve famine derives from a standing *de dicto* desire to do what is morally right, not the badness of famine and/or goodness of relief *in themselves*. When she is certain that individually contributing to famine relief *is* morally right, her motivation to contribute will derive from her *de dicto* desire to do what is morally right, not from *de re* aversion to famine and *de re* desires for its relief. However, Smith argues that even though Kyra's strong

[102] It seems to me that the same rationale can be applied for any desires vs. type/token situations. Instead of music, Dreier makes the case in culinary terms where a particular dish is the token and a given list of dishes, such as the dishes he likes.

de dicto motivation could account for the tracking condition, it is not what commonsense takes to be characteristic of good moral agents. He writes:

> "For commonsense tells us that if good people judge it right to be honest, or right to care for their children and friends and fellows, or right for people to get what they deserve, then they care non-derivatively about these things. 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."[103]

So, a *de dicto* motivation to relieve famine is fetishistic because instead of originally deriving from (the beliefs in) the base-level facts that make it morally right, it derives from a (belief in a) remote aspect of the action. This remoteness is the distancing suggested in Williams' one-thought-too-many argument based on the wife's case. In that case, the husband is a moral fetishist because his acting desire to save his wife derived from a *de dicto* desire to do what is morally right, instead of arising directly from the situation's base-level facts, i.e., that it is his wife.[104] Smith's intuition is that a good and strong-willed agent would frown upon Kyra's *de dicto* desire in the moral case for the same reason in the jazz case; she is not a genuine moral agent just like she is not a genuine jazz connoisseur. Since an externalist ends up committed to moral fetishism, Smith concludes that his Practicality Requirement remains the best candidate for explaining the tracking condition in good moral agents.

Smith's depiction of moral motivation is wrong in two ways. First, it underestimates the dynamics between *de dicto* and *de re* desires. Second, in a larger context, Smith's psychology of the good and strong-willed person is the psychology of moral saints and thus vulnerable to the same objection, *viz*. moral saints' lives are unappealing.

2.5. Strong-willed moral agent x Upright individual

[103] Smith, 1994 p. 75

[104] An important difference between this case and the ones involving the tracking condition is that is does not involve a *change* in moral judgment.

To be clear, the externalist has no qualms with *de re* desires to perform morally right actions. But neither does he with *de dicto* desires to conform with the requirements of morality. Sure, it is often the case that good people have *de re* desires derived from base-level facts about welfare, autonomy, harm, respect for persons, etc. But before a first order moral reflection, there is nothing about these *de re* desires that make them necessarily overriding from the get-go (assuming such force would derive from the supremacy of the base-level reasons they track). Think of someone under uncertainty about which of his *de re* moral desires he shall satisfy.

Simon: Suppose Simon lives in a small town and is that kind of good guy, usually imbued with *de re* desires to maximize net happiness. He cares *de re* about sentient beings' wellbeing. Simon doesn't think twice before going off his way to help everyone around him, even when he is in doubt about whether he really should. Simon was brought up as a fisherman and has no formal 101 Ethics. But Simon has a sense of his own moral fallibility, *viz.* he might be wrong about what is it that really ought to do for others. He wonders whether sometimes he should let people strive for themselves, since "Give a man a fish and you feed him for a day. Teach a man to fish and you feed him for a lifetime." And after all Simon holds simultaneous *de re* desires to respect people's liberty. So, Simon is torn and wondering whether he ought to respect people's autonomy against utility maximization.

Simon feels puzzled about his *de re* desires for wellbeing and liberty. As a consequence, he assumes an epistemic attitude, namely, he wants to know what are the moral requirements that actually apply to him. It's hard to deny that Simon's epistemic attitude before his conflicting *de re* moral desires is a genuine sign of a good moral agent. But note, it is Simon's standing *de dicto* moral desire which prompts him to engage in first-order moral investigation. And a *de dicto* desire can be a heavy task to sustain since it not only competes with other desires but also is vulnerable to all sorts of interferences, distractions and impediments. For example, Simon could surrender to his *de re* desire to end somebody's pain even knowing that respecting autonomy was the moral thing to do. Since this epistemic attitude couldn't take place had Simon not had a *de dicto* moral desire, the relentlessness of this desire is a legitimate quality of an admirable moral agent.

Furthermore, the lesson to be learned here is that the epistemic attitudes we take in face of our moral uncertainty and fallibility are both marks of a good moral agent.

Simon is a case of conflict between desires. What if Simon had absolutely no *de re* desire at all to do what morally requires? In other words, no desire arisen by (his beliefs in) the base-level facts which make the action morally relevant. Wouldn't be strange for Simon to act solely by a *de dicto* desire? The fact is that *de dicto* moral concerns also play a role in our wanting to acquire *de re* desires. Imagine someone like Russ:

Russ: A long-time retired cargo airplane pilot, Russ has always been motivated by *de re* concerns for the people around him. Living in a middle-of-nowhere town in Nevada, Russ has never been confronted with tough questions about morality's demandingness *vis-à-vis* individual sacrifice. One day, as a matter of historical warfare crossroads, the nation's destiny happens to be in his hands. But the only way available to end this deadly threat requires him to sacrifice his own life.

I find very difficult to morally blame Russ for not having a *de re* desire to commit suicide, even if for the nation. Now, imagine that in the heat of this tough call, Russ remembers his training soon after WWII, during which he had warfare ethics classes. One of the passages in Aristotle especially caught his attention. There it reads:

"For even if the good is the same for a city as for an individual, still the good of the city is apparently a greater and more complete good to acquire and preserve. For while it is satisfactory to acquire and preserve the good even for an individual, it is finer and more divine to acquire and preserve it for a people and for cities. (Nicomachean Ethics, 1094b)

That quote stuck in Russ's mind. Since then, Russ has sustained a desire to do what is morally right. In the lack of a *de re* desire, this *de dicto* desire is Russ's last resort for motivating himself to sacrifice everything he has. It would be admirable had Russ had a *de re* desire for the wellbeing of his country strong enough to lead him on. Notwithstanding, Russ's *de dicto* sense of morality strengthens his heart and comforts his mind in this moment of extreme fear. The lesson to be learned here is that in the absence of *de re* desires, a *de dicto* concern for what is right, call

it architectonic desires[105] can play a legitimate function in the psychology of good moral agents. Lillehammer for instance argues that a *de dicto* desire could even outweigh counter *de re* desires. Such as Russ's desire to avoid his death.[106]

One objection. Imagine that I know nothing about jazz, but I got curious and so I decided to seek for paradigmatic instances of high-quality jazz. Since I had no idea of how jazz sounds, I sought for the list of the most acclaimed jazz albums. From the get-go, all I had was a *de dicto* desire to listen to what is on the list. Smith would say that I am not yet a good jazz fan until I sustain *de re* desires for appreciating the albums in the list. My answer is that after getting familiar with the albums in the list, I have developed a number of *de re* desires to listen to a good number of jazz albums in the list for their intrinsic base-level facts. My jazz case and *Russ* are ones which show that *de dicto* desires can be *maieutic*.[107]

Maieutic desires act like second order desires which generates or triggers *de re* desires. For example, my *de dicto* desire that I *de re* desire to listen to good jazz makes me ultimately develop *de re* desires to listen to good jazz. In the moral case, a *de dicto* desire to do what morality requires generates *de re* desires for promoting wellbeing, respecting autonomy, etc. It is often the case the arise of a *de re* desire to ϕ comes from the *a posteriori* identification between what is right and ϕ. Only after finding out that *Kind of Blue* corresponds to good jazz, can I develop my *de re* desires to listen to it. Likewise, only after finding out that promoting wellbeing or respecting autonomy correspond to moral rightness can we develop, maintain or jettison *de re* desires for these things. Before one ask, *de dicto* desires don't need to keep maintenance of *de re* desires since the latter, once up and running, may well derive their force from whatever base-level facts. So, *de dicto*

[105] Desires to act according to one's architectonic reasons, in this case architectonic moral reasons. Dorsey, Darwall
[106] Lillehammer, 191, husband avoid betrayal, father turning in murderer son to police, etc.
[107] James Dreier, "Dispositions and Fetishes: Externalist Models of Moral Motivation" in *Philosophy and Phenomenological Research*, Vol LXI, No. 3, 2000.

desires are not a problem. Crucial for the point being made about strong-willed moral agents is the *attitude* people have when confronted to all sorts of moral ignorance, uncertainty, and fallibility.

One last point will close this not only my responses to the moral fetishism argument, but also this chapter. Here is a concern about what kinds of agent we are talking about: jazz is not the only normative realm in the musical universe, so isn't a jazz-only connoisseur a bad example of a well-rounded *music* connoisseur? Second, it is hard do get even close to say that jazz is *superior* to other genres. If I am an *upright* music enthusiast, I am sure I will have reasons and desires pertaining to a number of other musical normative systems needless to name. Moreover, the upright music enthusiast's *de dicto* desire to listen to good music will lead him to develop *de re* desires for music genres he does not even know yet. Since he is aware of his limited musical knowledge facing the vast number of genres, he jumps into further musical investigation. I think we all shall agree his *de dicto* desire to listen to good music and correspondent epistemic attitudes are marks of an upright music enthusiast.

The same goes with practical reason, since morality is not in the only normative system that matters. Note that in all the cases above nothing is said about whether either *de dicto* or *de re* moral desires are *also* desires to do what is all-things-considered correct. Think of *de re* desires. Good people have *de re* desires to promote wellbeing, respect autonomy, and other moral actions, but also *de re* desires to feel healthy, succeed professionally, participate in electoral campaigns, bow to the Queen of England, etc. You see where this is going. Likewise, besides a *de* dicto desire to do what is morally right, good people have *de dicto* desires to do what is right prudentially, professionally, politically, socially, and so on. Smith's view neglects sources of reasons other than moral. So, even if all the desiderative mechanism defended so far takes place within the perspective of morality or of any given normative system, that is just what happens within one normative

system among many recognized by the "upright individual".[108] An upright individual is one who is concerned about all the genuine practical reasons that apply to him. Hence, he may have a number of *de dicto* desires competing for becoming *the* desire to do what is all-things-considered correct.

All-things-considered desires are based on all-things-considered reasons. According to the attitudinal theory of an upright individual then, it is a sign of uprightness having the epistemic attitude of wanting to investigate what are his all-things-considered desires. Both *de re* and *de dicto* desires may exercise different motivational forces at different given times. So, bluntly speaking, an all-things-considered desire may play a regulatory role in deciding what one ought to do all-things-considered. Trying to find out which one of his desires would be best to satisfy leads the upright individual to weigh as many as possible reasons that apply to him at a given time. In this manner, part of what makes the uprightness of this individual is his epistemic attitude of wanting to know which desires to satisfy and which reasons ultimately apply.

[108] Dorsey, *Limits of Moral Authority*, pg. 33

Chapter 3 Queerness, Reasons, Alienation, Explanation

In the former chapter I tried to show that the motivational force of *moral* goodness depends on the motivational force of other areas[109] of normativity, such as economic, legal, prudential, moral, aesthetic, and professional, to name a few. So, according to my view, it is possible that nonmoral reasons outweigh or even neutralize moral reasons and thereby can prevent moral reasons from having an *ultimately deciding* motivational force. In this chapter I argue that even when one acknowledges that moral reason, they are not necessarily connected to motivation.

For motivational internalists, whatever we call moral must be necessarily connected with the agent's motivational set. In that manner, a judgment can only be qualified as moral if it is, *inter alia*, necessarily motivational. In the case of reasons, all reasons are then necessarily tied to motivation. In this chapter I discuss versions of reasons internalism. The argument can be summarized as follows:

1. Moral realists are committed to external reasons.
2. All external reason claims are false. Either because it alienates, or it can't explain action
3. Therefore, moral realism is false.

3.1. Conflict between objectivity and normativity and Universal Motivation

Mackie's metaethics is an instance of MJI-based problems for moral naturalist realism. First, a metaphysical problem: there are no "objectively prescriptive" concrete entities.[110] If they existed, they would be essentially motivational and since no objective items carry, in themselves,

[109] Mackie, "good in different contexts" p. 59-61
[110] Mackie, p. 26

any motivational property, such facts and relations would constitute outlandish entities. Second, apart from the possibility that intrinsically motivational objects and relations may simply not exist, there is an epistemological problem since we would not have cognitive access to observable objects and external relations of fact whose considerations are intrinsically motivating. Either one of the argumentative directions for moral realism is bound to fail. Hence, the semantic corollary is that moral (and normative) terms do not denote natural extensions.

Mackie opens *Inventing Right and Wrong* raising the conflicting intuitions about morality: while the sense of objective morality (that moral judgments can be true or false) *"has a firm basis in ordinary thought, and even in the meaning of moral terms*"[111], we have the intuition that "*there are no objective values*".[112] According to Mackie, morality may even be universal in the sense of intersubjectivity, but, more importantly, moral language expresses something that goes beyond mere intersubjectivity. Moral judgments aspire to describe objective states of the world. At the same time, moral language expresses something that goes beyond objectivity. Moral judgments are taken to have an internal motivating capacity. He contrasts this additional feature with the objectivity of morality:

> The ordinary user of moral language means to say something about whatever it is that he characterizes morally, for example a possible action as it is in itself, or would be if it were realized, and not about, or even simply expressive of, his, or anyone else´s, attitude or relation to it. But the something he wants to say is not purely descriptive, certainly **not inert**, but something that involves a call for action or for the refraining from action, **and** one that is **absolute, not contingent upon desire or preference or policy or choice, his own or anyone else´s**.[113]

Besides the internal motivational force of morality, Mackie claims that since morality's normative authority is universal, and hence its motivational force would apply universally. Everyone who correctly understands or honestly utters a moral judgment shall be equally motivated. It is this "**universal motivation**" which G.E. Moore *casually* mentions in *Principia Ethica*. Moore says:

---

[111] Mackie, p. 31
[112] Mackie, p.15.
[113] Mackie p. 33

"*it does seem to be true that we hardly ever think a thing good, and never very decidedly, without at the same time having a special attitude of feeling or will towards it; though it is certainly not the case that this is true universally*."[114] But that was within the context of denying that goodness is analytically definable as what is desired. According to Hurka, Moore never directly employed MJI in whatever way we know it as support for his open question argument. That is why I don't tackle that argument in this dissertation.

So, according to Mackie, at the heart of morality there is a tension between its motivational character and the universalizable character attributed to the motivational aspect itself.[115] On the one hand, moral language has this dynamic characteristic, which is its internal power to impel the agent to action and, at the same time, due to morality's universality, moral language seems to state propositions that apply equally to everyone independent of their attitudes, dispositions, or desires of the agents.[116] Mackie seems to assume that in order to guarantee universality, the defense of an objective dimension of ethics must postulate a domain outside the minds of those who consider a moral judgment. In this manner, he says, "*it will not be a matter of choice or decision whether an action is cruel or unjust or imprudent or whether it is likely to produce more distress than pleasure*".[117] However, satisfying the requirement introduces a difficulty for the moral naturalist realist, for now it is complicated to conceive of in natural terms items that are external to the mind and at the same time containing necessary motivational force:

---

[114] Moore, PE, parag. 79

[115] Because saying that moral facts are intrinsically motivating implies that the learning of such facts, to be genuine, must necessarily *equally* motivate everyone who learns them, otherwise they have not really learned the fact. However, I expect that my arguments lead us to think that the relativity of motivation, *viz.* people feel motivated in different degrees once they acknowledge a moral fact, should not preclude the epistemological legitimacy of such acknowledgement. Call that moral acknowledgement. As I said earlier, this bifurcation between either you know and are motivated, or you are not motivated and hence do not really know. descends from Socratic intellectualism.

[116] Perhaps worth noting David Ross argues that normative objectivity, moral, aesthetic, etc. does not preclude universal desires, but that objectivity would only make sense if there are desires that everyone could equally sustain in relation to a particular object, e.g. Everyone would desire to see Mona Lisa.

[117] Mackie p. 33

> But in satisfying this demand, it introduces a converse deficiency. On a naturalist analysis, moral judgments can be practical, but their practicality is wholly relative to desires or possible satisfactions of the person whose actions are to be guided; but moral judgments seem to say more than this. This view leaves out the categorical quality of moral requirements. In fact, both naturalist and non-cognitive analysis leave out the apparent authority of ethics, the one by excluding the categorically imperative aspect, the other the claim to objective validity or truth.[118]

Mackie concluded that it is impossible to find something in the world whose mere perception contains, in itself, a motivational element. But he seems to prefer to save internalism rather than objectivity. According to him: "*The claim to objectivity, however ingrained in our language and thought, is not self-validating. It can and should be questioned*".[119] As a result, Mackie advocates an error theory to accommodate the distinction between the two fundamental elements of morality: normative authority and internal motivation. We talk of moral qualities all the time, but despite our projecting prescriptive qualities or entities onto the natural world, they do not actually exist. However, since such an error theory seems to go against common sense, Mackie offers two arguments. Let us look at the two arguments below.

3.2. Mackie's Arguments

First, the argument from relativity (AFR):

> The argument from relativity has as its premise the well-known variation in moral codes from one society to another and from one period to another, and also the differences in moral beliefs between different groups and classes within a complex community. Such variation is in itself merely a truth of descriptive morality, a fact of anthropology (…). But it is not the mere occurrence of disagreements that tells against the objectivity of values. (…) **The argument from relativity has some force simply because the actual variations in the moral codes are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions**, most of them seriously inadequate and badly distorted, of **objective values**.[120]

This argument is based on the apparent lack of solution to moral disagreement. The inconclusiveness of moral disagreements is attributed to the fact that people's moral choices are not true or false based on the independent authority of observations or evidence that justify them.

---

[118] Mackie p. 33. In practical reason lingo, one saves justificatory (normative) reasons at the expense of explanatory (motivational) reason, and vice versa.
[119] Mackie p. 35
[120] Mackie, p. 37-37

Our moral choices reflect a diversity of substantial normative theories internally adopted by individuals. Deciding what moral conclusions to follow or which moral theories are appropriate does not depend on an objective reality, but only on qualities intrinsic to the observer or the group it is part of. This concedes a theoretical tolerance to morality which we do not find in science. In a certain sense, there is no room for willy-nilly preferences for scientific theories, since we are rationally required to adopt only those theories that best explain the facts or have greater predictive capacity, simplicity, and so forth. While adopting moral theories seem solely justified by the way a person chooses to live.[121] And life choices in turn will be justified by the experiencing of *different* intrinsic values that makes one's life worth living. Since those intrinsic values seem all sufficiently justified, it generates a larger variety of theories whose adoption are all equally justified. Nevertheless, these theories often conflict or are inconsistent with one another. Hence adjudication is necessary. Now, see how high the ambition of the Mackian moral realist is. To settle the matter, he needs to point out to some external object, fact, or state-of-affairs which is normatively authoritative *and* universally motivating in and of themselves.[122] But there isn't anything like that in the world. Hence, there is no way to decide which theory is best according to an external authority.[123]

From the relativity of value endorsements (or acknowledgements) follows the argument from queerness (AFQ) which has two parts. A metaphysical and an epistemological:

> **Metaphysical**:
> Plato's Forms give a dramatic picture of what objective values would have to be. The Form of the Good is such that knowledge of it provides the knower with both a direction and an overriding motive; something's being good both tells the person who knows this to pursue it and makes him pursue it. An objective good

[121] Korsgaard, *Sources* p. 71
[122] Williams, Ethics and the Limits, p. 140
[123] Tersman (2006) identifies this phenomenon of tolerance with the concept of latitude to explain why moral disagreement would be somehow deeper than in the sciences. Latitude is characterized by the notion that ethics, unlike sciences, requires less agreement with basic theories, thus tolerating a wider spectrum of differences and idiosyncrasies as to the choice of normative theories we adopt, since the final decision is an intimate forum.

> would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. Similarly, if there were objective principles of right and wrong, any wrong (possible) course of action would have not-to-be-doneness somehow built into it. Or we should have something like Clarke's necessary relations of fitness between situations and actions so that a situation would have a demand for such-and-such an action somehow built into it.[124]
>
> If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe.
>
> **Epistemological**: Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ways of knowing everything else…
> When we ask the awkward question, how we can be aware of this authoritative prescriptivity, of the truth of these distinctively ethical premises or of the cogency of this distinctively ethical pattern of reasoning, **none** of our ordinary accounts of sensory perception or introspection or the framing and confirming of explanatory hypotheses or inference or logical construction or conceptual analysis, or any combination of these, will provide a satisfactory answer; 'a special sort of intuition' is a lame answer, but it is the one to which the clearheaded objectivist is compelled to resort.[125]

Despite the realist usage of moral language, from a natural point of view, there is *nothing in the world*[126] whose observed perception universally, and *necessarily*, motivates all and every perceiver. Therefore, not only we have no epistemic access to moral facts, but more importantly, they do not exist. This leads to the impossibility of external reasons.

3.3. Mackie and Motivation

The queerness is based on internalism. However, it is difficult to point out what exactly are Mackie's views about moral rationalism and whether he defends MJI, or merely endorses. Looking solely to the section where he presents the queerness argument, Mackie makes several different internalist claims. For instance, when drawing the contrast between moral judgments and factual judgments he uses Plato's Form of the Good[127] and its overridingness. So, we have the impression that Mackie endorses some version of moral supremacy. For instance, when establishing the

---

[124] Mackie, 40
[125] Mackie, p. 38-39
[126] What is the connection between the natural fact that an action is a case of deliberate cruelty – say, causing pain just for fun – and the moral fact that it is wrong? It cannot be an entailment, a logical or semantic necessity. Yet it is merely that the features occur together. The wrongness must somehow be 'consequential' or 'supervenient'; it is wrong because it is a piece of deliberate cruelty. But just what *in the world* is signified by this 'because'? Mackie, p. 41
[127] P.24,

queerness of objective (moral) values, he says: "*the Form of the Good are eternal, extra-mental realities. They are a very central structural element in the fabric of the world. But it is held also that just* ***knowing*** *them or 'seeing' them will not merely tell men what to do but will ensure that they do it,* ***overruling*** *any contrary inclinations."* and "*the Form of the Good is such that* ***knowledge*** *of it produces the knower with […] an* ***overriding*** *motive*."[128] So Mackie seems to hold that moral judgments have overriding motivations built into them via *moral knowledge internalism*, i.e. pure knowledge of the good imbues the knower with overriding motivation.[129] But he never explicitly defends moral supremacy, besides just indirectly displaying some inclination towards a prerogative in favor of morality's normative significance over other institutions within practical reason. [130] Then right under the reference to Plato, Mackie brings in Humean metaethics to illustrate how motivation is built into moral judgments or moral reasons, as they are devices that "automatically influence the will." At this point Mackie speaks only of *some* motivation. But Mackie does not endorse the classic Humean theory of reasons, so he is not the usual kind of Humean *internalist about reasons*. Still on the same page, he speaks of facts with "to-be-pursuedness" and "not-to-be-doneness" built into them, a view we can call *normative (moral) fact internalism*.[131]

Olson argues that fact-internalism is the closest interpretation of what Mackie called queer.[132] In this view, the queerness is in the facts. For Olson, this view is more congenial to Mackie's because it allows one to judge that a fact has this moral property, without there being a

---

[128] Mackie, p. 40
[129] As also commented by Sinnott-Armstrong, 60 *World Without Values 2010*
[130] Mackie, p. 40. Finlay, Stephen. The Error in the Error Theory. *Australasian Journal of Philosophy*. 86:3, 2008, and Sinnott-Armstrong, Walter. Mackie's Internalisms. In Joyce and Kirchin. *A World Without Values*. Olson exegetically separates those two views and shows that Mackie endorsed both MR and MJI. Olson, Jonas. *Moral Error Theory*, OUP, 2014.
[131] Dworkin says this view is utterly metaphorical, 1996, p. 114
[132] Jonas Olson, In Defense of Moral Error Theory In Michael Brady (ed.), *New Waves in Metaethics*. Palgrave-Macmillan (2011)

necessary connection between making the judgment and being motivated. Indeed, Mackie remarks that "*the lack of motivation is not logically ruled out*" in moral judgments.[133] If the motivation is not in the judgment, then it must be in the object of the judgment. On the other hand, other commentators[134] think Mackie's queerness argument is best interpreted as resting on judgment internalism, where the judgment is a belief about the moral qualities of facts and actions. In this case, one's belief about, say, the wrongness of killing, is necessarily motivated not to kill. Interpretative skirmishes aside, a thorough reading of *Inventing Right and Wrong* will show that Mackie talks in ontological as well as in epistemological terms. Correspondently, Mackie often discusses *there being* reasons, when he talks about moral and prudential (and many others) reasons, their normative weight, and the possibility of, for example, prudential reasons overruling moral ones.[135] At the same time, Mackie often talks of moral judgment as knowledge or belief in moral facts, and how these epistemic states would necessarily motivate the agent. Hence, we can discuss Mackie's argument's on resting on either version of internalism. That notwithstanding, I think the best way to understand Mackie's internalisms is through his *institutional* view about practical reason.

3.4. Practical Reasons Internal to Institutions

Mackie's reasons internalism is mediated by the concept of institutions. Ultimately, he concludes that claims (propositions) about categorical (external) reasons are all false because the facts they refer to would have inbuilt to-be-pursuedness.

---

[133] Mackie, p. 69
[134] Sinnott-Armstrong, Copp, and Dreier, all in *World Without Values 2010*
[135] how strong a reason, how easily *overruled* by other considerations, may be a matter of dispute." Mackie, p. 75

Let us see how that works. Mackie argues that the meaning of 'good' is "*such as to satisfy requirements (etc.) of the kind in question*."[136] When we call something good, we mean that something satisfies or meets a particular system of standards, norms, or ends; and bad what violates them.[137] Mackie calls these systems "*institutions*". So, for Mackie, something has goodness only *in relation* to requirement-generating standards internal to institutions within practical reason. These institutions are reason-generators, and only within these institutions can a reason be valid. They are reasons for one to act only insofar as one desires to be part of the institution, which *entails* motivation to respect its internal requirements. For example, one has reason not to double dribble only insofar one desires to play basketball. Thus, if one desires to adhere to the basketball institution, then two internalist consequences follow. First all reasons are internal because they are ultimately based on the original desire to commit to the institution. Second, normative judgments when uttered from within institutions are necessarily connected to motivation, also derived from the original desire to adhere to morality: "*If the agent is sincere, then his endorsement of the rules of the institution means he is necessarily motivated to respect the rules of the institution and its demands or he is inconsistent"*[138] A lack of motivation in this case would amount to endorsing an institution and yet ask oneself whether one ought to endorse it.[139]

Only in that internal sense can propositions about moral goodness be valid, and thus generate reasons, *viz*. only *within* the institution of morality.[140] Accordingly, moral goodness would be that what satisfies quintessential moral standards within the institution of morality, which Mackie takes to center around exclusively other-regarding concerns. In this sense, a brute fact,

---

[136] Mackie, p. 55 his italics. Institutions are reason generators. Ex: chess, rock-climber, etc.
[137] Mackie and Joyce call "standards", Finlay calls "ends".
[138] Mackie, p. 68-70
[139] Mackie even mentions that this would be something like a re-run of the open question argument.
[140] Mackie, p. 25 "given any sufficiently determinate standards, it will be an objective matter, a matter of truth and falsehood, how well any particular specimen measures up those standards."

such as action ϕ saves the life of *p*, becomes a moral reason for someone once one desires to undertake the institution of morality. The generation of reasons is mediated by the institutions, but the agent must first desire to take part in that institution.

3.5. Problems for Mackie's Institutional Internalism

Sure, uncountable kinds of reasons are institutional in the Mackian sense, such as basketball reasons. Also, perhaps some moral reasons are institutionally generated (in the sense that they apply only if one adheres). However, the Mackian conclusion that all reasons are institutional is implausible. Take the drowning child case. If Jake is passing by and can save the child at insignificant cost, no one would say that if Jake has not adhered to morality, then there is no moral reason for him to save the child. Take a moral reason against killing innocents. We think that a moral reason like that exist (and must at least be considered), independently of one's adherence to morality. Of course, moral requirements can be built within system of rules or have their source somewhere else. But moral requirements may generate reasons for agents regardless of their adherence to the moral system. The same can be said about prudential reasons. There are reasons about my own well-being, if I don't endorse the health life institution. This is a way to say one has an *external* moral reason. I will defend external reasons in section x below.

There is another important sense in which reasons are non-institutional. Sometimes we need to decide whether we have reason to adhere to or to leave institutions. Of course, one first step is to make inter-institutional comparisons. But how can we decide what we have ultimate reason to do when two or more institutional reasons are perfectly tied up? There must be a way of assigning different normative weight to existing rival reasons. Second, even if institution M is comparatively superior to other institutions, we still have the rational ability to question whether we have decisive reason to adhere to institution M. There must be a way of assigning decisive

normative force to institutional reasons, or dimensions of practical reason, if you will.[141] We do that from the perspective of practical reason as such. So, we can ask questions such as:

1. The institution of morality contains requirement r, but does r *really* issue a reason R?

2. If requirement r entails a reason, is it decisive?

When we ask question 1 from within the institution, as a member of it, it will be most usual that r entails R. Eventually, interinstitutional and external facts may raise internal doubts about the requirements of a given system, and so a member may wonder about whether a given requirement does really issue a reason, because, say, the requirement seems outdated[142]. But we can also wonder about whether that requirement does generate a reason either in comparison to other institutional reasons or from the perspective of practical reason as such. Then, even if we confirm that the requirement entails or generates a reason, we can still ask, like in 2, whether that reason is decisive. One may wonder why that is necessary if 1 has already been answered. Imagine that the other institutions available at the time of decision provided incredibly poor reasons for action, perhaps none, and for lack of competition, so to speak, institution K was left with a minimally satisfactory rational advantage over other institutions. Given such a bleak normative landscape, I wonder, do I really have decisive reason to act according to K? Those questions will sometimes require that we try to look at the normative landscape from a non-institutional perspective or institutional detachment.[143]

---

[141] We question morality from the outside. When I ask there is a moral requirement to x, but do I really have reason to x? We question the normative authority of a given requirement not only from within the institution but also from outside the institution. (Joyce, *The Myth of Morality,* p. 45)

[142] This can sound self-undermining since to question whether one has reason to conform to an institution which one has already adhered seems incoherent. But it seems to me that intrainstitutional questioning is legitimate in cases of internal uncertainty. Joyce, p 50.

[143] For example, that can happen when one normative system oppress all others and seize information about open normative outlooks. That is one way to interpret Germans who often committed to the Nazi institution in conditions of very poor information.

Hence, I believe institutional adherence cannot be a constraint on what counts as reasons, at least not in the case of practical reason as such. If all reasons were institutional, then there would be no way to decide whether one has reason (perhaps decisive) to adhere to a given institution. Now, it is natural to ask whether practical reason is not just another institution. Well, if it were, taking its verdicts as reasons would require adherence to it *qua* institution. But practical reason is not optional in that sense, because questioning allegiance to practical reason is incoherent. We may diverge about the rational weight of all the normative systems we know of, but everyone asks external questions such, why should I act morally, what reason is there to conform to a political system, and so on, but we don't question our practical reason that way. It is like asking whether one has reason to do what one has reason to do. And especially, in the case of all-things-considered reasons, one can't decide that one has decisive reason for any given action, and yet wonder whether one has decisive reason for the same action. [144] Trying to provide further reason is pointless.[145]

3.6. Reasons and Alienation

My arguments against MJI implies that there are *external reasons, viz.* reasons which are *not* necessarily connected to motivation or not grounded on desire. I have argued that since moral

[144] There are doubts about the legitimacy of a "free unscripted 'ought'" Tiffany (2007), p. 233. Copp 2009, 2007; Worries involve an infinite regress about the source of normativity and a reductio based on the impossibility of normative comparison. That would happen because the view would be required to postulate ever greater sources normativity conferring institutions/dimensions of practical reason, such as reason-as-such super reason, super super reason, and so on. I think the regress can be avoided. Two options. One is to go naturalist and place the ultimate force on some empirically verifiable source. Hobbesian contractualism can be read like that. Or we can do that on conceptual grounds. Perhaps Mcleod's analogy with the US law is one way to do that. Reason-as-such works like the Supreme Court, whose job, based on its decisive power, is to determine what ultimately ought to be done. Super reason works like the US Constitution, who does not settle conflicts among reasons, but confers that power to the Supreme Court without ever competing with the dimensions of practical reason, e.g., aesthetics, politics, prudence, morality, etc. This strategy is strengthened by arguing that domains of practical reason lack independent normative significance (till totally sanctioned by reason-as-such) and that only reason-as-such is independently normative. I find this view quite attractive given my externalist inclinations and that questioning whether one has ultimate reason to conform to practical reason-as-such is incoherent. Asking for further reason is the same as denying the decisiveness of practical reason as such. That is incoherent, since there is no further court of appeal.

[145] Joyce, pg. 49, Sinnott-Armstrong, 2010

reasons are not necessarily overriding, moral judgments as conceived by MJI proponents are not necessarily motivational. But even if moral reasons are not necessarily overriding, they may still be necessarily connected to motivation. In fact, for *reasons internalism*, the connection to motivation is a pre-requisite for a reason to count as a moral reason, and in fact for *all* practical reasons. The *locus classicus* of the distinction between internal and external reasons, as well as one of the strongest attacks against the latter, is Bernard Williams' Owen Wingrave's case. The argument for moral reasons internalism based on the view that morality is necessarily motivational can be dismembered in two fronts: alienation and explanation.

The view that normative authority is necessarily connected to motivation is best formulated as the following:

Internal Reason: *R* is a reason for *x* to ϕ, only insofar ϕ satisfies desire *d*.

According to *internal reason*, considerations in favor or against any given action[146] are grounded on antecedent desires. All actions are merely instrumental to the satisfaction of the agent's desires. Thus, given that Owen has absolutely no desire to join the army, he has no reason to join the army. This view is not only applied to practical reason, but also to value theories, where value is the desired. In the latter, any ψ is valuable insofar ψ is desired by *x,* under the proper conditions. So, not only practical reason, but, in fact, all value, are subjected to the distinctive authority of desire.

Internalism about reasons and value allows Williams to open two fronts or argumentation against externalism. The first is the explanatory front. All actions are explained as means to the achievement of a given end, i.e., value, where value just is the satisfaction of a desire. In this manner, Owen not only has no reason to join the army, but he also doesn't value joining the army at all. If that is true, then what could be more intolerably alienating then doing something you don't

[146] Dancy, "reasons are facts that count in favour of an action" *Practical Reality* 2004, and "How to Act for a Good Reason" in *Philosophy of Action: An Anthology*, 2015

want, neither value? That's the second flank of attack led by Williams. Externalist views command actions alien to the agent's near and dear values.

Let's tackle the alienation issue first. According to Williams, given Owen's "*subjective* motivational set", he does not have any legitimate reason to join the army, even if there are objective considerations in favor of doing so. Any consideration *not* counting in favor of the satisfaction of Owen's motivational set has no normative weight in the agent's deliberation. So, from the point of view of the agent, it's as though external reasons don't exist. Now, why is the agent's subjective motivational set so normatively significant for this view? The fact that the military life is external to Owen's motivational set *distances* him from his values, his deepest convictions, and his "ground projects". Hence joining the army leads him to a life which he does not want to live. To live a life which one does not want to must be intolerably alienating. This concern about the nature of practical reasons has been expressed by various philosophers:

Williams (in the context of his critic to utilitarianism):

"It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own projects and decision and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and the source of his action in his own convictions. It is to make him into a channel between the input of everyone's projects, including his own, and an output of optimific decision, but this is to neglect the extent to which *his* actions and *his* decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified."[147]

Railton:

"Is it true that all normative judgments must find an internal resonance in those to whom they are applied? While I do not find this thesis convincing as a claim about all species of normative assessment, it does seem to me to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him."[148]

Rosati:

[147] Smart and Williams, *Utilitarianism: For and Against*, p. 116-17
[148] Railton, *Facts and Values*, p. 47

"The principal intuition supporting internalism about a person's good, as aptly expressed by Railton, is that an individual's good must not be something alien – it must be "made for" or "suited to" her. But something can be made or suited to an individual, the thought goes, only if a concern for that thing lies within her motivational capacity: what is good for her must connect with what she would find "in some degree compelling or attractive, at least if [she] were rational and aware." In this way, there must be a "fit" between an individual and her good."[149]

In order. Although Williams is talking about moral reasons and Railton and Rosati refer to a person's good or value, the concern can be extended over practical reason in general. But now it is clear that the concern about alienation has become a constraint on what counts as a genuine reason for action. A consideration that fails to be connected to the agent's life projects, that does not resonate within the agent's interests, or does not fit the agent's motivational capacity is not really a genuine reason for that agent:

Alienation Constraint:
Consideration R provides person P with a reason to ϕ only if R fits, resonates, engages P's subjective motivational set.

The objection concludes that external reasons cannot satisfy AC. I find this objection compelling, because if it is correct, external reasons lead the agent to live a life which he does not value. Hence AC is constraint about what considerations can be genuine practical reasons. But external reasons are considerations that count in favor of an action that are, in principle, independent of the agent's motivations. So, it seems that considerations of this kind are ruled out from the agent's deliberation from the get-go.

3.7. Response

Since AC is seen as a justified constraint, theories which satisfy AC would be better than those which don't. So, the fact that internalism satisfies AC counts in favor of the view. I think it is still an open question whether internalism satisfies AC, because that what the agent is alienated

[149] Rosati, *Internalism and the Good for a Person* 298-99

from is still up for grabs.[150] Notwithstanding, AC cannot be an independent argument for internalism, for the fact that accounts of alienation presuppose an account of reasons for action.[151] Hence AC cannot be a constraint on what counts as practical reason. In this section I explore the second route.

Descriptively, alienation is distance. To be alienated from something is to be distant from something. [152] But obviously it is not any distancing that matters. So, the alienation we are talking about has a normative component, in other words, certain types of alienation matter and others don't. Deciding which ones are normatively significant requires some theoretical tasks.[153] For example, we need to determine what is the agent distancing himself from. Historically, the point of reference has been the agent's *self*. Thus, the alienated agent is in some way distant from his own self. From the start, we can envisage that rationalists and sentimentalists will diverge about what does the concept of self refers to. Yet it is possible to craft internalist conceptions of practical reason in both ways insofar one keeps the nature of reasons for action under the constraint of being necessarily linked to motivation (be them determinations of rationality or desires).

Now, certainly there are aspects of our own selves which we are not proud of. We know we would be better off away from those aspects, and perhaps we desire to, in one way or another, alienate ourselves from them. So, besides a comprehensive conception of the self, determining the normativity of alienation requires pointing out just what aspect of the agent's self-identity is normatively significant. Once that step is taken, we are in position to say that considerations in favor of a particular action ϕ would *not* constitute genuine reasons for action when ϕ-ing drives the agent away from this normatively significant aspect of his self-identity.

---

[150] 2nd order desires, practical identity, etc.
[151] Tiffany, *Alienation and Internal Reasons for Action*
[152] Railton: "estrangement, distancing, or separateness … resulting in some sort of loss" p.93
[153] Tiffany, "alienation problem" the task to determine just what alienation matters.

I think a conceptual distinction may shed some light on the evaluation of AC. We can think of alienation as subjective or objective. If we give the Wingraves' argument a second run for its money, we may see that, in a sense, Owen is alienated from certain *objective facts about himself and the world*. In other words, that Owen is a Wingrave, that male Wingraves have traditionally been military men, that Owen's continuing the tradition would be a source of family honor are all factual considerations that count in favor of Owen's joining the army, regardless of his actual desires. These considerations are all true about Owen, and though he might not like it, they are *reasons* for him to enlist. So, from this point of view, Owen is objectively alienating himself from his own family values, and, hence, part of who he is, part of his own self. Objective alienation is therefore a relationship between a person and some objective condition or state-of-affairs in the world.

Nonetheless, it is worrisome that the military life does not connect in any way with Owen's subjective mental states, *viz*. Owen does not desire to enlist, does not believe in the worth of a military life, nor values it. Unless Owen changes his mind, joining the army will lead him to a miserable life. So, the internalist claims to have an advantage for saying that Owen has no reason to join the army because that would be *subjectively* alienating, while externalists must *a fortiori* accede to there being reasons for Owen, nonetheless. Okay, subjective alienation is more intolerable than objective alienation, so it seems to be the dominant intuition. However, arguments from desires to eat plovers' eggs[154] to counting blades of grass[155] have well shown that mental states possess unequal normative significance. So, the AC needs to be refined to that the constraint is placed under the right sort of mental state.

---

[154] G. A. Cohen, *On the currency of egalitarian justice*
[155] Rawls TJ: 432

Desire theorists circumvent the familiar objection by tying the desires to aspects of personal identity and well-being or to the cognitive procedures the agent went through. So, what aspects of personal identity, well-being, or which cognitive procedures are used to assign normative significance to mental states, desires in this case? So, various philosophers, such as Frankfurt[156] and Lewis[157] have defended that one's personal identity is a sort of cluster of second-order desires, *i.e.,* desires to desire. The normativity of first-order desires depends on whether they are sanctioned by second-order desires. For example, Steve is in the mindset of quit smoking, and so he desires to have the desire to quit smoking. If Steve really is in such mindset, he has the relevant second-order desire. So, when he first-order desires to smoke, his second-order desire will revoke this desire, and (hopefully) even give rise to a first-order desire for whatever leads him away from smoking. So, we have a new AC:

Second-order Desire AC:
Consideration R provides person P with a reason to ϕ only if R fits, resonates, engages P's second-order desires.

For this constraint then, considerations in favor of an action generated by first-order desires that do not align with the agent's second-order desires will alienate him from whom the agent wants to be or identifies with. It is easy though to raise interesting conflicts between first and second order desires which. And now my favorite shot kicks in. Deciding the normative weight of second-order desires presuppose a theory of practical reason. That is the doom of the *Second-order Desire AC.*

Let me introduce Philip. After over decades of military service, Lieutenant Philip has spent his last 5 or so years in retirement. Unfortunately, right when he thought he would enjoy quality

[156] Frankfurt, Freedom of the will and the concept of person.
[157] Lewis, Dispositional Theories of Value

time with his family, his son, a jet pilot died in combat over the Balkans during a classified (and shady) mission whose details have never been disclosed to Philip. Today Philip spends most of his time helping to grow his three, now orphan, grandchildren. As a result, Philip has been in doubt about the rectitude of the military institutions he vowed to honor. Deep down, Philip values the military life and still has a second-order desire to have first-order desires to act according to these values. But today Philip has lost his interest in that kind of life and does not want to deal with that no more. He has systematically rejected invitations to any related events such as Memorial Day and all. Philip even has second-order desires (to desire) to attend to these events, not just because military duties also apply to the reserve, but especially when he gets calls from his veteran friends. But he kindly dismisses them all saying that some memories are too painful for him to revisit and that, even more importantly, he is always too busy caring and providing for his grandchildren. Philip has intense first-order desires to avoid painful memories and to care and provide for his grandchildren to the best of his ability. But he still holds a perhaps strong second-order desire to live up to his military requirements, desires with which he had fully identified for a long time, even now amidst grief for his son.

Next Memorial Day comes around and this year Philip's son is going to be remembered during the event. Even if we concede that first-order desires can also be reason-generators, the view will claim that the reasons generated by second-order desires are normatively weightier. So, Philip has most reason to attend to the event. Should we say that Philip ought to attend to the event regardless of his contrary first-order desires? I think not. However, according to the second-order view, Philip acceding to his first-order desires alienates from him his a very important aspect of his personal identity.

Is Philip's second-order desire a legitimate expression of his "true self"? Given his actual position, that seems up for grabs. We can argue that Philip would rightly regard his second-order desire as expressive of who he really is, even though he also has justified, but contrary first-order desires. After all, he still identifies with military values. Likewise, we can claim that it would be intolerably alienating to demand Philip to betray his actual feelings. Moreover, it is equally possible that at this point Philip has developed a second-order desire to be a dutiful grandfather, even amidst the grief and desolation for having lost his son. Isn't this actual and more intense second-order desire to be a dutiful grandfather more genuinely representative of Philip's personal identity? Philip loves his grandchildren. Besides that, they depend on him now more than ever. So not only he has first-order desires to care and provide for them, but also related second-order desires. He has reason to alienate himself from an institution that has caused him so much pain and remorse. We have come down to a standoff. Which of the two second-order desires (or even first-order desires if you will) are more representative of this actual personal identity? Well, I find intolerably alienating any value system that requires one to repress either option under penalty of irrationality.

I am not interested in deciding which line of action is all-things-considered rationally advisable for Philip. Rather, I want to show that deciding about the normative authority of desires of either order (and which desires are alienating) is derivative on what we believe Philip *has most reason to do*. If one believes that the military institution is often overly demanding and is obliterating Philip's feelings, one will probably be inclined to say that Philip ought to never mind the military life and turn back to his breadwinner duties. If we believe that Philip ought to live up to his military requirements, then we will assign authority to his military-related first and second

order desires. We should think that desires that lead him away from his military duties as alienating (at least military-wise).

So, the resonance intuition cannot be an a priori constraint on what counts as genuine practical reason, because to decide which desire presupposes practical reason, rather than the other way around. Therefore, the alienation constraint cannot work as an independent support for internalism. I think that once we make these distinctions, the alienation constraint does not look so *decisive* in favor of internalism anymore. I am not saying that resonance does not matter. I just think that the constraint has lost most of its normative significance in defining the nature of genuine practical reason.

3.8. Alienation and Practical Identities in Defense of Internal Reasons

The alienation constraint can also be used to argue that reasons are only (ontologically) *internal* to the agent's practical identity. One could argue that whatever is the source of considerations with normative significance, they only form reasons when they fit with one's practical identity. Or one can argue, like Korsgaard, that practical identities themselves are the sources of reasons.[158] She writes:

> "Practical identity is a complex matter and for the average person there will be a jumble of such conceptions. You are 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. And all of these identities give rise to reasons and obligations. Your reasons express your identity, your nature; your obligations spring from what that identity forbids."[159]

I find confusing that she mentions identities which are optional, such as being someone's friend, a member of a profession, etc., and identities that are not optional, such as one's nature (at least physical nature, which certainly is reason-giving). But she clarifies that "*the conception of*

[158] Korsgaard, *Sources,* 1997, 101
[159] Korsgaard, 1997, p. 101

one's identity in question here is not a theoretical one, a view about what as a matter of inescapable scientific fact you are. It's better understood as a description under which you value yourself, a description under which you find your life worth living and your actions to be worth undertaking."[160] So this view seems to do very well in terms of respecting the alienation constraint, since any consideration potentially normative only becomes a reason once it fits a self-personal description valued by the agent. As a result of the view, we can say that the reasons are internal in two ways. They are internal in the sense that they are grounded on an internal feature of the agent.[161] And they are internal in the sense that if one values one's own practical identity, then one will be motivated to act according to its internally generated reasons.

Now if we use practical identity to re-run new alienation constraint, we can set it as follows:

PIC: Consideration R provides person P with a reason to ϕ only if R would be a concern for an agent with P's practical identity[162]

So, potentially normative considerations must meet PIC, or they will alienate from his valued practical identity. However, PIC does not fare better at supporting an independent alienation constraint. I think it is possible that endorsed practical identities generate reasons, but many, if not most, normatively significant considerations are reasons for any one *independently* of one's practical identity. Hence, if reasons exist prior to practical identities, then PIC cannot be an *a priori* constraint on what ultimately counts as reasons.

Our practical identities conception we have of ourselves is forged both subjectively and objectively. We have internal states that are important for how we ought to live, and they contribute

[160] Korsgaard, 1997, p. 101
[161] Tiffany reads Korsgaard as an externalist about reason grounds because practical identities are grounded on the mere fact of being a human being, which would commit human agents to respecting the humanity of others. (Tiffany 2003, p. 400) But not only does he make explicit his internalist use of practical identity, also this interpretation is congenial to her own "test" for practical identities with normative powers.
[162] Tiffany

to one's formation of character. But our practical identities are often formed by what we have most reason to do. External facts to one's practical identity can constitute reasons for one to change one's practical identity, instead of alienating him from it. Let us remember of Philip. The military life is one of his endorsed practical identities. But now imagine that Philip did not, at least in principle, give up that life and assume a new identity because only now he endorses and values the practical identity of being a grandfather. An important fact, namely, his son died on a war mission in shady conditions, was a strong reason for his deciding to abandon the military life. That reason is independent of his endorsing either practical identity. Had this fact not taken place, it would be hard to explain why suddenly Philip decided to abandon the military. His existential change tracks the reason, instead of creating it. These identities may add or withdraw normative strength[163] to the reason, but the fact will work as a reason for either identity, independent of Philip's prior commitment to the military life. Therefore, that is a reason that applied to Philip regardless of endorsed, and still valued, military identity.

That said, I don't see how one could argue that losing a beloved son is an alienating reason. We could suppose that Philip tries to think about reasons to attend to the event. He will consider facts about the military life, such as his oath to the flag, patriotism, tradition, hierarchy, etc. But all military reasons seem to alienate him from his parental identity. At the same time, facts about the well-being of his family constitute reasons for him to alienate himself from the military. Now, whether one finds the former or the latter identity more normative significant will derive from our intuitions about what Philip has most reason to do. If the practical identities were tied, so to speak, since he endorses both and values both, then, only reasons existing independent of his identity will help decide one way of the other. In fact, appealing to external reasons seems to be the only way

[163] Dorsey, *Normative Significance of Self*. 2016

out of the "Cohen's mafioso problem" [164] in which Korsgaard gets entangled. To convince the mafioso, we would cite facts about respect for persons, justice, human flourishing, and so on. In other words, people can assume practical identities which they have reason to abandon. In turn, these reasons exist and are available to the mafioso prior to his deciding to endorse a new identity. Would these reasons be genuinely alienating as internalists want it to be? I don't think anyone would think so. In the case of Owen Wingrave, for instance, he may be alienating himself from the Wingraves or from the other identities he would assume depending on available reasons. He has does not merely desire to be pacifist, he believes that there are reasons for assuming that standpoint and not the other way around. Likewise, were there a risk of his own country being devasted, this fact would certainly be a reason for Own to ultimately join the army. Since reasons exist independently of practical identities and in fact can lead to one's changing his self-conscious identity, I conclude that PIC cannot stand as an independent constraint on what counts as reasons.

3.9. External Reasons and The Explanatory Constraint

Practical judgments are beliefs about there being normative reasons. Reasons are facts that count in favor of actions. Reason statements carry propositions whose truth can be verified. So, for example, we can verify that action Φ would promote utility, that action Φ respects autonomy, or advances aesthetic, political values, and so forth. But if practical judgments are *beliefs* about there being reasons, then it is possible this justification does not necessarily imply, entail, elicit or form a motivation/desire/intention to act.[165] According to the traditional belief-desire system[166],

---

[164] *Sources, p. 254*

[165] I'm using this variety of terms to encompass those doubtful about the Humean belief-desire system such as Nagel, Korsgaard, and Shafer-Landau and hence often use either the over general "motivation" or explicitly choses in favor of "intention".

[166] Anscombe, Smith

the belief is not a sufficient condition for the action. So, the externalist needs to explain how a practical judgment gives rise to action, besides justifying it.[167] The challenge for the externalist is also presented as reductio. When the externalist dissociates normative reasons from motivation, there will explanatory reasons which have no normativity and normative reasons without explanatory power.[168] This externalist burden can be put in the form of a constraint about what counts as genuine reasons for action:

> ***Explanatory Constraint***:
>
> Normative (Justificatory) reasons must identify with or be able to become explanatory (motivational) reasons.[169]

Now, we can extract the argument employing EC in favor of internalism:

1. A practical reason (in the sense of justificatory reason) must be capable of explaining action that was performed for that reason (EC)
2. Actions can only be explained by motivationally relevant attitudes, i.e., desires.
3. All practical reasons must be internal.

Williams to go back and forth between the two readings of premise 2, in justificatory and explanatory terms. Practical judgments are motivational and produce action, exactly because reasons are grounded in desires. But I believe that premise 2 should be read merely in explanatory terms, as an instance of the believe-desire system. In that manner, justificatory reasons remain being factual considerations, independent of desires, that count in favor of actions. For example, the fact that a comedy show is fun is a reason for anyone to see one, regardless of

---

[167] "if there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action." Williams, *Internal and External Reasons,* p. 18; Korsgaard, *Sources*, p. 251-2; Nagel, *The Possibility of Altruism*, pg. 34-5

[168] Smith, 1994, Brink *Significance of Desire*

[169] Smith, 1994; Ulrike Heuer, *Reasons for Actions and Desires*, 2004. Brink, *Significance of Desire*, Dancy, J. *How To Act For a Good Reason*, 2000.

whether one desires to see one. But, to show how justificatory reasons can become explanatory reasons, I propose to understand practical judgments in terms of beliefs about there being reasons. I judge that I *have* a reason to respect autonomy when I believe that there is a reason to respect autonomy.

When I believe there is a reason, that belief may contingently attract motivation. But in my view, this motivation is contingent, *inter alia*, on the reason passing from a *pro tanto* reason to an all-things-considered judgment, i.e. a judgment of plain practical reason as such.

A qualification about the content of reasons. There are two kinds of reasons in place. Take the reason to attend to the Nutcracker next Friday at Lied Center. This reason is derivative from there being a *prudential reason* to experience aesthetic value. Accordingly, there are two beliefs in operation, a belief about there being a reason to attend to the Nutcracker and a belief there being a prudential reason to experience aesthetic value. Let me put it in terms of moral reasons. Take the reason to help the homeless on Mass st. Then assume any traditionally solid moral theory to claim that there is a reason to help the homeless on Mass st, say, respect autonomy as an end in itself. Hence, the value of autonomy generates reasons and when the agent believes in there being a reason to respect autonomy, he will also believe in there being a reason to help the homeless on Mass st. So, the reasons are the things believed, rather than the mental state of the agent. But they will motivate once they are *believed* by the agent.[170]

Now to complete the explanatory story, I just add the proposal that desires are (contingently) responsive[171] to beliefs about there being reasons. To be rendered intelligible, desires need to be about objects of relevant rational interest. What could be said in terms of well-being about the act of counting blades of grass? Not much, everyone agrees. But there is a lot to

[170] Some inspiration from Ralph Wedgwood, *Intrinsic Value and Reasons for Action*, 2009
[171] Heuer, *The Identification Thesis*, Brink *Significance of Desire*.

be said about the wellbeing of the homeless on Mass st., autonomy, equality, and the Nutcracker. It is the factual characteristics of these things that explain why we desire them. Thus, the desires themselves need reasons to be rendered intelligible. In this manner, we have the ability to deliberate about what desires we shall sustain and pursue. So, deliberation about what reasons there are and their comparative weight to practical reason can give rise to new motivation. The explanatory constraint does not require that justificatory reasons and explanatory reasons be one and the same. Explanatory reasons may be desires (plus the belief that Φ-ing satisfy the desire). The normative reasons they must only be *potentially explanatory*.

Aren't you talking about dispositions then? one could ask. I suspect that dispositional accounts are question-begging because they explain the origin of desires in terms of the disposition to have that desire. To say that there is a disposition is to say that there is a pattern. But the appealing to a disposition does not explain why there is a pattern. It just states that there is one. (Their hope is to confirm it physiologically). Why do we have desire d? Because usually when we get acquainted with x, we tend to desire it. So, inductively, we have a disposition to desire x. What explains the disposition D? We have disposition D, because we often have desire d. [172] One way to stop the cut the circularity is to explain how the disposition is acquired. However, there are two problems for this attempt. First, many dispositions may be acquired in quite wrong ways, and second, quite wrong dispositions may be acquired.[173] Deciding which dispositions are normatively relevant will take us back to external reasons.

The desire only explains the action, but does not justify it. The desire follows from the belief that on there being a reason to Φ-ing. The agent believes in there being a reason to Φ, because Φ is in accordance with the requirements generated by value V. Say I believe in the value of

[172] Heuer, p. 56
[173] Kant and Mill extensively insist on their own versions of "moral training".

autonomy. Respecting and advancing autonomy generates requirements. That there is an autonomy-requirement not to Φ is a reason for me not Φ. Take trespassing. My belief in the value of autonomy gives me reason to respect and advance autonomy. Autonomy-generated requirements prohibit trespassing. Hence there is a reason not to trespass. Then, if we endorse the belief-desire system, we can say that in principle I would have a *de dicto* desire to act according to autonomy reasons, and a *de re* desire to avoid trespassing. Now, these desires are all contingent. Sure, Kantian or Millian training can make these desires more habitual. But the only motivation that seems more binding is the desire to act according to all-things-considered reasons. It seems to me that if the agent actually engages in rational deliberation, then he is willing to act according to the resulting reason for action.

Furthermore, all this account is assuming that desires in functional terms and that they are the only thing that explains. And often the concept seems to be treated as some individuated physiological state/process that physically propel actions. But questions about the phenomenology of desire left open: functional and physical states are the same? Very difficult to prove. How could it be that desires have always exactly the same neurological makeup in every agent to whom we attribute a desire? Is it a desire to see the Jayhawks that draw people to the game? Perhaps. But it is impossible to say that everyone's brain states are equally the same. It is more natural to infer that many different brain states give rise to desire, e.g. the eagerness for a felt quality, but also fear, love, beliefs, etc. So it is hard to find one single brain state matching the functional concept of desire. If all that the internalists were worried about was naming this brain state that propels action, then externalism has better prospects to guarantee resonance and explanation.

3.10. **Objection: But isn't valuing ultimately grounded in desire?**

The way to show how external reasons can become explanatory is by arguing that the reason is internalized via belief in there being a reason. As I said, the reasons reflect requirements generated by value V. But, if value V is grounded on desire (or any other non-cognitive state), then the reasons generated by the requirements are ultimately grounded on desire, and hence, *internal*. The response is to adopt a cognitive theory of what is to value something. In this manner, the reasons remain connected only to beliefs. Hence valuing V is good for *x* to the extent that *x believes* that V is good.[174]

The literature in value theory is extravagantly prodigal in its attacks to the desire-based theory of value.[175] Moreover, we have seen that desires themselves need reasons to be rendered intelligible. One needs to believe in the qualities of value V in order to desire V. Notwithstanding, people can have false beliefs about V. So, not all beliefs about what is good correctly tracks what is *actually* good.[176] But these beliefs can be achieved by further axiological investigation, based on coherence among beliefs of various kinds, strength of beliefs, and their comparative influence in people's lives. For instance, the value of autonomy is much stronger and more central to anyone's life than the aesthetic value of listening to Kind of Blue. More importantly, a belief-based theory of valuing has deals better with cases where people desire, but do not value, and cases where people value, but do not desire.

Let us think about how things could have gone different for Owen. Owen has no reason to enlist, not because he does not desire it, but because he does not *believe* in the values that justify

---

[174] Influenced by Dorsey, *Subjectivism without desire*

[175] Nagel, Korsgaard, Cohen, Anderson, Brink, Shafer-Landau, to name a few

[176] Which has intrinsic features, *viz.* formal and substantial qualities. The ultimate metaphysical shape of the good for the purposes of this section, I assume to be determined by a theory of supervenience. Shafer-Landau, *Moral Realism,* p. 72-79.

going to war. Of course, he does not desire it either. Now suppose Owen loves his family, and he does not want to disappoint them. So, he has struggled to develop a desire to desire to enlist. Ultimately is overcome by the desire to desire to enlist and goes ahead with that. Does Owen *believe* in the values that justify the war? Certainly not. His value judgment in operation is a belief about the value of familial ties. And even under Williams' account, suppose that Owen found a subjective element in S that makes him better see the reason to enlist. He loves his family and want them happy. If they are happy, he is happy. So, he now has a reason to enlist because enlisting is grounded on a subjective element of S. Yet, does Owen now value military life? Of course not, because he still lacks the belief in the values that justify wars. So now we have a reason to Φ grounded in an element of S which leads him to an action that he does not value. Therefore, Owen desires to desire enlisting, but does not value it.

Also, we value things we don't desire. Let us recall Phillip. After losing his son as war casualty during a suspicious mission, he has no desires for military life anymore. The prospect of going to the military event causes him frustration and angst. He lacks any desire whatsoever to be around the military. Thus, many internalist would stand to claim that Phillip does not value the military life anymore. But wait a minute. Phillip is an awarded military officer. He lived the military all his life and for all this time believed in the values that justify (just) war. Beliefs like that, that make one go to a war are not beliefs that vanish from night to day, even at the loss of a beloved son. I believe it remains plausible to say that Phillip still values the military life. A desire-based view is too quick to declare Phillip does not value the military anymore. Perhaps, after long time passes and his mourning sits, Phillip may feel like returning to his military duties little by little, as emotions allow. Therefore, Phillip values the military, does not desire it.

3.11. **Chapter Conclusion**

I hope to have shown that the two most important constraints used to defend internalism about reasons have no independent support against externalism. In the following chapter, I am going to introduce synthetic moral realism.[177] as the most viable semantic response to the OQA. There are some different strains within that program, so I also will sharpen which view exactly I am holding. Next, I introduce the Moral Twin Earth problem, the deadliest semantic challenge for synthetic moral realism.

[177] Boyd, Brink, Railton, and Copp

Chapter 4 Semantic Objections to Moral Realism: Open Question and Moral Twin Earth

I open this chapter introducing *synthetic naturalist moral realism* (moral realism henceforth) as a response to the OQA. Then, we shall see that the realist project falls victim to a contemporary version of the OQA, the *moral twin earth argument* (MT). With those cards on the table, I proceed to discuss the internalist and externalist semantic aspects of the problem before we enter the contemporary state-of-affairs of the debate. Finally, I develop a theory for mitigating the problems raised by MT.

4.1. The Open Question Argument: The Dead Still Walks

The importance of Moore's OQA is that given the conservativeness of logic[178], conceptual identity matters for inferences, either deductive or inductive.[179] In order for arguments using moral judgments such as that "Nazism is bad" as premises, the judgment itself must represent a true identification between the properties of Nazism and the badness (moral badness in this case). If the judgment does not represent an actual relation, it is false and obviously cannot be employed as premise. Moore attacks metaethical naturalism by claiming that it commits what he calls the *naturalistic fallacy* based on his *open question argument* (OQA). This argument is based on the supposed absence of synonymic relations between *moral* terms and *naturalist* language The conclusion of OQA is that moral entities are neither reducible to, nor derivable from non-moral entities, that is, those found or posited by scientific or metaphysical enquiry. In short, 'good' cannot be defined.[180] There are however some different interpretations of what were Moore's philosophical virtues or vices that led him to such position. Let us first see how Moore arrived at his exotic conclusion.

Moore claims that 'good', when used to refer to a quality of an object, including abstract ones, is a simple notion.[181] It is simple because it denotes something that does not have different properties and qualities such as the thing denoted by the complex notion of 'horse'. In this kind of

[178] Logic does not allow that the conclusion of an argument states something of another category than the one to which the premises of the argument belong. Therefore, one cannot extract moral conclusions from a set of purely empirical premises. Similar applications of the conservativeness of logic in philosophy prohibiting inferences from premises of one kind to conclusions of another kind are that one can't get a universal sentence from particular ones and that one can't get a sentence about the future from sentences about the past or present. Hence, at a minimum, moral conclusions cannot be deductions, because in that case ought-relations would be derived from merely is-relations, and introduction or new relations or affirmations are logically debatable, to say the least. For more on the conservativeness of logic cf. Pigden, C. (ed.) *Hume on Is and Ought.* Palgrave Macmillan, 2010

[179] Frankena 466-467

[180] PE, p. 58/7

[181] PE, P. 59/7.

analysis, definitions follow a hierarchy in which more complex notions are defined in terms of more primitive ones.[182] But for Moore, 'good' refers to something that is not decomposable into smaller parts, whose names would in turn be amenable to similar decomposition.[183] That something is a quality which applies to any given whole that may instantiate it, like in "this typewriter is good", and in that manner, they can truly apply to it or not. Either this typewriter is good, or it is not, and, in this case, it is easy to grasp the significance of a question about whether an adjective does or does not apply to a substantive. Moore clarifies that he is not talking about *things* that have goodness, but about *goodness*.[184] In this case, the thing to which the term 'good' refers, namely, goodness, has no parts, and so, if definitions of wholes are simply enumerations of their necessary, sufficient, and ultimately indivisible parts, then 'good' is indefinable. Therefore, every attempt at defining good mistakes a quality for the whole to which it applies, *viz*, it confuses goodness with things that have goodness. Moore calls this misidentification the *naturalistic fallacy*.[185]

The open question argument enters the scene to show that goodness is indivisible. Goodness, as a quality, may apply to pleasure, the desired, human flourishing, rationality, personal character, etc., but these wholes do not reduce to this single quality, nor is the quality the whole. 'Good' applies to these things insofar they have, *inter alia,* the quality of goodness, but they obviously are not goodness itself. For example, rationality may be, among other things, good, but it is not goodness itself. It probably is mix of mental abilities and neurological conditions. If we identify 'good' with 'rationality', reducing the former to the latter, we get "rationality is

[182] "A definition states what are the parts which invariably compose a certain whole." 61
[183] PE, p. 60/8 and "Good" has no parts. 61
[184] PE, p. 61/10 "'Good', then, if we mean by it that quality which we assert to belong to a thing, when we say that the thing is good, is incapable of any definition, in the most important sense of that word."
[185] It is "naturalistic" because Moore's first target were identifications of goodness with natural objects. But it can apply to abstract ideas as well, a topic he discusses in chapters 3 and 4 of PE.

rationality" as a result. Identifications like these reduce the whole that contains a quality to the quality.[186]

Moore suggests a test by which the confusion can be grasped. Whenever a definition of 'good' is offered, "it may be always asked with significance" whether the thing that has goodness does really *possess* such quality.[187] However, in the Moorean world, every definition of 'good' confuses the quality of goodness with the whole that instantiates it via identification using the verb 'is' form of the verb to be, like F is G. But it is crucial that in many occasions the verb 'to possess' is much more representative of Professor Moore's profound insight. Accordingly, since F also has qualities H, I, J…, "it may be always asked, with significance", whether F does or does not *possess* G, the truth of which would allow us to say F is G. The question whether F is G *qua "*does F instantiate G?" is intelligible, whose answer may be genuinely informative, exactly because it is an open question whether F *possess* G or not. When F does possess G, we may say F is G, which is fine, but Moore's concern emerges when we say that G *is* F, thus reducing F to G. If the reduction succeeds, then we are not saying F *possess* G, but that F *is* G, and nothing else. If that's the case, then we can replace "F is G" by either "F is F" or "G is G". But the question of whether F is F or G is G is obviously closed. But we all think that the question whether F *possess* G or not is meaningful, given that it is possible that F, though having H, I, J…, may not have G. Therefore, it is an open question whether F possesses G or not.

Using the example of rationality again, when we say, 'rationality is good', we are saying rationality *possess* goodness. But rationality also has other qualities H, I, J…, and so if by saying 'rationality is good', we are not meaning that rationality *has* goodness among other qualities, but just that 'rationality' and 'good' are synonyms denoting one and the same entity, whatever it is,

[186] PE, p. 65/12
[187] PE, p. 67/13

then it would indeed be a *closed question* whether rationality is good or not, since that question could be replaced by 'Is rationality rationality?' or 'Is rationality rational?'.[188] However, asking whether rationality is good or not remains a meaningful question, since saying that rationality can be evil remains a meaningful possibility.[189]

4.2. Historical Criticisms

4.2.1. Problems with Moore's Notion of Analysis

The first wave of critic, captained by Frankena, attacks Moore's (Butlerian)[190] notion of analysis and its theoretical consequences. Frankena raises three arguments:

Argument 1

Frankena claims that the naturalistic fallacy commits the mistake of treating two properties as one.[191] But how does Moore's view multiply extensions? According to Frankena, Langford[192], and Lewy[193], Moore's notion of analysis implies that, for instance, 'brother' and 'male sibling' denote different extensions, because they are *intensionally* different. In that sense 'John is a

---

[188] "But whoever will attentively consider with himself what is actually before his mind when he asks the question 'Is pleasure (or whatever it may be) after all good?' can easily satisfy himself that he is not merely wondering whether pleasure is pleasant. And if he will try this experiment with each suggested definition in succession, he may become expert enough to recognize that in every case he has before his mind a unique object, with regard to the connection of which with any other object, a distinct question may be asked. Everyone does in fact understand the question 'Is this good?' When he thinks of it, his state of mind is different from what it would be were he asked 'Is this pleasant, desired or approved?' It has a distinct meaning for him, even though he may not recognize in what respect it is distinct." PE, p. 68/10

[189] Perhaps an example of an almost similar consideration, if it was not for what would come next about the will: "Understanding, wit, judgment, and the like, whatever such talents of mind may be called, or courage, resolution, and perseverance in one's plans as qualities of temperament, are undoubtedly good and desirable for many purposes, but they can also be extremely evil and harmful..." Kant, Immanuel. *Groundwork*. Revised Edition. Cambridge University Press, 2015, 4:393. p. 7. It may be worth noting that metaphysical conceptions of goodness, such as Kant's, equally fall victim of the OQA, see PE, Chapter IV.

[190] *Viz.* PE's epigraph: "Everything is what it is, and not another thing." In 58/6, "If I am asked 'What is good?' my answer is that good is good, and that is the end of the matter." And preface to 2nd ed. p.6

[191] Frankena, W.K. The Naturalistic Fallacy. *Mind*. v.48 n. 192. 1939, p.471.

[192] Langford, C.H. The Notion of Analysis in Moore's Philosophy. In Schilpp, P.A. *The Philosophy of G.E. Moore*, Northwestern University Press, 1942.

[193] Lewy, Casimir. G.E. Moore on the Naturalistic Fallacy. *Proceedings of the British Academy*. V. 50, 1964, pp.251-262

brother' is not intensionally identical to 'John is a male sibling'. In the moral case, 'good' is intensionally different than everything, but 'good' itself. They argue that making conceptual intension as the sole criterion of synonymy makes the OQA invalid, since that would render all synonymies impossible or irrelevant, called the paradox of analysis.[194]

But Lewy thinks since it is possible to logically doubt whether F is G in a way that we cannot about F is F, he concludes that if read merely from the intensional point of view, the OQA is valid. But his view, like Frankena's, is to include extensionality as a criterion of synonymy, so that the *truth-value* of the propositions would not be altered if we replaced expressions denoting the *same extension*.[195] Similarly, Frankena claims that historically those who have defined moral concepts in natural terms would not be committing that mistake, if all they argue is that two words mean one and the same extension.[196] In this manner, we would not be identifying two different extensions, but *discovering* that two linguistic items refer to the same extension. And Langford argues that we identify F with G via the conscious habit of recognizing that they usually denote the same extension.

Argument 2

The second argument is that Moore's Butlerism would rule out all definitions of any term whatever, since the criterion of intensionality prevents that any concept be analyzed in terms from any other region of language, and so *ipso facto* 'good' cannot be expressed by any words or phrases from non-moral vocabulary.[197] Indeed, Moore considers that by 'Good is good, and that is the end of the matter' he meant that 'good' is "different from everything that we express by any *word or*

[194] The paradox of analysis is exposed in detail by Langford p. 323. Lewy, p. 260-1, Lewy tells that in a number of discussions, it transpired that Moore insisted on intensional identity as the ultimate grounds for identity *per se.*
[195] Lewy, p. 260
[196] Frankena, p. 472
[197] Frankena, p. 472

phrase other than the word 'good'"[198] He believes that if this thesis is true, it would follow that 'good' is unanalyzable, since any analysis would use other linguistic items. However, after admitting that 'good' is colloquially expressed by various other words, he maintains that 'good' is unanalyzable, if those words and phrases *contain an analysis*. In other words, if those propositions are meant to be genuine analytical propositions, and not to express their function in ordinary language. However, as Moore points out, 'good' may still be identical with some natural property, since natural properties may be unanalyzable.[199] At this point Moore refines the view to the thesis that the analysis of 'good' cannot be given in *natural* terms, leaving the possibility of an analysis in other terms.[200] But what is Moore's argument for the difference between moral terms and natural terms? That is where Frankena's third argument kicks in.

Argument 3

The most ardent of Frankena's arguments is that Moore *assumes* that moral properties are different in kind from non-moral properties, and from there argues circularly that moral properties are not definable in non-moral terms.[201] In the case of morality, arguing that such identifications cannot be found because moral terms are exclusively moral begs the question, since the question at stake is showing why moral terms cannot be defined in non-moral terms.[202] Now, Moore needs to argue for the indefinability of goodness in *natural* terms before he concludes that all tentative

[198] PE preface 2nd ed. p. 6-9
[199] Preface p.14 and P. 254 in Lewy, Casimir. G.E. Moore on the Naturalistic Fallacy. *Proceedings of the British Academy*. V. 50, 1964, pp.251-262
[200] This is represented by Moore saying in the *Preface* that he ultimately holds to his proposition 3. 'Good' is not identical with any natural or metaphysical property. p. 16-17
[201] Frankena, p. 467
[202] Frankena 473

definitions are flawed. So how exactly is the question begged? Let us look at Moore's four attempts to distinguish between natural properties (parts) and intrinsic value (properties).[203]

Moore first characterizes natural properties and objects in epistemological terms, existing as "objects of experience"[204] that we can "touch and see and feel"[205], and which can be known by "empirical observation and induction"[206]. We can settle empirically what things possess the property of goodness. But what goodness itself is, cannot be settled empirically. Two responses: reasoning about what goodness is seems to include material propositions and implications, such as goodness in virtue ethics and consequentialism. Second, there are similarities between value enquiry and scientific enquiry, such as explanation, prediction, and mitigation of vagueness.[207] Moore seems to be implicitly assuming the absence of any extension[208] of the property of goodness, while still attached to his intensional criterion of identity of the OQA in the first chapter, and so taking as clear that questions about goodness cannot be settled empirically.

The second characterization is metaphysical; natural properties of an object are "rather *parts* of which the object is made up than mere predicates which attach to it", and thus *exist in time by themselves*.[209] Goodness, though a property of the object, is not part of it, and so by contrast does not exist in time.[210] Moore dismisses this characterization after agreeing with Broad that at least some natural parts cannot be thought to exist by themselves in time.[211] But this point was

---

[203] Moore, The Conception of intrinsic Value, pp. 280-208; Preface, pp. 12, 13, 22, 23. And Sturgeon, Nicholas. Moore on Ethical Naturalism. *Ethics*. Vol. 113, No. 3, Centenary Symposium on G.E. Moore's Principia Ethica. April 2003, pp. 528-556.
[204] PE, s25/p90
[205] PE, s66/p161
[206] PE, s25/p91
[207] Sturgeon, p. 545-6 and Sturgeon *Moral Explanations* and Sayre-McCord.
[208] Moore says "evidence" in the first preface ix and some other times,
[209] Including our minds, thoughts, and feelings 26/92
[210] Moore, 26/92-3 The parts are such that "if they are all taken away, no object would be left, not even a bare substance: for they are in themselves substantial and give to the object all the substance that is has. But this is not so with good."
[211] C. D. Broad, Schilpp, pp. 59, Moore, Reply, p. 581-582

instrumental for introducing Moore's distinction and admitted confusion between parts and properties of natural objects.[212]

The third characterization also uses the distinction between parts and properties of natural objects. Moore defends that goodness, or intrinsic value, is *internal* to the objects which possess the property of goodness.[213] Internality is not mere intersubjectivity[214], i.e. objectivity among thinking minds, but objectivity in a stronger sense; from the point of view of the universe. Intrinsic value (property) depends solely on the internal parts[215] of the things that possess it, yet it is not itself an internal part.[216] The reason why Moore insists that intrinsic value is not internal part nor a natural one is his confusion between "intrinsic" and "internal". His reasoning is that according to common sense, if something is intrinsic, then something is internal. What is internal to objects is natural. Good is not natural. Therefore, intrinsic value is not internal.[217] The third premise is question-begging. And when Moore entertains twice the question what is it that characteristic of intrinsic value that prevents it from being an internal part, he answers "I can't see *what* it is"[218] and "I confess I cannot say" but only can "vaguely express" that by enumerating all the internal parts of an object, we give a complete description of it, and not mentioning the intrinsic value it possesses does not make the description less complete as it would, were it missing an internal part.[219] So, intrinsic value occupies an intermediary conceptual space, goodness is never internal to an object, and yet always dependent on the internal parts of its possessor.

---

[212] Moore preface 2nd p. 13
[213] Moore, The Conception of Intrinsic Value, p. 282
[214] Moore calls the non-subjectivist views he is criticizing "objectivist", but it seems to me that "intersubjectivity" would be more accurate.
[215] Moore says "intrinsic nature" and "intrinsic properties", but per his own explanation, it seems more accurate to say "internal parts"
[216] Moore, Preface 2nd p. 22; CIV, p. 282, 286, 295,296
[217] 297
[218] It's MJI, I will go back to this starting with Frankena, Rosati, and then Mackie proving that it is MJI.
[219] 297

This characterization invites for contextualism. Since an objects' description is complete without mentioning the eventual intrinsic value that attaches to it, it is possible that in some contexts it attaches and in others it does not. But Moore does not welcome any contextualism, since he does not think goodness is relational, even less so relative to circumstantial factors, because that is a form of subjectivism.[220] And as he has previously argued, subjectivism conflicts with strong objectivist intuitions, rendering the simple forms of the view unattractive.

The last characterization is disciplinary, according to which natural properties are the object of the natural sciences and psychology.[221] First, the distinction carries an overphysicalist view of natural sciences, whose business would be to reduce concepts to their physical extensions. However, the natural sciences work with several not completely reducible concepts, such as the notion of health[222]. But even if that view was supported, not saying why moral concepts are not reducible to natural terms is again question-begging.[223]

These criticisms show that Moore believes, without sufficient argument, that there is something special about moral properties that fall outside the purview of natural sciences. Moore is defending that moral properties are indefinable in natural terms, before demonstrating what exactly about moral properties prevents the translation. He admits he cannot see or say what that is. Frankena says that is an "unfounded assumption", a "prejudice", and a "moral hallucination" and to assume that against those who have tried such translations is "both unamiable and profitless".[224] Earlier in this dissertation I have mentioned that according to Hurka, there is no

[220] PE section 19, see organic units
[221] Preface, p. 13, PE section 26, p.92
[222] PE, p. 43/S27 he assumes an eliminativist stance about notions like "health". He thinks health is after all an ethical term, and the natural sciences illicitly borrow the term. Yet, it seems that we can settle empirically whether someone is healthy or not. So, later in the CIV, he says that health possesses goodness, but does not reduce to it. So the natural sciences can investigate all the other internal parts of health, expect for its goodness.
[223] Why is it natural? Because it is object of natural sciences. Why is it object of natural sciences? Because it is natural.
[224] Frankena, p. 475

textual evidence that Moore's prejudice was to quite implicitly assume that morality is *necessarily* motivational.[225] Hence, I have tried to avoid *this* prejudice about Moore. This point is instrumental for specifying this dissertation's core thesis: MJI cannot be a deciding factor for the identity or the lack thereof between moral and natural terms.

4.2.2. There Is No Goodness

Geach argues that "there is no one description, 'natural' or 'non-natural', to which all good things answer. The traits for which a thing is called 'good' are different according to the kind of thing in question: a knife is called 'good' if it is UVW, z stomach if it is XYZ, and so on.[226] 'Good' does not stand for any property, or set of properties, rather it determines different properties depending on the object it applies. "Even when 'good' or 'bad' stands by itself as a predicate, and is thus grammatically predicative, some substance understood; there is no such thing as being just good or bad, there is only being a good or bad so-and-so."[227] The corollary of this view is that any attempt of defining goodness by itself is vain, since moral terms are never legitimately used in the Moorean, predicative, sense.[228]

'Good' is used in the attributive sense in 'a good knife' and in the predicative in 'this knife is good'.[229] An adjective is predicative if a phrase 'is an Fg' ('F' = adjective; 'g'= noun), like 'is a red car" can be logically split up into 'is a car' and 'is red', otherwise 'F' is attributive. The particularity of attributive adjectives is that they alienate the sense of the noun they modify. So, if F is attributive, then either an Fg is not a g, or it does not follow that from 'x is Fg', 'x is g' and 'x is F'. Examples: artificial leather, apparent heart attack, former wife, decoy duck, forged banknote,

[225] Geach touches this point and disagrees with Moore: "this influence (motivation) on action is not the logically primary force of the word 'good'" p. 37
[226] 37
[227] Geach, p. 34
[228] Geach p. 35. And that would make the OQA a pseudo-problem.
[229] Geach, Peter. Good and Evil. *Analysis* 17/2: 33-41, 1956

putative father, negative growth. We cannot validly move from 'x is artificial leather' to 'x is leather' in the same way we can from 'x is artificial insemination' to 'x is insemination', since in the latter x remains being insemination, artificial or not, while in the first artificial leather will never be leather.[230]

Geach's view has problems. First, Geach uses this distinction *per se* as an argument not to respond, but to dismiss Moore's OQA as a pseudo-problem. But we need an argument showing that moral terms are used solely in the attributive sense, because simply *stating* the distinction is question begging. Geach disdains the predicative good as a "peculiar philosophical use of words" whose meaning and use cannot be known.[231] Attributive use aside, moral terms are used in the predicative sense all the time. Anyone can say "friendship is good", "starvation is bad" and so forth. The use of predicative good being legitimate or not is different question. But the ordinary use cannot be denied. So, Geach offers two arguments for the illegitimate use of predicative in ordinary language.

He argues that for predicative adjectives like 'red', we can "pool independent information" that something is red and is a car to derive the proposition 'this car is red'. The same does not apply to moral terms. We cannot combine the independent knowledge that something is good and that it is a car to conclude that this car is good. In this manner, knowledge of 'good' and 'bad' is always *dependent* on our first determining what is the object of qualification. Therefore, goodness does not exist independently of the object which has it. But at this point of his argumentation, the only reason Geach seems to have for thinking that there is no independent knowledge of moral terms is that he believes in the OQA: "Such theories of goodness are open to well-known

[230] Geach's example: From 'x is a forged banknote" we cannot extract 'x is a banknote'. p. 33
[231] P. 36

objections: they are cases of the Naturalistic Fallacy."[232] On the other hand, he also thinks that the naturalistic fallacy is pseudo-problem, since any use of predicative good is attributive good in disguise. Two things can be said here. First, to say that something is good or bad requires that we know at least a sufficient number of the properties of the object under qualification. So, moral qualification seems to be at least in part dependent on the thing to be qualified or we could not be predicating at all. Second, Geach does not yet have an argument for why there is no independent knowledge of moral terms, except for accepting an argument that for him is pseudo-problem.

Geach's last argument is that the use of moral terms in both predicative and attributive senses would generate grammatical anomalies such as 'a bad good g' or 'this good g is bad'. That happens if we add 'good', in the attributive sense, to the name of a sortal, and then add 'bad' in the predicative sense. For example, we can agree that the AK 47 is one of the best assault rifles ever invented, and so 'g is a good assault rifle'. But we could also agree that weapons, no matter how good, are bad. But saying that something is good in the attributive sense does not prevent us from saying the same thing is bad in a predicative sense. So, that generates anomalies such as 'this good assault rifle is bad'. Three points in response. First, the intelligibility of the claim that AK-47s are, however good, still bad, proves that moral terms can be used as predicative too. Second, Geach has not *shown* that moral terms alone are not definable. In fact, the anomaly can be avoided by replacing one of the moral terms by their relevant definitions, which could be supplied even by the context, for lack of a more reflected one. Lastly, the use of 'bad' in this case does not function as an *alienans*, since from 'bad AK-47" we can extract 'AK-47' without the object losing any of the characteristics associated with the noun.

[232] P. 35

4.3. Most Promising Response to OQA: (Synthetic Naturalist) Moral Realism

I believe that the most promising semantic account of moral terms is the moral realism represented by Boyd[233], Sturgeon[234], Brink[235], Railton[236], Sayre-McCord[237], and Copp[238]. It became known for extending the Kripke's causal theory of reference[239], Putnam's semantic externalism[240], and Quine's coherentism[241] to the analysis of moral language to respond to the OQA.[242]

As I have argued, that action φ promotes, respects, advances etc. value V is a reason to φ. Then, I argued that valuing is best understood as believing value V is good for *x*. Some examples of things believed to be valuable (good) for *x* can be: autonomy, pleasure, perfection, equality, etc. However, there is a threat of circularity at this point. The analysis of value in terms of beliefs about value runs the risk of circularity. To say that V is valuable because *x* believes that V is good for *x* seems circular and uninformative.[243] If we ask why is V good? Answer is "because *x* believes that

---

[233] Boyd, R. How to be a moral realist. In: Sayre-McCord (ed.) *Essays on moral realism.* Ithaca, NY: Cornell University Press, p.181-228. 1988.
[234] Sturgeon, N. What Difference Does It Make Whether Moral Realism Is True? *Southern Journal of Philosophy*, v.24, 1986; Moral Explanations. In Sayre-McCord (ed.) *Essays on Moral Realism*. Cornell University Press, 1988.
[235] Brink, D. Externalism Moral Realism. *Southern Journal of Philosophy*, v.24, 1986; Realism, Naturalism, and Moral Semantics. *Social Philosophy and Policy* v. 18, 2001; Moral Realism and the Sceptical Arguments from Disagreement and Queerness. *Australasian Journal of Philosophy*. 62:2, 2006; *Moral Realism and the Foundation of Ethics.* Cambridge University Press, 1989.
[236] Railton, P. 'Naturalism and Prescriptivity'. *Social Philosophy and Policy* 7 (1):151 (1989); *Facts, Values, and Norms: Essays Toward a Morality of Consequence*. CUP, 2003.
[237] Sayre-McCord, G. The Many Moral Realisms. *Southern Journal of Philosophy* v.24, 1986; Moral Theory and Explanatory Impotence. In Sayre-McCord (ed.) *Essays on Moral Realism*. Cornell University Press, 1988;
[238] Copp, D. Moral Realism: Facts and Norms. *Ethics* v. 101, 1991; *Morality in a Natural World*. Cambridge University Press, 2007.
[239] Kripke, S. *Naming and Necessity*. Harvard University Press, 1972.
[240] Putnam, H. The Meaning of "Meaning". In: *Mind, Language and Reality*, Cambridge University Press, 1975.
[241] Quine, W.O. On the Nature of Moral Values. In *Theories and Things*. Harvard University Press, 1982; Two Dogmas of Empiricism. In: *From a Logical Point of View,* Harvard University Press, 3rd Ed, 1980. Epistemology Naturalized. In *Ontological Relativity and Other Essays*, Columbia University Press, 1969.
[242] Boyd: "It has occurred to many philosophers that naturalistic theories of reference and of definitions might be extended to the analysis of moral language. *If* this could be done successfully *and* if the results were favorable to a realist conception of morals, then it would be possible to reply to several anti-realist arguments." (1988, p. 199)
[243] Brink notes the circularity: "Here we invoke the very value we are analyzing in our analysans. It is true that, on this view, we analyze V, not in terms of (characteristics of) V, but in terms of beliefs about V. But if we accept the not unreasonable assumption that any story about what makes a belief a belief about V must eventually advert to V, then it appears that this sort of analysis is ultimately circular" Brink *Significance of Desire*, p. 21

V is good*"*. But then we must ask, "and why does *x* believe that V is good?*"* There seems to be no other answer than going back to "because V is good." An account of why V is good presupposes that V is good. To stop the circularity, we need to offer an account of the reference of "good". Reductionist naturalism will simply *identify* reference directly with cluster of natural facts and properties. On the other hand, some philosophers accept what Wedgwood calls the quietist view, for which indeed no substantive and non-circular analysis of goodness (value) is possible.[244] The first option, as we know, faces Moore's open question argument. The second is unattractive because sounds like table-pounding. My externalist inclinations lead me to think that there must be some naturalist, though not reductive, end to the problem of circularity. In other words, something extra-linguistic facts about the things valued that accounts for value statements, which does not use the notion which is trying to explain/analyze.

The way out is to say that the goodness of, say, perfection, autonomy, etc. is *constituted* by a number of natural properties and facts which multiply realize goodness.[245] Properties in the cluster can be weighted in terms of its sufficiency and necessity in realizing goodness. For example: the natural properties that constitute the goodness of a given token of autonomy (an action that instantiates autonomy or that promotes it) will vary from agent to agent, place to place and so forth. However, certainly there are natural properties that can't be absent. For example, some degree of absence of physical impediments seems a necessary member of the cluster of natural properties that constitute the goodness of autonomy. Identifying which of these facts and properties matter more than others for the cluster to realize value will depend on how or whether the beliefs about these properties and facts cohere with other evaluative beliefs, and ultimately, beliefs about extra-linguistic reality.

---

[244] Wedgwood, Nature of Normativity, p. 18
[245] Brink, Moral Realism, pg. 158, Wedgwood, *Intrinsic Values and Reasons for Action*, pg. 328

One way to motivate this view is to start from the Kripkean attempt to a non-circular semantic theory of truth. According to the causal theory of reference, meaning consists in how a term denotes a property or relation that causally regulates the use of this term in a given context or possible world.[246] Boyd argues that moral terms have synthetic definitions just like in science and that according to his version of causal theory, "*reference is itself an epistemic notion and the sorts of causal connections which are relevant to reference are just those involved in the reliable regulation of belief* "[247] The epistemic notion of reference is nothing more than to say that the historical-causal connections that unite the extension to the speaker control or regulate his beliefs. Boyd describes the process as follows:

> Roughly, and for nondegenerate cases, a term *t* refers to a kind (property, relation, etc.) *k* just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term *t* will be approximately true of *k*. Such mechanisms will typically include the existence of procedures which are approximately accurate for recognizing members or instances of *k* (at least for easy cases) and which relevantly govern the use of *t*, the social transmission of certain relevantly approximately true beliefs regarding *k*, formulated as claims about *t*, a pattern of deference to experts on *k* with respect to the use of *t*, etc. When relations of this sort obtain, we may think of the properties of *k* as *regulating* the use of *t* (via such causal relation), and we may think of what is said using *t* as providing us with socially coordinated *epistemic access* to *k*; *t* refers to *k* (in nondegenerate cases) just in case the socially coordinated us of *t* provides significant epistemic access to *k*, and no to other kinds (properties, etc.)[248]

[246] The vocabulary of possible worlds is used to clarify the modal notions of necessity, contingency, and possibility. According to this vocabulary, a truth is necessary if, and only if, it is a truth in all possible worlds; on the other hand, a truth is contingent if, and only if, it is a truth in the present world, but not a truth in all possible worlds; and yet, a truth is merely possible if it is a truth in some possible non-actual worlds. The vocabulary of possible worlds can be used even with the rejection of the theory of modal realism defended by Lewis in Lewis, D. *On the Plurality of Worlds*. Oxford: Basil Blackwell, 1986. We can work on a conception where this vocabulary is just a way of talking about modalities, without committing ourselves to the very existence of possible non-actual worlds, as Kripke did. Possible worlds also serve to determine the domains of the different types of modalities as I have done in Chapter 1 of this dissertation. For example, it is logically possible that the sentence "water is not H_2O" is true. There are logically possible worlds in which water is not H_2O. However, in worlds that are not merely logically possible, worlds in which we consider not only the laws of logic but also the nature of things (the physical, chemical, etc., laws to which they are subjected), it seems that water is necessarily H_2O; it is plausible to think that being composed of two molecules of hydrogen and one of oxygen in a given relation is a property that water has in all possible worlds in which such a natural category exists. Thus, in a given domain of possibilities, the sentence "water is H_2O" seems to express a necessary truth and in another domain of possibilities expresses a contingent truth. In the case of metaethics, a moral sentence expressing a necessary truth would be one that is true in all possible worlds where the sentence is uttered.

[247] Boyd, 1988, p.195.

[248] Boyd, 1988, p. 195

In response to OQA, this view asserts that just as in causal theories of reference for proper nouns, names, and natural kinds, evaluative terms behave like rigid designators.[249] In other words, relation between evaluative and the non-evaluative terms is synthetic and *a posteriori*. Notwithstanding first impressions, this view does not claim that evaluative properties are *immediately* constituted by natural properties, but with functional properties. These functional properties, in turn, are made of, or constituted by, homeostatic bundles of natural properties. The individuality of a property of this type is given by the causal function it exercises in conserving homeostatic sets of conditions of the realization of value: "*considerations of human well-being play a significant role in determining what is said to be 'good'*"[250]. A functional property becomes the reference of an evaluative term when it causally regulates the use of that term. The physical embedding of those functions is its ultimate extension, *viz.* the homeostatic bundles of natural properties. These bundles sustain evaluative beliefs and take the responsibility for the justification of the judgments expressing those beliefs.

The relation between the most basic naturalist facts and properties and evaluative facts and properties is that of supervenience, not of synonymy. I understand supervenience, like Brink, as a nomological/lawlike relation between the properties of the two linguistic realms. However, since the basis of which natural properties that ultimately realize value, this view comes with a distinction between strong supervenience and weak supervenience. The first claims that the set of organized combination of natural properties *necessitates*, in the metaphysical sense[251], their

[249] Rigid designation is a feature of terms which denote the same particular in all possible worlds where this particular exists, non-rigid designators, in turn, denote a single particular in only a few possible worlds. For example, the name "water" denotes H_2O in all possible worlds in which H_2O exists. On the other hand, the description "The author of Metaphysics" denotes Aristotle only in some possible worlds, since it is logically possible that Aristotle might not have written the Metaphysics or that it was written by somebody else. Two applications of rigid designation in value theory: Hurka, T. *Perfectionism*. OUP, 1993; and Slote, M. 'A New Kind of Reference Fixing' in *Moral Sentimentalism*. OUP, 2010.

[250] (1988, p. 211)

[251] A posteriori, not logical, not conceptual, Brink, Moral Realism, p. 175

supervening properties. Weak supervenience settles for a set of sufficient number of properties within specified contexts and perhaps not too different possible worlds.[252] Although it may be difficult to investigate and find a strong supervenience base given the number of variables, it seems that any involuntary restraint of physical movement and social coexistence are part of the bundle of properties that necessitate the badness of slavery not just in the actual world but also the badness of systems considerably different from our own in every nearby possible world where some autonomy exists. Now, note that from the point of view of an externalist, none of the bundles must necessarily include the agent's motivation. The functionalist wrinkle of the view just goes as far as to say that these properties play significant functions in considerations of well-being. But this notion of well-being does not need to be grounded internally on desires or the like.

What exactly is the role of homeostasis in this view? According to moral realism, a moral term, or any evaluative term for that matter, denote a collection of properties contingently, but homeostatically clustered in nature. Homeostasis is the equilibrium among the properties in the cluster. For example, health is homeostatic per excellence. But it would be utterly difficult to determine exactly all the properties that must figure in the cluster of natural properties that multiply realize health. So, the balance of properties will often need to be taken as a whole, instead of a discrete list. Once a cluster assumes functions in evaluative language, it exercises normative force or authority in human action.

Before I move on to its problems, moral realism is incomplete without an account of equilibrium and convergence, and one about moral motivation and reasons for action. At this point the accounts are brief because I will develop them after I introduce the moral twin earth objection.

[252] Brink, Moral Realism, p. 160

4.4. Equilibrium and Convergence

The best prospect for moral realism is to rely on a sort of combination between supervenience and coherentism[253], of which Brink's coherentism is an example[254], drawing on the works of Brandt[255], Rawls[256].This hybrid coherentism is not very different than coherentism in noncontroversial natural sciences.[257] For moral realism, even if scientific beliefs and value beliefs are not exactly alike, a difference that the view seeks to clear away, Boyd, Brink, and many others to count argue that both, to a large extent, rely on reflective equilibrium, socially and epistemologically, for successive approximations to the truth.[258] Just like reflective equilibrium in value theory, natural sciences require a division of theoretical labor, which Boyd exemplifies with the dialectical relationships between current theory, observation, and methodology as well as the dialogue between members of scientific communities and learned societies. Like the concepts of mass, energy, beauty, and health, the homeostatic sets of natural properties converging into the goodness of autonomy, perfection, equality, and so on are constituted by terms whose definition are in constant reformation vis-à-vis dialectical scrutiny leading to approximations to the truth.

So, the justification of any belief will be dependent of its relation to the totality of other beliefs the agent holds. The belief p is justified when p is a genuine member of a maximally

---

[253] I have in mind something like Quine's suggestion: "The observation sentence is the cornerstone of semantics. For it is fundamental to the learning of meaning. Also, it is where meaning is firmest. Sentences higher up in theories have no empirical consequences they can call their own; they confront the tribunal of sensory evidence only in more or less inclusive aggregates. The observation sentences, situated at the sensory periphery of the body scientific, is the minimal verifiable aggregate, it has empirical content all its own and wears it on its sleeve." 'Epistemology Naturalized' in *Ontological Relativity and Other Essays.* Columbia University Press, 1969. p.89

[254] Brink, D. *Moral Realism,* pg. 122-

[255] Brandt, R. *A Theory of The Good and The Right*. OUP, 1979; 'The Science of Man and Wide Reflective Equilibrium', *Ethics*, 100: 1990, 259–278.

[256] Brandt and Rawls have significant epistemological disagreements about the foundations of reflective equilibrium but are equally interested in some form of coherence or equilibrium. Rawls, J. *TJ, PL, Kantian Constructivism,* and *Restatement*.

[257] Put in another way: "Counterfactuals are the building blocks of moral behavior as well as scientific thought." Pearl, J. and Mackenzie, D, *The Book of Why*. Basic Books, 2018.

[258] Boyd, pg. 189-90.

coherent system of beliefs. But as justification *is* inferential, so the system must include second-order beliefs distinguishing kinds of beliefs and about how they are to be justified. First-order beliefs are about extra-linguistic world. Second-order beliefs in turn are about the relationship between first-order beliefs and the world and about the relations between the first-order beliefs themselves, the nature of belief formation, degree of certainty, fallibility, and so on. For, example, it is a second-order belief that one can be justified though based in false beliefs.[259]

Now the evaluative beliefs are themselves *inferential.* What characterizes an argument as inferential is the fact that there is a connection between A and B. Given this connection, we can say that since we have B, the best inference we can do is to conclude that A. A does not follow logically from B; however, given the causal connection between the two, the existence of A is the best explanation for the existence of B. For example, if we have the conditional sentence "If it rained, then the street is wet", and we observe that the street is wet, we concluded that it must have rained. As there are circumstances where the street is wet and it has not rained (for example, the street was wet by water sweeper truck), we must ensure the reliability of the inference by identifying other effects produced by A, for example, wet roofs, which could not have been done by a water sweeper truck. This application of inferential reasoning in this context is warranted by the fact that it is used in favor of scientific realism, where truth that integrates successful theories as to its predictions, or the postulation of certain entities and laws, provides the best explanation of phenomena observed.[260]

But these epistemological mechanisms plus the plurality of values that I have been assuming invites for significant divergence about beliefs in supervenient properties. There are

[259] Alvin Goldman, "What is Justified Belief" In George Pappas (ed.), *Justification and Knowledge*. Boston: D. Reidel. pp. 1-25 (1979)
[260] Lipton, P. *Inference to the best explanation*. Routledge, 2005

costs in my view and one of them is an embedded assumption in favor of value pluralism. Given that, even if people find some degree of equilibrium among these values within their own lives, these intrapersonal equilibriums are too fine grained and so it is not difficult to imagine that people will arrive at conflicting intrapersonal equilibriums given the different significance they internally assign to these values.

All of these equilibriums will have some degree of imprecision or vagueness about membership in a homeostatic set, such that axiological terms are "open textured" concepts.[261] We do not always know, or may not know yet, exactly what are all the natural properties embedding happiness, autonomy, perfection, so there is some indeterminacy about what items fall into the extension of these terms.[262] According to moral realism, this indeterminacy is constantly mitigated via dialectical scrutiny about each of these concepts, not different than in natural sciences. Lifestyles and communal agreements are equally open to similar scrutiny. By lifestyles, I mean people's private homeostatic set of values. The more we learn about what are all the natural properties and functionings pertaining to each of these homeostatic sets, we are led to adjust on extensions, first and second order beliefs, and even overall lifestyles, and so on. But given that the process of pursuing equilibrium is affected by time, place, etc., we can expect substantially different resulting equilibriums in different possible worlds.[263] If that is correct, axiological disagreements do not necessarily indicate the wholesale absence of shared, causally fixed, extensions for axiological terms. For instance, the homeodynamic process may be taking place where background conditions have led the working conceptions to be at different stages of reformation, hence development.

---

[261] Boyd, pg. 196
[262] Extensional vagueness about properties and relations may or may not be eliminable. There is a new optimism in that regard. For instance, Fuller, S. and Lipinska, V. *The Proactionary Imperative*. Palgrave Macmillan, 2014.
[263] Boyd, p. 197

4.5. Morality, Motivation, and Reasons

Internalism and externalism about motivations and reasons for action are two opposing views about the *modality* between morality and its motivational force in the first case, and morality and rationality, in the second. Some philosophers have tried to maintain both moral realism and MJI. Those of Kantian inclination have defended that moral beliefs are necessarily motivating without a corresponding desire, and some of Humean inclination try to establish some necessary (or at least highly regular and reliable) connection between moral beliefs and desires.[264] However, the moral realism I am interested in developing is externalist in two ways. First, moral judgments are not necessarily connected to motivation. Second, even if moral reasons necessarily hold ineliminable justificatory residual[265], they are not necessarily overriding. But why should a moral realist endorse motivational judgment externalism (MJE) and anti-moral rationalism?

Internalism has been one of the most pervasive premises in anti-realist arguments, such as the Is-Ought Thesis and the Queerness Argument. If it turns out to be false, moral realism is discharged from the burden of having to accommodate it. But moral realism is not only interested in rejecting internalism, but also in providing the correct account of moral motivation and rationality. Externalism in both cases is, I submit, closer to the truth, conceptually and empirically. In moral realism's seminal work, Boyd touches on the motivation and rationality of morality very briefly. But he just assumes that, like in constraints to rational choice over scientific matters, it is possible to truly understand them, and not have any correspondent motivational response. Likewise, it is perfectly rational to question their *ultimate* justification in relation to other rational

[264] Given that historically they diverge between motive of duty vs. sympathy over the source of moral motivation. Brink, p. 12-15 1997

[265] Still provide sufficient justification though superseded by other stronger reasons.

choice constraints.[266] However, Boyd does not substantially defend either one, but leave them as legitimate conceptual possibilities.[267]

A second reason for moral realism to go externalist is that it takes the *all-things-considered* rationality of morality and its motivational force to be questions that can only be known *a posteriori*.[268] Because we need to know the degree of rational force of other options at stake at the moment of decision, such as the degree one would be justified in acting prudentially, politically, etic. instead of morally, those degrees cannot always be told ahead of time. Even if it was a prediction, it would still be an induction or abduction (inference to the best explanation). It is important to say that even if MJI were a conceptual truth, we can still articulate the question whether a moral demand one is motivated for, is all-things-considered rationally justified.[269] And if there is at least the logical possibility that a moral reason, regardless of being internally motivational, be overridden by competing reasons, say a professional reason, or aesthetic, or prudential, then moral rationalism, is conceptually false. If moral rationalism is conceptually false, then a deliberator may well hold *bona fide* moral judgments without being motivated by the judgment. Otherwise, we would not be able to say that the deliberator sincerely holds several reason-judgments at a given time of deliberation. When someone deliberates about what one ought to do, all things considered, this deliberator believes there to be several different and often conflicting reasons for action. On its face value, the fact that the agent ultimately acted according to a, say, political reason, says nothing about whether the agent did or did not believe in there being

---

[266] Boyd pgs. 186 and 214-6. Like Mackie, Boyd takes facts about practical reason to inform facts about moral motivation, which is a strategy I subscribe all along this dissertation.

[267] On MJE: "there could be nonhuman cognizing systems which could understand the natural facts about moral goodness but be entirely indifferent to them in choosing how to act." On MRE: "It is hard to see how the naturalistic moral realist can escape the conclusion that it would be *logically possible* for there to be a human being for whom moral judgments provided no reasons for action." Boyd p. 214.

[268] Brink, p. 28 Externalist Moral Realism, 1986

[269] It might be held that the recognition that one course of action is morally preferable to another *necessarily* provides a reason for action (even if not a decisive one) Boyd p. 214

other reasons, moral, prudential, legal, etc. equally applying to the deliberation under demand. In absence of contrary evidence, I have no grounds to doubt the deliberator when he reports to believe in there being reasons moral, prudential, legal, even though he is/was not motivated to act according to their prescriptions, since he ended up acting according to a political reason. We do not need to say that all the other reason-beliefs held by the agent simply have their previous justification deleted. When talking about the reasons that lost the competitions, some will say that even though the agent recognizes demands/obligations/requirements, they do/did not generate reasons. But I find that implausible. If the agent acknowledges that a given field of practical reason contains an obligation, surely that is a reason to act in some way, even if he does not elect that reason as the all-things-considered reason. I suggest that we can accept its residual justification, and yet let other reasons exceed in rational force.

I discussed MJE and anti-moral rationalism in the first and second chapters, but if some version of anti-moral rationalism, even if weak, is closer to the truth, then so is MJE. And if those two points are correct, then MJI cannot be a constraint on the meaning of moral terms and hence no impediment to fixing the reference of *moral* terms. Whether an all-things-considered reason is necessarily motivational[270] is a question I leave for a future endeavor. For now, it suffices to argue that a lack of *moral* motivation does not necessarily imply a rational deficiency, at least before all-things-considered procedures take place.

4.6. The Moral Twin Earth Argument

Recall the internalist thesis that motivation and rational authority are necessary parts of the meaning of moral terms. If those aspects are only contingently integrated in the meaning of these

[270] Cf. Weatherson, *Normative Externalism*

terms, the determination of reference will occur freely without the controlling action of a general property[271] whose presence fixes the reference:

> There is supposed to be a single objective property which we´re all talking about when we use the term 'good' in moral contexts. But people´s moral concepts differ profoundly. How can it be maintained that our radically different concepts of 'good' are really concepts of one and the same property? Why not a different property for each significantly different conception of the good? Don´t the radical differences in our conceptions of the good suggest either a noncognitivist or constructivist conception of the semantics of the moral terms?[272]

The moral twin earth argument (MTE) suggests that these radical differences are due to the fact that semantic moral realism neglects the motivational aspects inherent in the meaning of moral judgments. That is, the proposed process of determining the reference is not able to capture the normative aspect, which would function as a fixer of the reference since it is the common element to all moral qualifications. The moral twin earth argument espouses an explanatory challenge like Moore's OQA, now revised to attack synthetic identity/definition.

MTE is an application of Putnam's thought experiment Twin Earth.[273] The goal of the experiment is to show that a strictly internal state of a cognizant living creature does not, fully, determine extension. It goes as follows: we have Oscar 1, inhabitant of the Earth and Oscar 2, inhabitant of the Twin Earth. The two Oscars are internally identical. If there are no distinct psychological states without a corresponding difference of physical states, then both Oscars are in exactly the same psychological states when they entertain the same thoughts. Now consider the word 'water' and its uses. The psychological state in which Oscar 1 is when he understands 'water' is identical to the psychological state in which Oscar 2 is when he understands 'water'. Thus, Oscar 1 and Oscar 2 associate the term 'water' with the same intension.[274] However, the experiment

---

[271] Such as Mackie's notion of "universal motivation/prescription" internal to moral terms, discussed in chapter 3.
[272] 1988, p. 213
[273] Putnam, H. The Meaning of "Meaning". In: *Mind, Language and Reality*, Cambridge University Press, 1975.
[274] Worth noting that anti-descriptivist semantic theories hold that intension determines extension. In this case, there is a strong relation between knowing a term's intension and linguistic competence, since without knowing the term's intension, the speaker won't be able to determine its extension, and hence not able to use the word. Semantic externalists defend or assume that linguistic competence does not require high knowledge of a term's intension. For

establishes that the extension of 'water' is different for Oscar 1 and Oscar 2. On Earth, the extension of 'water' is H_2O, while on Twin Earth it is XYZ. For the perceptual system of humans and twinhumans, H_2O and XYZ are indistinguishable at normal temperatures and pressures, they taste alike, they equally quench thirst, fill oceans, rivers, lakes, and perform all the same global hydrologic cycles. This characterizing is to suggest that the two different extensions even have the same descriptions.

Now, if a spaceship from Earth visited Twin Earth, the supposition at first will be that 'water' has the same meaning on Earth and on Twin Earth. But this supposition would be corrected when it is discovered that 'water' on Twin Earth is XYZ. Symmetrically, if a spaceship from Twin Earth ever visits Earth, then the supposition at first will be that the word 'water' has the same meaning. The supposition will be corrected when it is discovered that 'water' on Earth is H_2O. Putnam then mobilizes linguistic intuitions to argue that since 'water' refers to different extensions in the mouths of each Oscar, they do not have the same meaning. The words have the same intension, but they mean different things. When the two groups learn that extensional difference, they realize it would be silly for them to think that they have differing views about the real nature of water.[275] In other words, there is no possibility of genuine disagreement on matters about "water".

Before Henry Cavendish's *Three Papers Containing Experiments of Factitious Air* in 1766[276], and the chemical nomenclature of water on Earth was established by Lavoisier et. Al. in 1780s[277] when the natural elements of water were not yet known, residents of Earth and Twin

---

more on this point, Dowell, "The Metaethical Insignificance of Moral Twin Earth" in *Oxford Studies in Metaethics*, v. 11.

[275] New Wave, p. 166

[276] The Royal Society Publishing, Vol. 56. 1766.

[277] American Chemical Society International Historic Chemical Landmarks. Antoine-Laurent Lavoisier: The Chemical Revolution. http://www.acs.org/content/acs/en/education/whatischemistry/landmarks/lavoisier.html (accessed January 23, 2020).

Earth, would be sufficiently justified to suppose that they understood 'water' exactly alike[278], though the extensions of 'water' in each case remained H_2O on Earth and XYZ on Twin Earth all the way. After 1780s, had the Oscars met again, they would certainly think that they do not mean the same thing, since their words refer to different things. Thus, Putnam concluded, the *meaning* of words of the same epistemic nature of 'water' is not merely a function of the strictly internal properties of the speaker, such as mental states or semantic rules.

Now let us see how MTE goes. Horgan and Timmons have used MTE not just against moral realism[279], which they call new wave moral semantics, but also Michael Smith's moral rationalism[280] and Frank Jackson's moral functionalism[281]. In the case of moral realism, they argue that if the term 'good' rigidly designates the unique property that causally regulates the use of 'good', then we can construct a similar scenario for moral language. Suppose human axiological terms on Earth are regulated by certain natural properties, and that, as a matter of empirical fact, these are consequentialist properties whose functionings is captured by some specific consequentialist normative theory Tc. Next, suppose that the residents of Twin Earth possess a quite similar axiological vocabulary, employing words like 'good' and 'right' like humans for the same purposes. But on Twin Earth, twin-axiological terms are causally regulated by natural properties whose functionings are captured by a deontological normative theory Td.[282] Hence,

---

[278] Such as making twin inferences and exercising propositional attitudes employing 'water'.
[279] Horgan, T. and Timmons, M. Troubles for New Wave Semantics: The Open Question Argument Revived. *Philosophical Papers*, Vol. 21, n. 3, 1992; and Troubles on Moral Twin Earth: Moral Queerness Revived. *Synthese,* Vol. 92, No. 2, 1992.
[280] Horgan, T. and Timmons, M. Troubles for Michael Smith's Metaethical Rationalism. *Philosophical Papers*, Vol. 25, n. 3, 1996.
[281] Horgan, T. and Timmons, M. Analytical Moral Functionalism Meets Moral Twin Earth. In Ian Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*. Oxford University Press, 2009.
[282] H&T. Troubles for New Wave, p. 163-4

'good' rigidly designates different natural properties that causally regulate their use on Earth and Twin Earth.

Now, in the case of 'water', Oscar 1would hold to his *paradigm,* as Putnam names the set of factors determining H_2O to be its extension, and say "XYZ is not water. The thing Oscar 2 drinks and fills lakes, etc. is just coincidentally called by the same word. So, we, born on Earth cannot call XYZ water, because the word does not refer to the same thing. But note that Oscar 1's paradigm includes *inter alia* a physical-chemical theory, which both Oscars believe to be the correct theory. It seems that it is this very same theory which Oscar 1 employs to identify the elements of 'water' on Twin Earth. And if Oscar 2 agrees with Oscar 1, they would probably conclude that the fact the words are the same is just a cosmic coincidence. Perhaps, inhabitants of Twin Earth would slowly start calling XYZ somehow else, or vice-versa. But note that it seems that inhabitants of both worlds are employing the same physical-chemical paradigm.

In the case of MTE, different moral paradigms generate different moral intensions, endorsed by inhabitants of each world. Horgan and Timmons argue that once Oscars 1 and 2 realize that the words are regulated by different properties, they would not have the same attitude as to the case of 'water'. Each one would hold fast to their own paradigms, and Oscar 1 would see no reason to revise the intension he attributes to 'good', and the same for Oscar 2. They have different substantial theories about the nature of morality in a way that they do not about the nature of water. The fact Tc and Td are technically tied assumes probative value for the view that 'good' does not rigidly designate either C or D. Were 'good' to rigidly designate either one, we expect that one of the parties would revise their moral paradigm, in favor of the designation which is closer to the truth. *If moral terms do not rigidly designate anything, then it will always be an open question whether anything is actually 'good'.*

Let us see how MTE is a revision of OQA. Moore argues that while it is tautological to ask, "Entity e possesses natural property N, but is it N?" It is always meaningful to ask, "Entity e possesses natural property N, but is it good?". In the case of 'water' we would have "Liquid l is H_2O, but is it 'H_2O?" as uninteresting and "Liquid l is water, but is it H_2O?" was a meaningful question until 1780s. In the case of morality stipulated by MTE we have:

> On Earth:
> Entity e possesses natural property C picked up by Tc, but is it good?
>
> On Twin Earth:
> Entity e possesses natural property D picked up by Td, but is it good?

Horgan and Timmons argue this is moral theory stand-off that cannot be resolved if causal semantic naturalism is true. Then they say this putative unresolvable moral disagreement constitutes strong empirical evidence against causal semantic naturalism, since our intuition is that the disagreement can be resolved. If 'good' rigidly designates C on Earth and rigidly designates D on Twin Earth, then they have different meanings. Horgan and Timmons say: "if causal semantic naturalism were true, then recognition of these differences ought to result in its seeming rather silly, to members of each group to engage in inter-group debate about goodness – about whether it conforms to normative theory Tc or to Td."[283] However, Horgan and Timmons say that "such inter-group debate would surely strike both groups not as silly but as quite appropriate, because they would regard one another as differing in moral beliefs and moral theory, not in meaning."[284] I agree. But I do not believe causal semantic naturalism prevents an inter-group debate, and eventually an agreement. Why, can't Oscars 1 and 2 debate about substantial views of morality and eventually converge on a sufficiently homeostatic paradigm? They were able to endorse a

[283] New Wave, p. 166.
[284] New Wave, p. 166

sufficiently similar physical-chemical paradigms before they could understand their use of 'water' was referring different extensions. I perfectly agree. So, the question is: why can't both groups converge?

Horgan and Timmons seem to believe in two explanations for the resilience of the moral disagreement. First, and less troublesome, is the difference in the psychological hardware intrinsic to Oscars 1 and 2: "Twin Earthlings tend to experience the sentiment of guilty more readily and more intensively, and tend to experience sympathy less readily and less intensively, than do Earthlings."[285] It is fine that differences in psychological hardware factor in normative theorizing. But that does not necessarily preclude intergroup discussion and eventual homeostatic convergence, unless we are shown exactly why particular differences would do so. Maybe the differences would be too drastic and extreme. But in times of war, life is the ultimate good, and insofar both species want to preserve their own lives, and can somehow communicate, there is a way open for convergence. Movies such as District 9 and Alien Covenant depict such scenarios nicely.

There is a second reason for the resilience of the disagreement. Horgan and Timmons build up the function of moral terms supported on both MJI and MR. The function of *moral* terms is to point out to overriding considerations. See what they attribute as the intension of moral terms:

> "After all, the uses of these (moral) terms on Moral Twin Earth bear all the 'formal' marks that we take to characterize moral vocabulary and moral practice. In particular, the terms are used to reason about considerations bearing on well-being; Moral Twin Earthlings are normally disposed to act in certain ways corresponding to judgments about what is 'good' and 'right'; they normally take considerations about what is 'good' and 'right' to be especially important, even of overriding importance in most cases, in deciding what to do, and so on."[286]

So, it seems to me that MTE stands as a strong argument against moral realism if we establish that 1) Td and Tc are incorrigibly psychologically incompatible; 2) each group endorsing the views

[285] New Wave, p. 165
[286] HT, New Wave, p. 164

hold them to be necessarily motivational and overriding; and 3) that causal semantic naturalism prevents definitional revision. If these tenets are true, then homeostatic convergence is unlikely, and the question about what the natural properties designated by 'good' are remains open.

Hence, the substitution of Td-good by Tc-good would generate differences in the functionings of 'good', for example, inferential discrepancies or different truth-values. Thus, according to a causal semantic naturalism perspective, MTE shows that moral realism leads to moral relativism[287], for even conscious of those differences, the inhabitants of the two remain in disagreement.

In my view, even though at first the use of moral terms denotes different extensions, since they employ different normative theories, speakers can still participate in a genuine moral debate and eventually reach a minimally sufficient homeostatic convergence. The inhabitants use terms and theories with the same referential intentions, and that seems good raw material to continue on the development of the semantic dimension of moral realism.

To say that referential intentions are the same in both contexts means that moral terms are used with the intention to refer to the same extensions, that is, qualities of actions, characters, institutions, policies, and so on. MTE leads us to realize that there are two dimensions that contribute to the determination of the reference in the case of morality, namely, our referential intentions associated with the term, and the extensional properties of the term in the actual world of its expression. If these same referential intentions can be expressed in different possible worlds

[287] "This is also what the causal semantic thesis would seem to imply, inasmuch as the same 'M' is causally regulated by different N-properties on Earth and Moral Twin Earth. However, this would appear to imply interplanetary semantic relativism…If so, it appears that the semantic relativistic interpretation of Moral Twin Earth is inconsistent with moral realism. Moreover, interplanetary relativism might seem problematic for the moral realist if it is just the first step toward intraplanetary relativism." Brink, D. Realism, Naturalism, and Moral Semantics. *Social Philosophy and Policy*, 2001, p. 166-7

and shared intergroup, we may come to perceive more clearly what remains associated with the moral terms in any possible world, regardless of how the actual world will be.

4.7. Initial Steps to Overcoming the Moral Twin Earth Argument

In Putnam's *Mind, Language and Reality* we find a sketch of the psychological foundations that play some role in determining reference. Putnam is not very clear on this subject, but we find him saying something like the following:

> Imagine a race of creatures - perhaps higher animals - who are just beginning to evolve pre-speech behaviors. They notice middle-sized material objects of various kinds, and occasionally they need to call one another's attention to various of these objects. One especially clever creature develops the habit of *pointing with his finger* at an object and saying a noise, which sounds like 'Lewkthis'. Other members of the tribe imitate this (after recent observations of chimpanzee behavior both in the wild and in experimental situations, this is not so implausible); and soon the creatures develop a full set of what I shall call Gricean intentions (cf. Grice, 1957, 1968, 1969); that is intentions to call attention to an object by pointing and uttering ' Lewkthis'; the intention being to do this partly by recognition of this very intention (1975, p. 284).

> Thus, the principle describes intentions which actually exist and are for the most part honored in the linguistic community - and it is a normative principle; we should honor it, for otherwise stable reference to theoretical entities would almost surely be impossible (1975, p. 275).

Here we see how referential intentions would have been formed along the evolution of language. The story emphasizes the point at which language developed from our will to refer to objects, relations, and events. Our words depend in part on the world; but at the same time, language is the fruit of our own willingness to speak about certain things and not others. Note that Oscar 1 and Oscar 2 have the same referential intentions associated to ‘good’ as well as same are the referential intentions associated to ‘water’. However, although Horgan and Timmons admit that Oscars 1 and 2’s referential intentions are the same, they do not believe they are sufficient to fix the same reference[288]:

[288] Indeed, referential intentions are not sufficient. But they may work as a starting point for homeostatic convergence.

> A defender of moral realism might attempt to avoid Boydian causal semantics and appeal, say, to our **linguistic intentions** regarding our uses of moral terms like "good" and "right" in trying to give a realist account of moral reference. Again, the Moral Twin Earth argument is a recipe that can be adapted and applied to various realist attempts to nonarbitrarily pin down the referents of moral terms including proposals that do not make use of causal regulation stories. (1999, p. 66-7)

Perhaps they believe that referential intentions do not guarantee the *modal* character of necessity traditionally attributed to the relationship between a moral judgment and its motivational qualities or its overridingness. In other words, it seems that for Horgan and Timmons, neither the objectivity nor the normativity of moral judgments is explained by the introduction of the notion of referential intentions. But regardless of the theoretical value of the notion of referential intentions, this presupposition compromises moral realism with internalist perspectives, according to which moral judgments are necessarily motivating and/or overriding. But moral realism is not internalist in either way. Instead of tacking this point head-on, Timmons says:

> First, I do think that the point of Moral Twin Earth is that moral discourse is essentially evaluative discourse. Second, I agree with Brink´s assessment of the various internalist theses he considers – they implausibly represent the connection between morality and action. **But from the fact that standard versions of internalism are implausible, we should only conclude that the more standard forms of internalist should be rejected** (1999, p. 68).

So, again we have MRa and MJI at the center of contention between moral realism and all other views, which are internalist or rationalist one way or another. But before I go back to that topic, let me deploy the notion of referential intentions in favor of moral realism.

Brink[289] and Copp[290] propose some adjustments in Boyd's conception of causal regulation. For them, meaning is determined by two factors: 1) Intension, and *inter alia*, the referential intentions and 2) the object(s) that is in causal contact with the speaker at the act of denotation: "*In general, the referent of a term is determined by the conspiracy between the present nature of the stuff in the samples we use in explaining it, on the one hand, and our referential intentions, on the*

[289] 2001
[290] 2007

other hand."[291]. It is the matching between these two factors that determines the extension of a term. In order to arrive at the content of the referential intentions Brink starts from the observation that, "*one way in which referential intentions function ... is to provide continuity of reference among community of inquirers*"[292] This continuity takes place through the sharing of referential intentions, that is, we make approximations to a uniform use of words by departing from the intention to talk about the same thing. As Brink says:

> The intentions of the moral inquirers to say and think things about the same features of people, actions, and institutions as each other, will block ascription of different meanings or reference to fellow participants in a common moral inquiry[293]

Shared referential intentions is a first level of playing field to depart towards homeostatic convergence. It opens the floor for speakers to debate on what kind of object may satisfy their referential intentions. In that manner, the method for discovering which homeostatic clusters of natural properties that multiply realize functional moral properties depends on a dialectic that starts from the intentions of the speakers during their initial use of moral vocabulary. These intentions help to guide the investigation and determination of the objects that happen to fall within the extension of those terms.

It could be said that employing referential intentions in the dialectic of determining the base level natural facts may only make the divergence even worse, where each one would endlessly point out to what best suits him as an object of his intentions. Rather, it helps to *mitigate* moral disagreement. Referential intentions do not suffice to fix the reference however, because to function as a first approach for the dialectic process over what natural properties fall into the supervenience base, the intentions must mean some property sufficiently generic and abstract to

---

[291] Copp, 2007, p.
[292] Brink, 2001, p. 171
[293] Brink, 2001, p. 173

constitute the initial content of the referential intentions.[294] Once we have that content sufficiently established, we may proceed to try to bring together the natural properties within a homeostatic set. Moral realists differ in some respects about this criterion, but all admit that it must be somewhat generic. Here is a way to express this requirement:

> To apply this account to the moral case, we need some parallel descriptive specification of the referential intentions of moral inquirers. This requires a descriptive formulation of the moral point of view, but it must be a description that is sufficiently abstract so that a wide variety of views might be thought to satisfy this description. Moreover, what best satisfies this description must be a matter of substantive moral theory.[295]

And here how we obtain these generalizations:

> Our actual use of moral predicates is imperfectly guided by our (perhaps implicit) acceptance of moral principles that identify morally relevant factors. We identify principles by looking for patterns in our actual and counterfactual judgments that employ those predicates; we test these principles by drawing out their implications for real or imagined cases and comparing these implications with our own existing or reflective moral assessments of those cases.[296]

Brink suggests that through this process we identify a general property or principle. Thus, to the extent that referential intentions are abstract enough to be similar enough in the two contexts of evaluation, it is possible for the debaters to identify the differences in the normative theories that ground their own terms or moral judgments and to adapt to one another or even to adopt one of the two to the detriment of the other in order to denote the same extensions or the same truth-values, that is, to begin building the basis of a possible future moral agreement. Imagine, for illustration, that the concept of 'justice' picks out different extensions as when employed by utilitarians, egalitarians, and, libertarians and thus denote different extensions. However, we can say that there is still something in common and abstract associated with 'justice' that survives through the different normative conceptions. This *something abstract* transpires in the referential

[294] Copp, 2007, p. 207
[295] Brink, 2001, p. 172
[296] Brink, 2001, p. 168

intentions of speakers who, when talking about 'justice', are intending to speak of the *same thing.* This sharing opens up an interpersonal dimension of moral language. According to Brink:

> This understanding of the concept of morality admits of many very different conceptions. It admits of a Kantian interpretation in terms of the universal law formulation of the categorical imperative, as well as a Humean interpretation in terms of affective responses from the general point of view.[297]

That is a great advantage to moral realism, for the use of referential intentions, supporting Boyd's epistemic notion of reference allows us to explain how two debaters discussing a moral issue/dilemma can, while sustaining different moral theories, discuss the moral qualifications of the same natural event, and eventually make trade-offs between their moral beliefs. Though they hold different beliefs, those beliefs are sustained with shared intentions to refer to the same extension. From a sharing of similar referential intentions, that is, as to the general principle that controls the use of the moral terms in question, the debaters become apt to identify the differences in the normative theories that they endorse, make theoretical fine-tunings, and eventually find a possible common extension for their moral qualifications, i.e. reach a moral agreement.

Copp equally emphasizes that we need to start off from an abstract property inherent to moral terms that can be accepted as inherent in those terms when they represent different normative concepts. This general characteristic allows the initiation of possible translations between the moral and the natural terms, without the terms yet denoting equal entities: "To understand this point, it needs to be understood that translation is more like trying to find someone who looks enough like you to pass muster in a Police lineup than it is like trying to find your identical twin."[298] Copp uses the term 'milk' to show that even though the terms corresponding to the two groups differ as to the property they denote, they do not necessarily differ in meaning. In short, it shows

[297] Brink, 2001, p. 174
[298] Copp, 2007, p. 213

that functional aspects are at least equally relevant to determining reference than only the constitutional aspects of extension.

> Clearly, then, the "physical properties" that determine whether two samples [of milk] bear the same$_L$ relation to each other need not be chemical in nature. Milk is in fact *a liquid that is produced by the mammary glands of female mammals that have recently given birth, in order to nourish their young*. It is presumably this 'functional' and 'genetic' property that, on Putnam´s account, would be relevant to determining whether a glass of a white liquid bears the same$_L$ relation to the milk in my fridge (Putnam 1975, p. 239, 241 *apud* Copp, 2007, p. 205)

This notion of genetic or functional property seems to be the same notion that Brink had in mind when he said that the referential intention refers to a property abstract enough that it can be employed by everyone and even people who hold different normative theories. In the case of MTE, the same referential intentions come into action, because the function of the terms is, in both cases, to assess actions, characters, institutions, policies, and so on. Hence, the parties are not necessarily in disagreement about the meaning, since the functional aspects associated with the terms are the same: "By construction of the Twin Earth scenarios, Earthling moral terms and Twin Earthling twin-moral terms **play the same role** in people´s lives on the two planets. The term "wrong" in Twin English is like the English term "wrong" in that it is standardly and conventionally used to express negative appraisals of actions."[299]

What roles exactly? Both Brink and Copp bet on survival and flourishing. Let us see below each of the respective formulations:

> Brink:
> He [moral realist] might claim that what is essential to moral properties is the causal role which they play in the characteristic activities of human organisms. In particular, the realist might claim that moral properties are those which bear upon the maintenance and flourishing of human organisms. Maintenance and flourishing presumably consist in necessary conditions for survival, other needs associated with basic well-being, wants of various sorts, and distinctively human capacities. People, actions, policies, states of affairs, etc. will bear good-making moral properties just insofar as they contribute to the satisfaction of these needs, want, and capacities (Brink, 1984, p. 121-2)

> Copp:

[299] Copp, 2007, p. 214

> My own society-centered moral theory suggests an account of the semantic intentions with which speakers use moral terms. The theory provides an account of the content and truth conditions of moral propositions. It identifies the property of rightness – that is, the property of being the right action in a context C – with the property of being required by the code of rules, whatever it is, the currency of which in S actually would best contribute to S´s ability to meet its needs – its needs, *inter alia*, for social stability, for peaceful cooperative interaction among its members, and for its members to be able to contribute to the overall flourishing of the society (2007, p. 237)

So, these formulations in terms of referential intentions are a great help to guarantee some initial and sufficient objectivity of reference determination and suggest that MTE may not be effective against naturalistic moral realism. That is the case if inhabitants of both worlds have the intention to engage in the a posteriori investigation of what natural properties clusters allow for inferences to evaluative judgments closer to the truth. Note however that there is no indication that Brink and Copp build into those formulations the putative necessarily motivational forces or overridingness of moral considerations. It seems to me that all that Copp and Brink can say in this sense is that people's intentions are to denote facts that when believed by people, can possibly make their beliefs in these facts generate related attitudes.

4.8. Is Convergence Necessary to Close the Question?

Convergence is not necessary to close the question. True, the moral realist treatment of MTE invites the anti-moral realist argument from disagreement (AFD), since it seems that rigid designation requires convergence between Oscar 1 and Oscar 2 about which moral theory prevails. Naturalist moral realism claims that moral properties are identified with, or are caused by, or supervene upon, natural properties. The relations of identification, causal inference, and supervenience are very strong, given that in principle they are relations of necessity. That means that whenever a particular natural property N obtains, its correspondent moral property M necessarily obtains too, and vice versa. Hence, the argument claims that if moral realism is true, then fully informed rational individuals, would identify the same moral properties with the same

natural properties, or would infer the same moral properties from the same natural properties. Exactly like two equally informed and rational chemists would converge on the belief that the liquid they see is water, if they both agree that the liquid is H_2O. John Rawls writes about such expectation:

> "It is natural to suppose that a necessary condition for objective moral truths is that there be sufficient agreement between the moral conceptions affirmed in wide reflective equilibrium, a state reached when people's moral convictions satisfy certain conditions of rationality."[300]

But, AFD goes, if such convergence does not obtain, even under ideal conditions of rationality, then it must be due to the fact that moral terms do not designate rigidly. If they did, fully informed rational individuals would eventually converge[301] via some version of reflective equilibrium. However, empirical evidence seems to be that fully rational individuals often agree about all nonmoral properties and facts of a given case, and yet identify it with different moral properties or infer different moral facts from it. Some philosophers thus have argued that, in the absence of a convergence theory, moral realism is false:

> "…it is of course evident that moral disagreements can be and frequently are attributable to confused thinking, factual ignorance and sheer prejudice. But the [realist's] obligation …is to show that such deficiency *has* to be involved in the generation of any such dispute. Any student of morality who has come to feel, therefore, that a substantial body of the principles that inform our ordinary moral thought are essentially contestable, and that no rational or cognitive deficiency is needed to sustain the clashes on things like sexual morality, the value of individual freedom, the moral status of animals and the ethics of suicide and mercy-killing, which are freely exemplified within and across cultures, won't give much for the realist's chances."[302]

Obviously, non-ideal widespread convergence is daydreaming, so the question here is whether *ideal* convergence is necessary for moral realism. I argue that moral realism does not require convergence. Indeed, even the total obliteration of nonmoral ignorance and irrationality

---

[300] Rawls, Collected Papers, p. 290
[301] Identify same M with same N or infer same M from same N.
[302] Crispin Wright, p. 199 Saving the Difference

would not entail moral convergence. But the absence of convergence does not undermine moral realism. Let us see in more depth the reasons to believe moral realism requires convergence.

The first reason we need to clear is the putative a priori character of morality. If moral identifications and moral inferences can be made merely by aprioristic rational methods, then in the absence of any defects in process and material, any rational individual will make the same identifications and the same inferences. Sure, in past discussions along this dissertation, I have defended morality as a synthetic enquiry, but there is more to be said about the problems of aprioristic convergence.

It is sometimes assumed that the whole truth about an a priori domain can be fully known by any impeccably rational creature. But the fact that a priori truths are known a priori does not entail that rational minds will know all the truths there are. In other words, it is not true that rational creatures necessarily know a priori everything there is to be known a priori. That would be to assume that rational creatures are necessarily *omniscient* about the a priori domain of truths. However, rationality does not entail omniscience about the a priori.[303] Justified false beliefs and other classic failures of rationality are the proof. Given that rational creatures are not omniscient it is possible that they arrive at incompatible moral views, even if morality is a priori. So, the fact that impeccably rational creatures do not grasp the moral truth, that by itself does not prevent the moral truth to exist *per se*.

A second reason for believing moral realism requires convergence is the view that the moral realist is committed to truth as accurate representation. The aim of moral judgment is to accurately represent an objective state of affairs. The convergence assumption then is that two equally rational creatures, equally geographically situated, should represent a given objective state

[303] McGrath, S. Moral Realism without Convergence, *Philosophical Topics,* vol. 38, n.2 2010, pp. 59-90

of affairs exactly alike, or something is wrong with process, material, or both.[304] Either both represent equally, or one is defective. But again, unless we also assume both creatures to be exactly equally omniscient, whose convergence is utterly trivial, realism does not get undermined by the lack of omniscient convergence. If we drop the assumption that the notion of fully rational includes omniscience, then a number of extrinsic factors may lead two equally situated rational creatures to arrive at different moral representations of a same extension. There may be moral facts too distantly out of range or beyond rational discrimination. The only representational convergence a realist is committed is the utterly trivial convergence among omniscient rational individuals. Nontrivial convergence is not necessarily entailed by the truth of any domain. As some philosophers[305] have argued, even perfectly rational thinkers may not be in position to know the entire truth within a domain.

Does taking seriously that there may be unknowable or yet unknown moral truths undermine moral realism? Only if one was required to know all of the moral truths there are in order to make one particular true moral statement. But clearly that is implausible. We can make true statements in any and every domain of knowledge without first having to know all and every truth there is in that domain.

The last reason to believe moral realism requires convergence is to conceive of the objectivity of moral truth *in terms of* convergence. In other words, moral truths just are the propositions all fully rational individuals in equilibrium converge on. Says Michael Smith:

> "The claim I defend is that normative reasons (…) are objective in the sense that, via a conversational process involving rational reflection and argument, we are each able to come up with an answer to the question,

[304] Says Wright: "…that representationally functioning systems, targeted on the same subject matter, can produce divergent output only if working on divergent input of if they function less than perfect."
[305] Williamson, T. *Knowledge and Its Limits*, OUP, 200; McGrath, S. 2010; Arpaly, Nomy. *Unprincipled Virtue* Wedgwood, Suikkanen, Jussi. "Williamson and (Moral) Realism on PEA Soup http://peasoup.us/2008/03/williamson-and/ March 27, 2008 and "Knowability of Moral Truths Again" on PEA Soup http://peasoup.us/2008/05/knowability-of/ May 19, 2008. Both accessed Feb 02, 2020.

"What do we have normative reason to do if we are in such and such circumstances?" and our answers to this question, provided we have each reflected properly, will all be one and the same. **We will all converge on an answer** of the form "It is desirable that we do so-and-so in such-and-such circumstances."[306]

"A careful mustering and assessment of the reasons for and against our particular moral opinions about such dilemmas and issues is therefore the best way to discover what the moral facts really are. If we are open-minded and thinking clearly then such an argument should result in a convergence in moral opinion, **a convergence upon the *truth***."[307]

"The rationalist thinks that the existence of reasons presupposes that under conditions of full rationality we would all have the same desires about what we are to do in the various circumstances we might face; **that absent such convergence we should say that there are no reasons at all**."[308]

That is the strongest necessity we could establish between our knowing truths and the truths there are. In this case of morality, the only truth(s) there is(are) *is/are* the moral proposition(s) on which all rational individuals would converge via reflective equilibrium. Not only fully rational individuals would eventually or often converge on the truth in the moral domain, rather, Smith argues moral facts *are* that what fully rational individuals would converge on. And, if fully rational thinkers would not all converge on a putative moral claim, then it follows that such claim is not a moral claim at all.

But this view builds reasons and convergence way too close.[309] It excludes the possibility of there being moral truths beyond intersubjective rational convergence. Since it is plausible that fully rational individuals are not omniscient not infallible, there might be parts of axiological reality not yet accessed by rational individuals or that are forever beyond rational understanding. Intersubjective convergence is not the same as convergence on all the truths there are to be known, especially natural truths.

The only way I can imagine for Smith to maintain such strong convergence theory is to build omniscience and infallibility into full rationality. Omniscient rational individuals would

---

[306] Smith, *Ethics and The A Priori,* p. 262
[307] Smith, *Moral Problem* p. 5
[308] *The Moral Problem*, p. 198 and 164-77
[309] Copp, "Belief, Reason, and Motivation: Michael Smith's "The Moral Problem" *Ethics* 108-1, 1997 p. 44-5

indeed not fail to know all the truths there are. So, the view is that omniscient infallible rational individuals will know all the truths, because they will converge on all the truths. But doesn't that seem too much to require from rational creatures, even *fully* rational ones? Smith's view also is instructive for its vulnerability to MTE.

Smith's conceptual rationalism claims that moral reasons are the reasons which every fully rational individual would converge on via reflective equilibrium. But as Smith admits, it is possible that there may not be such convergence in which case there would be no moral requirements at all. Thus, substantive rationalism is the view that *there are* such moral reasons. Horgan and Timmons argue that if conceptual rationalism is true, then substantive rationalism is false, given that there is no such a thing as substantive perfect convergence. They show that substantive rationalism is false by applying MTE.

So, take substantive rationalism: There are moral reasons only if *all* fully rational agents would converge in their desires about what to do in those situations calling for a moral response.[310] Now ask the question: Would fully rational individuals in reflective equilibrium *all* converge on the same moral judgment about Φ in same circumstances C? Well, suppose the natural property possessed by the action on which all fully rational individuals converge is captured by consequentialist moral theory Tc. Next, imagine Twin Earth where all fully rational individuals converge on a deontological moral theory Td. Then suppose the two groups encounter, learn about each other's moral convictions and remain in divergence. Neither group is moved to change their moral paradigm. Imagine that one of the reasons the groups diverge is due to their different emotional sensibilities, such as experiencing sympathy and guilt differently. Since this resilient[311]

[310] These reasons are not to be found "in the nature of man nor in the circumstances of the world in which he is placed, but solely a priori in the concepts of pure reason." Kant, *Groundwork*

[311] Resilient moral disagreement illustrated: Horgan and Timmons say "there is no reason to think" that the divergence would be overcome if Earthlings and TwinEarthlings sought reflective equilibrium. Instead, the two groups have each

divergence between fully rational individuals is conceptually coherent, and a genuine metaphysical possibility, MTE shows that "Smith's substantive rationalism is false: it is not the case that all metaphysically possible moral judgers would converge on the natural property they would want acts to have were they fully rational."[312] Moreover, Horgan and Timmons think that it is "wildly implausible to insist that at least one of the groups must not have satisfied the requirements of being fully rational."[313]

But note that for moral realism, it is not necessary to defend that one of the groups is not fully rational. First because moral realism does not depend on full convergence. Moral realists will either make approximations to the moral truth or enlarge their knowledge of moral truths having referential intentions as starting point for future reflective equilibrium, if any ever comes to fruition. Second, the externalist features of moral realism, motivational and rational wise, asserts it to be perfectly possible that two fully rational individuals disagree about what ought to be done in circumstances C. The externalist about moral motivation and moral reasons, may accept that an all-things-considered normative reason is the reason on which all fully rational individuals would converge in reflective equilibrium. What he does not accept is that all-things-considered reasons is always *necessarily* a *moral* reason. But the latter is consistent with Smith's substantive rationalism, which I discussed in chapter 2.

---

reached their own reflective equilibrium. Illustrated by the divergence between Putnam and Nozick on welfare versus individual liberty as grounding of government spending. Given the level of philosophical sophistication and moral sensibility, it is very unlikely that any movement in the direction of yet greater unity and coherence will bring the two together on this issue. Putnam, p. 164 *Reason, Truth, and History*

[312] Horgan and Timmons. "Troubles for Michael Smith's Metaethical Rationalism". *Philosophical Papers*. 25-3, 1996

[313] P. 211

4.9. Vagueness and Unknown Reference

It may be argued that since there is no convergence on which moral theory fully rational individuals employ to pick out natural properties, there will be vagueness about which of those properties are to be picked out. And since an ultimate convergence is unlikely, it is possible that some evaluations of some actions, characters, etc. will remain unknown. But, the ability of causal theory to allow for some degree of unfamiliarity with referents is one of its virtues. In a descriptive theory, the referent of a term is any object that satisfies the description that the speakers associate with the terms. Thus, if a speaker does not associate the correct description with the term, the term will not denote as expected. On the other hand, Kripke considers several ways in which the speakers are ignorant or hold false beliefs about objects, which they nonetheless denote successfully.

Let us suppose that the description associated with the name "Gödel" is "the person who has proved the incompleteness theorem". Then let us consider that a man named Schmidt, not Gödel, proved the incompleteness of arithmetic. Gödel, somehow, was associated with a work that was falsely attributed to him. According to Kripkean semantics, if that was the case, the term "Gödel" would denote Gödel and not Schmidt, even though it is actually Schmidt who satisfies the description "the person who has proved...", a description we mistakenly associate with the term "Gödel." In descriptivist theory, the term "Gödel" would denote Schmidt and not Gödel, for it is Schmidt who satisfies the relevant description. The causal theory explains how the term "Gödel" would still denote Gödel, despite the wide neglect of Gödel's intellectual accomplishments, and even if most of what we believe about Gödel is false. This is because even though we do not have the correct descriptions of the extension, *there is a sharing of the intention to talk about Gödel*.

In the same way this explanation can relate the moral terms to the general and abstract descriptions to which we arrive from our referential intentions. In this way, referential intentions, along with causal contact with portions of events that contain that abstract property, transmit the reference successfully without our being part of or being positioned in the correct causal chain.

4.10. Horgan and Timmons' Critique of Copp

Copp's response to MTE has two parts. In the first, Copp accuses MTE of presupposing that the difference in reference necessarily leads to a difference in meaning. In the second part, Copp seeks to show that naturalists can tell a story of how the inhabitants of the Earth and the Twin Earth may refer to the same things. Timmons and Horgan argue, against the first part, that MTE does not assume that the difference in reference entails a difference of meaning, but rather a difference of attitudes would be responsible for the difference of meaning in the moral case.

Copp emphasizes that there are attributes general enough in the functional role of the moral terms in both worlds to plausibly assume that there are common characteristics in the referential intention between the two groups. But like I said before, if referential intentions are employed according to a multiplicity of substantial moral theories, we will have difficulty explaining how the intentions of speakers from different moral communities could point out to a single moral property. Here we return to the problem of establishing the content of referential intentions since we admit that they play an indispensable role in setting the reference.

The problem spotted by Timmons and Horgan is that offhand talking about flourishing and impartiality is vague. This is because different and incompatible theories are equally compatible with the generic notions of flourishing and impartiality. A serious problem with impartiality is that it lacks sufficient determinations and cannot serve to set the reference of moral terms. Flourishing suffers from the same problem. As Timmons and Horgan state:

> On both Earth and Moral Twin Earth there are too many natural properties available to serve as referents of moral terms like 'good' and 'right' – specifically, the functional property definable as the 'good'-role property by a consequentialist theory *Tc*, and the functional property definable as the 'good'-role property by a deontological theory *Td*. Both properties are compatible with generic ideas of flourishing and impartiality. It looks like appeal to referential intentions that the two groups share, involving notions like flourishing and impartiality, is too weak to pin down the referents of moral terms; moral indeterminacy results.[314]

In view of these considerations, Copp assumes that he is not entirely satisfied with his own view, since pure causal theories of semantics would prefer to explain the determination of reference independently of psychology. In general, what Copp suggests is that the introduction of the notion of referential intentions to show that the causal theory of reference does not lead to moral relativism entails a problem of its own within the philosophy of mind. In order to explain how we have cognitive access to referential intentions attributed to other people we need abductive inferences about the content of these intentions. Critics of semantic theories that employ the notion of referential intentions say that if these theories cannot explain this cognitive access, then they too cannot describe the content of intentions. Therefore, since intentional theories argue that intentions are essential for the fixation and mutual recognition of the reference, critics accuse it of making it impossible for listeners to fix and recognize the reference of words uttered by the speaker because they do not have cognitive access to the mental states intrinsic to interlocutors. But this objection isn't right. We all face epistemic difficulties in determining what any particular person means for a number of reasons. The interlocutor may by lying, for instance. But this does not entail that the theory itself cannot describe, attribute, and predict the content of intentions or the meaning of terms given those intentions. They will rely on nomological, lawlike mechanisms about the formation of intentions or associative relations to determine those contents even if particular cases happen to be hard ones.

[314] Timmons e Horgan, 2000, p. 145

Notwithstanding, I do not think this problem constitutes an insurmountable obstacle to the moral semantics proposed by Brink and Copp. My position is based on Borg's work.[315] If she is correct, the non-inferential theories created to explain our cognitive access to the referential intentions attributed to other people show us that the criticism of Timmons and Horgan is not as strong as it seems at first.

Back to my reasoning, we have seen that the notion of referential intention introduces an additional problem. To know what are the referential intentions of a speaker we need to make inferences about the mental contents of the speaker, merely based on the observation of his behavior. This means that a semantic theory based on the notion of referential intentions requires "enriched contexts"[316]. In other words, in order to understand what the speaker's words denote successfully we need to know not only the physical and social context of the speaker, but also to "know" abductively what are the mental states of the speaker that underlie his linguistic intentions.

However, Borg argues that we can avoid enriched context objection by narrowing of the gap between the directly accessible aspects in a context of utterance and the seemingly implicit intentional states of the speaker in this context. In this way, according to Borg, the referential intentions would become accessible without the need for inferential enrichment heavy-duty work by the listener or observer. This type of perspective is known as a minimalist conception about the attribution of referential intentions. According to these theories, cognitive access to the semantic content of a referential intention does not require direct access into the speaker's mind. She proposes that the observation of the speaker's behavior sufficiently reveals that content. This type of conception can be captured in a theory called "non-inferentialism" according to which:

---

[315] Emma Borg. 'The place of referential intentions in linguistic content' (2009), *Manuscrito: CLE: Unicamp* 32: 85-122. Special edition on semantics/pragmatics.

[316] Enriched context means all the necessary knowledge one would have to acquire about another person before one could infer that person's referential intentions.

> The openness to the states of others exhausts our methods for assigning (certain) mental states to them. Thus, there is no inferential step to be taken between seeing A behaving in manner p and assigning to A the mental state m associated with p – we simply *see* another´s mental state in their behavior. (BORG, 2009, p. 100)

Borg suggests that non-inferentialism allows non-inferential (direct) access to the speaker's referential intentions. There are several non-inferential ways of explaining our access to the referential intentions: *simulation theory, mirror neuron hypothesis, theory-theory mindreading* and *bodyreading.*[317] For instance, in simulation theory, we attribute mental states to other people through an empathic process where we place ourselves in the "others' shoes." Thus, we use our own intentional mechanisms in a process in which we "pretend" to be the other person to conjecture what the reasoning would be if we upheld the same beliefs and desires of this person.

However, *simulation theory* could be criticized for not completely rejecting the need for abductive inference for access to other people's mental states. Minimalist non-inferential theories actually propose that we can *see* the referential intentions expressed directly in people's behavior. Some may argue that the *simulation theory* is not good, because when it requires a fancy process of imagination, it reestablishes the need to enrich the context through abductive inference, making the process again indirect. But this is just a safe net argument I am offering for cases of extreme ignorance of someone's referential intentions. And most of the cases individuals are in fairly sufficient conditions to know each other's referential intentions.

Mirror neuron hypothesis, on the other hand, does not require this kind of complex reasoning. According to this theory, when we observe a person's behavior, our brain recreates/mirrors the same neuronal process that led that person to behave that way. The *theory-theory mindreading* and *bodyreading* attest that we understand the actions of other people by classifying these actions according to psychological laws (for example, "if A wants x and believes

[317] 2009, p. 101

that realizing y is a way to reach x, then A will perform y"). These two perspectives have been extensively investigated through experiments with children in language development up to four years. These studies show that children develop a theory of assigning primitive referential intentions about, for example, what their nannies are willing to communicate when they point to some object.

Of course, since all of these options are based on the attribution of referential intentions through observation of behavior without the use of abductive inferences (as proposed by non-inferential minimalist semantic theories), some similarities in the gesture and behavioral movements of the speaker may lead to mistaken assignments of intentions to speakers. Could this be evidence that referential intentions are not fully expressed by behavior and thus destabilize minimalism? Borg recognizes this question and cites as an example of this problem the statement of Baldwin and Baird:

> The surface flow of motion people produce in most, if not all, cases is consistent with a multitude of different intentions. Thus when observing others in action, we rely on other sources of information – knowledge about human behaviour in general, specific knowledge about the particular individual involved, knowledge about the situation – to help to disambiguate which among the many candidate intentions is relevant in any given case…The upshot is that discerning intentions is a complex enterprise; it is knowledge driven as well as rooted in structure detection (Baldwin and Baird 2001: 175-6 *apud* Borg, 2009, p. 112)

This problem seems to show us that at different levels all theories exemplified so far require a context enriched by abductive inferences about the mental states that underlie the referential intentions. However, the function of referential intentions cannot be discarded from the process of identifying the meaning of sentences and therefore the abductive reasoning must be accommodated by minimalist semantic theories avoiding the problem posed by non-inferentialism: speaker intentions "*do* play a part in fixing a reference for a demonstrative *and* they are recoverable only via rich, abductive means, but features".

This could be done through the assumption that referential intentions play a role in determining reference when they are used to denote items, properties, or events from which we may draw a generic and abstract idea that can be offered in functional or descriptive terms. This proposal is found in the philosophy of language and resembles to a large extent the application of referential intentions in the moral case by Brink and Copp, though they do not mention works in this area.

4.11. **Chapter Conclusion**

We have seen that tradition in metaethics presents us with the open question argument as the first argument of semantic nature in favor of anti-realist or non-cognitivist theories that Moore employed against analytic naturalistic realism. Boyd then attempted to answer the argument based on Kripke's synthetic naturalist semantic theory. However, Horgan and Timmons rephrased Moore's argument according to Putnam's Twin Earth thought experiment applied to the moral case to refute Boyd's synthetic moral naturalism. Next, Brink and Copp resort to a concept found in Putnam's own naturalistic semantics, namely, the notion of referential intentions, to respond to the Moral Twin Earth argument. Finally, we have seen that this strategy maintains the determination of the moral reference safe from arguments such as the Moral Twin Earth.

References

Alvin Goldman, “What is Justified Belief” In George Pappas (ed.), *Justification and Knowledge*. Boston: D. Reidel. pp. 1-25 (1979

Brandt, R. *A Theory of The Good and The Right*. OUP, 1979

Boyd, R. How to be a moral realist. In: Sayre-McCord (ed.) *Essays on moral realism.* Ithaca, NY: Cornell University Press, p.181-228. 1988

Copp, D. Moral Realism: Facts and Norms. *Ethics* v. 101, 1991

Emma Borg. ‘The place of referential intentions in linguistic content’ (2009), *Manuscrito: CLE: Unicamp* 32: 85-122. Special edition on semantics/pragmatics.

Frankena, W.K. The Naturalistic Fallacy. *Mind*. v.48 n. 192. 1939, p.471

Horgan and Timmons. “Troubles for Michael Smith’s Metaethical Rationalism”. *Philosophical Papers*. 25-3, 1996

Fredrik Björklund, Gunnar Björnsson, John Eriksson, Ragnar Francen Olinder and Caj Strandberg. “Recent Work on Motivational Internalism” in *Analysis* vol. 72 No.1, January 2012

Horgan, T. and Timmons, M. Troubles for Michael Smith’s Metaethical Rationalism. *Philosophical Papers*, Vol. 25, n. 3, 1996

Horgan, T. and Timmons, M. Analytical Moral Functionalism Meets Moral Twin Earth. In Ian Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*. Oxford University Press, 2009.

Horgan, T. and Timmons, M. Troubles for New Wave Semantics: The Open Question Argument Revived. *Philosophical Papers*, Vol. 21, n. 3, 1992

Langford, C.H. The Notion of Analysis in Moore’s Philosophy. In Schilpp, P.A. *The Philosophy of G.E. Moore*, Northwestern University Press, 1942

Lewy, Casimir. G.E. Moore on the Naturalistic Fallacy. *Proceedings of the British Academy*. V. 50, 1964, pp.251-262

Putnam, H. The Meaning of “Meaning”. In: *Mind, Language and Reality*, Cambridge University Press, 1975

Thomas Nagel, *The Possibility of Altruism*, Princeton, 1970

McGrath, S. Moral Realism without Convergence, *Philosophical Topics,* vol. 38, n.2 2010, pp. 59-90

Copp, “Belief, Reason, and Motivation: Michael Smith's "The Moral Problem" *Ethics* 108-1, 1997 p. 44-5

Christine Korsgaard, *The Sources of Normativity*, Cambridge, 1996

John Mackie, *Inveting Right and Wrong*, Penguin, 1977

Tresan, Jon. Metaethical Internalism: Another Neglected Distinction, *The Journal of Ethics*, 2009

Fletcher and Ridge, *Having It Both Ways* 2014 Tresan, J, chapter 5 *Diachronic Hybrid Moral Realism*

James Dreier, Dispositions and Fetishes: Externalist Models of Moral Motivation, *Philosophy and Phenomenological Research*, Vol. 61, No. 3 (Nov., 2000), pp. 619-638

Shafer-Landau, *Moral Realism: a defence*, part III chaps 5 and 6. and “A defense of Motivational Externalism”, *Philosophical Studies*, 2000

David Brink, “Moral Motivation” *Ethics* 107 (1997): 4-32

Brink, D. Externalism Moral Realism. *Southern Journal of Philosophy*, v.24, 1986

David Brink, *Moral Realism and The Foundation of Ethics,* CUP, 1989

David Brink, “The Significance of Desire”, *Oxford Studies in Metaethics,* 3 (2008): 5-46.

David Brink, “Moral Conflict and Its Structure”, *The Philosophical Review*, Vol. 103, No. 2 (Apr., 1994), pp. 215-247

Zangwill, Nick. “Motivational Externalism: Formulation, Methodology, Rationality, and Indifference.” In Gunnar Björnsson, Fredrik Björklund, Caj Strandberg, John Eriksson, Ragnar Francén Olinder (editors). *Motivational Internalism.* Oxford University Press, 2015

Sigrun Svavarsdottir, “Moral Cognitivism and Motivation”, *The Philosophical Review*, V. 108, n.2 1999.

James Dreier, “Ecumenical Expressivism, Finessing Frege”, *Ethics* 116 (January 2006): 302–336

Zangwill, Nick. “Motivational Externalism: Formulation, Methodology, Rationality, and Indifference.” In Gunnar Björnsson, Fredrik Björklund, Caj Strandberg, John Eriksson, Ragnar Francén Olinder (editors). *Motivational Internalism.* Oxford University Press, 2015

Fredrik Björklund, Gunnar Björnsson, John Eriksson, Ragnar Francen Olinder and Caj Strandberg. "Recent Work on Motivational Internalism" in *Analysis* vol. 72 No.1, January 2012

Michael Stocker pg. 744 (1979) "Desiring the Bad: An Essay in Moral Psychology," *Journal of Philosophy*
76(12): 738-53

Kristján Kristjánsson *Virtuous Emotions*, OUP, 2019

Daniel Eggers "Moral Motivation in Early 18th Century Moral Rationalism", *European Journal of Philosophy*, 2019

Alfred Mele, pg. 733-4 (1996) "Internalist Moral Cognitivism and Listlessness," *Ethics* 106: 727-53.

Dale Dorsey, "Subjectivism Without Desire", *The Philosophical Review*, Vol. 121, No. 3 (2012), pp. 407-442

Dale Dorsey, *Limits of Moral Authority*, OUP, 2016

Steven Swartzer op cit and "A Challenge to Humean Externalism" *Philosophical Studies,* 2018

Wittgenstein. "Lecture on Ethics". *The Philosophical Review*, vol. 74, n.1, 1965

Michael Smith. *The moral problem*. Blackwell, 1994

Philip Pettit, "Humeans, Anti-Humeans, and Motivation" *Mind*, New Series, Vol. 96, No. 384 (Oct., 1987), pp. 530-533

Brad Inwood, *Ethics and Human Action in Early Stoicism*, OUP, 1985

Robert Louden, *Kant's Impure Ethics*, OUP, 2000

Michael Stocker, *Plural and Conflicting Values*, OUP, 1990

Brian Weatherson, *Normative Externalism*, OUP, 2019

.