

ABSTRACT

This dissertation exposes certain metaphysical commitments at the heart of liberal state neutrality theory and, in particular, it establishes the extent to which this theory assumes a certain a priori idea of human nature. It first explores the disagreements between neutralist authors (such as Ackerman, Nagel, Larmore, Barry, Nozick, Habermas, and Rawls), in order to show how in their disagreements there can be seen the acknowledgement that a too stringent understanding of neutrality is theoretically impossible or politically impractical. In this impossibility or impracticality we see the hole through which the material content of liberalism sneaks into the principle of neutrality. The argument moves on to explain that liberal neutralists reject metaphysics for political thinking because they see it as turning the state into an oppressive institution. From this, I argue that we need a metaphysical approach that highlights the idea of human nature that is posited a priori for the principle of neutrality to make sense or be operative. Three concepts that ground neutrality are selected to investigate the metaphysical content that they carry into this idea of human nature: intersubjectivity, reasonableness, and autonomy. Finally, the dissertation examines in detail the metaphysical content that these three concepts carry into the public sphere. From intersubjectivity (in Habermas), we obtain ideal preconditions of communication that form a universal structure of rationality. From reasonableness (in Rawls), we get the description of the moral powers that are presupposed in the idea of the person that underwrites liberalism. From autonomy, a metaphysical freedom proper of Christian theology and European idealism. These metaphysical commitments prevent alternative doctrines, political or metaphysical, to enter the public sphere. This way liberalism becomes oppressive on its own terms. By means of neutrality, liberalism normalizes and indeed trivializes a certain anthropology with metaphysical undertones, but, by the same token, it also normalizes other, non-metaphysical

positions that are in dispute in the background culture. In this way, national identities and cultures, chauvinist and sexist social hierarchies, and discriminations against minorities of all kinds can be secured in their purported supremacy by cloaking themselves in neutrality.

METAPHYSICS OF STATE NEUTRALITY

A Critique of Liberalism

by

Jordi Graupera Garcia-Milà

July 2017

Submitted to The New School for Social Research of The New School University
in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Comittee:

Dr. Omri Boehm, The New School for Social Research (advisor)

Dr. Alice Crary, The New School for Social Research

Dr. Simon Critchley, The New School for Social Research

Dr. Peter J. Diamond, New York University

To Sara: we made it.

ACKNOWLEDGEMENTS

This dissertation owes a great deal to the conversations I've had with people from The New School for Social Research. My classmates Philip Schauss and Matt Hackett have read, criticized, discussed and helped edit every word I submit here, and many more that died in the process. Their influence in my thinking cannot be overstated, like my feeling of gratitude. Juan Menchero, Pablo Bustinduy, Anna Matthiesen—all from The New School—and Blanca Serrano—from the Institute of Fine Arts at NYU—are fellow travelers in this quest to say something meaningful and clear. P.J. Gorre, Jonathan Chalier, Aaron Jaffe, Michael Becker, Cille Varslev: they have heard enough. What I learnt from Richard J. Bernstein, from his lectures on Kantian Pragmatism or on the question of religion in the public sphere, to the talks in his office, informed the first aim and set of questions of this project. Alice Crary's guidance questioned it and shaped it. The advice that made it into a reality came from Omri Boehm, whose involvement I cannot thank enough,—he made clear to me what I did not know I was saying—, and to whom for this and other reasons I dare to say that I am his disciple in many ways, including the way we disagree. Carles Boix and Alicia Adserà have been an inspiration all these years, and still are. My students at Parsons, NYU, and Saint Francis College contributed more than they know. Not only the readings and lectures I prepared for them taught me a lot and made it into my argument, albeit implicitly; but also, the necessity to make myself clear to them every week made me question many things, including some of the ideas I thought I knew well.

The wonders of our time have made it possible that friends from the Old World shared this burden too. They deserve their stake in the joy. Joan Lloveras heard about it while in Freiburg, read the first draft of the idea in China, and he claims he is going to read the final product now that he's back in Barcelona. Rubèn Guilanyà is writing this same dissertation from the other side of the mirror. Philosophers Cristian Palazzi and Aida C. Rodríguez, partners in crime, disciples of Begoña Roman and Josep Maria Via i Taltavull, as I am, share an *ethos*. From the days at the Philosophy Department at Universitat Ramon Llull, I must also mention professor Carles Llinàs: he first introduced me to the philosophical problem of neutrality and whispered the words “Koselleck, *Critique and Crisis*,” to which this dissertation sometimes is an echo. From the real world, Fran Segarra, Yara Carmona, Edu Mas, Alicia Garrido, Arnau Vidal, Noemí Duran, Joan Sanahuja: they are good at pretending they are interested every time. In my pocket, Mario Núñez, Marc Martí, Xavi Cortés, Manel Vidal, Arturo Puente, Carles Foguet, Albert Lloreta: they are the technical support. Writers Enric Vila and Bernat Dedéu, they are literally brothers in arms, and they care about the truth.

The workers of the bookstore-cafe *Topos*, in Ridgewood, Queens, and the workers of The New School and NYU have provided the material conditions for the actual writing. None of these conditions were taken for granted. I am thankful.

My parents are alive in this text, if nowhere else. My sisters, Mariona and Isabel, are everywhere, including this text. Francesc and Mercè, my *other* parents, are always there. My wife Sara sees where I don't look and listens to what I don't say: she opens the world with me. My daughter Aurèlia Ray also teaches me.

In spite of these contributions, all mistakes are mine.

TABLE OF CONTENTS

Acknowledgements	iii
------------------	-----

INTRODUCTION **An Awareness of What is Not Missing**

1. The Main Claim	1
2. Practical Consequences	5
3. Dissertation Outline	10

CHAPTER I **Some Consequences of an Impossibility**

1. The Principle of State neutrality	18
2. The Separate Sphere of the Political	21
3. Right vs Good	24
4. First Dichotomy: Relative or Absolute Neutrality	28
4.a Relative Neutrality	29
4.b Absolute Neutrality	36
4.c The Problem of Practicality	41
5. Second Dichotomy: Narrow or Comprehensive Neutrality	43
6. Third Dichotomy: Aim or Effect	47
6.a Aim	47

6.b Justification	50
6.c Effects	52
6.d Consequences of Disregarding Effects	54
7. An Alternative Narrative	58
8. Impossibilities and further debates	63

CHAPTER II

Metaphysical Carriers

.Introduction	70
1. What do they mean by metaphysical?	77
1.a. Post-metaphysics	77
1.b Comprehensive Views	80
1.c Metaphysical Commitments	84
2. Ideology Critiques and Critique of Metaphysics	90
3. Pragmatism and Social Reason	92
4. Kołakowski's and Rorty's external critique	98
5. World-views in Habermas	103
6. Rawls's method of avoidance	109
6.a. Universal Truth	109
6.b. Moral conceptions	113
7. Autonomy and moral powers	121

CHAPTER III

Intersubjectivity's Metaphysical Content in Habermas

1. Habermas's Intersubjectivity as Detranscendentalization of Reason	131
2. Validity Claims and Rawls Critique of Habermas's Comprehensiveness	139
3. Walzer's Critique of Habermas's Thickness	145
4. Gordon's Critique of Habermas's <i>Einverständnis</i>	155
5. Rorty's Critique of Habermas's 'better argument'	163
6. Metaphysical empathy (Michael Kelly and Seyla Benhabib)	169
7. An integrative perspective	179

CHAPTER IV

Reasonableness's Metaphysical Content in Rawls

1. From Comprehensive to Political	182
2. Exclusion and Inclusion	186
2.a Freestandingness as the Exclusion of Metaphysics	186
2.b Overlapping Consensus as Moral Inclusion	189
3. Reasonableness	193

3.a Indetermination and Self-containment	193
3.b The Anthropology of a Social <i>Ethos</i> and an Imperfect <i>Nous</i>	195
3.c Reasonability as Practical Reason	199
3.d The Essential Aspects of a Moral Person	201
4. The Original Position	205
4.a The Original Position Determines the Outcome	206
4.b The Original Position Is Not a Contract	209
4.c The Method of Reflective Equilibrium	211
4.d Dworkin and the Natural Right to Equal Respect	224
4.e Formal Comprehensive Views	230
4.f An-Na'im and Rawls's Theological Tradition	231
4.g Rawls's Answer	233
5. The Idea of Public Reason	237

CHAPTER V

The Co-dependence of neutrality and Autonomy

Introduction	249
1. neutrality and Autonomy in the Political Domain	252
2. Modern Autonomy: Kant, Levinas, and Post-metaphysics	257
2.a Henrich and Korsgaard on Kant's Pure Reason	257
2.b Kant's noumenal Causality	260
2.c Emmanuel Levinas: Autonomy and the Theology of Freedom	266
2.d The Post-metaphysical Project of Autonomy	275
3. Autonomy versus neutrality	281
3.a Raz's Perfectionism: Autonomy over neutrality	281
3.b Raz's Critique of Rawls	285
3.c Barry's Neutralism: neutrality over Autonomy	287
4. Rawls and Larmore: Non-metaphysical Autonomy	292
4.a Rawls's Political Domain	292
4.b <i>Citoyen</i> versus <i>Homme</i>	294
4.c Larmore's 'modus vivendi' liberalism	295
4.d The Limits of the Strategy	299
5. Nagel's Paradox	305
5.a Higher Order Impartiality	305
5.b Autonomy as Belief	308
5.c Consequences of the Paradox	312
6. Markus Gabriel: Liberal Naturalism and the Defeat of Freedom	316
6.1 Metaphysical Quietism	316
6.2 The Defeat of Metaphysics	318

CONCLUSIONS	321
WORKS CITED	327

Introduction

An Awareness of What Is Not Missing

“Beth Hamishpath” —the House of Justice: these words shouted by the court usher at the top of his voice make us jump to our feet as they announce the arrival of the three judges, who, bareheaded, in black robes, walk into the courtroom from side entrance to take their seats on the highest tier of the raised platform.

First sentence of Hanna Arendt’s *Eichmann in Jerusalem*.¹

1. The main claim

This dissertation will be successful if it manages to expose certain metaphysical commitments at the heart of liberal state neutrality theory and, in particular, if it establishes the extent to which this theory assumes a certain a priori idea of human nature.

The idea that the state should remain neutral among competing “conceptions of the good”² and that in their regular business, state officials and representatives should not act or produce legislation out of deep-seated metaphysical or religious beliefs —so far as possible, as Rawls says³

¹ Arendt, Hannah. *Eichmann in Jerusalem: a report on the banality of evil*. New York, NY, U.S.A.: Penguin Books, 2006, 3.

² This is a common expression in the literature, as we will see. Also: “comprehensive view of the good” or “metaphysical doctrines.”

³ “The idea is that in a constitutional democracy the public conception of justice should be, so far as possible, independent of controversial philosophical and religious doctrines.” John Rawls, “Justice as Fairness: Political not Metaphysical”; *Philosophy and Public Affairs*, Vol. 14, No. 3. (Summer, 1985), 223

— is at the core of the Western Liberal state, both historically and analytically.⁴ Nevertheless, I submit that all forms of contemporary political liberalism that implicitly or explicitly embrace some idea of neutrality as a characteristic of the state —or any other institution⁵— are thereby committed to a comprehensive and unifying conception of man and the good. As I maintain throughout, these are metaphysical commitments.

When the tension between the private world of self-creation and the public world of solidarity — to use Rorty’s terminology⁶— is resolved in practice or argument with the help of neutrality, does it give rise to an anthropology? When the vocabulary of liberalism is actualized through neutrality, what kind of meaning does it convey? Does this meaning refer ultimately to a meaning of human nature? If we answer affirmatively to all these questions, and take contingency seriously, then by means of its grammar and metaphoric production —and the institutions they inform—, liberalism is *creating* a human nature. I defend that this is seen through the lens of their endorsement of the idea of neutrality in politics and the vocabulary that crops out of it. The main purpose of my investigation is not whether this is right or wrong, but rather, to show that by claiming not to have metaphysical commitments, liberalism manages to smuggle a metaphysical paradigm into the system. Subordinate to this purpose, however, I offer a further critique, namely: if it is true that liberalism is fostering a particular understanding of human nature, then

⁴ The first explicit reference of political neutrality in Western international law is the Treaty of Augsburg in 1555, where the Emperor of the Holy Roman Empire declared himself to be neutral among Catholic and Lutheran princes. For a historical-philosophical account of neutrality and the historical background whence it emerged, see Reinhart Koselleck, *Critique and Crisis: Enlightenment and the Pathogenesis of Modern Society* (Cambridge, MA: MIT Press, 2015), and Paolo Prodi, *Una historia de la justicia: de la pluralidad de fueros al dualismo modern entre consciencia y derecho*. (Madrid, Spain: Katz, 2008).

⁵ ‘Institution’, in a broad sense; i.e., from principles of ideology to particular public actions and customs.

⁶ Rorty does not consider himself to be a neutralist liberal.

it is oppressive in the same sense that it says that illiberal, metaphysically committed, ideologies are.

To keep this introduction concise, I will explain here the main argument along Kantian lines. The principle of state neutrality is a political principle that articulates the way in which the state must rule over its citizens given the fact of plurality in social life. There is a plurality of beliefs among citizens regarding the best way to conduct one's life and the best kind of relationship between people. neutrality aims at ensuring that the state treats these beliefs and these relationships fairly, where "fairly" means that the laws and officials that shape the state cannot make any appeal to external standards of justice or rationality. In this context, an external standard of justice would be *transcendent*, demanding either a kind of revelation, as found in most if not all religious doctrines, or a metaphysical theory of the Good, such as the Platonic Idea of the Good. These are metaphysical views that remain in dispute in pluralist societies. To settle these disputes, the state cannot appeal to an understanding of rationality either; it cannot say that such and such views are irrational because one particular view is *the* rational one. This is so, in part, because what counts as rational is also hotly debated, and this dispute ultimately boils down to conflicting metaphysical positions. But, most importantly, because contemporary liberalism belongs to the tradition of modern philosophy that has accepted that reason cannot settle metaphysical questions, including the nature of rationality itself, as opposed to what science is capable of doing with empirical questions. The most it can do is to describe the limits of the use of reason. In the same way that "the unconditioned" (Kant's word for the "absolute") exceeds the limits of theoretical reason, the liberal state is also constrained by the multiplicity of metaphysical positions. There is doubtless a connection between the empirical fact of plurality in metaphysical

views and the inevitably antinomy of reason that Kant exposes in transcendental reflection. This means that the state cannot appeal to a strong notion of truth nor to the intrinsic superiority of a certain idea of justice in order to decide which laws and policies it produces.

The only possible alternative to dogmatic metaphysics, it seems, is democratic consensus. The state must be anchored in what its citizens agree is the most just way of dealing with public affairs. This is the reason why neutralist authors require an account of intersubjective communication in two senses: first, an account of what constitutes a genuine consensus as opposed to imposition or indoctrination –i.e., that reasons are exchanged and negotiated according to rules that treat them as equally valid as long as they are held to be open to criticism; and second, an account of consensus that excludes those positions, which, even if they were agreed upon, would trespass the limits that neutrality has set for the state. As for the former, my claim is that the idea of intersubjective consensus that neutrality demands harbors metaphysical commitments when it defines the kind of rationality that is at play in these exchanges of reasons. The latter, on the other hand, is achieved by means of the notion of reasonability: reasonable views are, in large part, those that accept the limits imposed by neutrality and are subject to modification. My claim is that these limits are related to an idea of human nature stemming from the modern understanding of the subject. This idea of human nature, I further claim, is an expression of those same metaphysical commitments.

However, in the same way that those theoretical limits that excluded metaphysics from the domain of science also excluded metaphysical positions from the domain of the state, so that the main understanding of the principle of state neutrality describes it as a principle of restraint, the

other side of the coin, namely practical reason, affects the very thing it seeks to protect. What neutrality is meant to protect is freedom.

Because the way in which our beliefs shape our decisions and actions is understood as the expression of freedom, the state must protect that self-determination, insofar as it does not infringe on the freedoms of other subjects and abides by the limits that are established through inter-subjective consensus and the principle of reasonability. In other words, the neutral state understands citizens as possessing a capacity for moral autonomy, even if the actual subject that lies behind the political notion of “citizen” does not believe that we are actually autonomous in that sense. Since neutralist authors are aware that a post-Kantian understanding of autonomy, and the duty to protect it, can be seen as a metaphysical position or as the protection of a conception of the good, they try to offer an account of autonomy that in no way relies on metaphysical formulations. My final claim in this dissertation is that they fail at this attempt. In other words: I will try to show, first, that intersubjectivity, reasonability, and autonomy are the three fundamental notions that neutralist authors need in order to meet the challenges that the principle of neutrality poses; second, that these three notions carry the metaphysical weight of neutrality; and, third, the particular content that they harbor.

2. Practical consequences

Why all this is of importance can be illustrated in three instances: a) contemporary liberalism is at the heart of most Western states’ self-definition and it is so strong a force that even those who reject liberalism like to cloak themselves in its rhetoric so as to look better. It is easy to see already

in the critiques that liberal neutralists level at the every-day practice of liberal states that in reality liberalism falls short of its promise which is b) to not be attached to any “comprehensive view of the good.” This promise is due to the fact that c) historically, the identity between a comprehensive metaphysical perspective and the business of the state resulted in both subtle and not so subtle forms of violence and oppression. As stated, my claim is that the failure to meet its own standards is not due to the imperfection in the application of the theory/promise —as is sometimes claimed for all sorts of ideologies of the state— but, rather, to a failure already evident in the theory itself, which, in this argument I am submitting, might turn out to be as committed to old metaphysical paradigms as the practice of the states themselves seem to be.

In practical terms, however, the aim of neutrality is not only to secure freedom, but also to ensure stability. As a tool, the idea of political neutrality serves two ends, the tension between which it is meant to resolve: security v. liberty, or, in a slightly different sense, peace v. freedom.

For the sake of security, neutrality shows its restrictive side, its function as a limit concept. It excludes what is non-neutral; indeed it defines it. In the literature, this concerns very concrete goals, like excluding Nazi and Communist ideologies, old-European religious fanaticism, or imperial colonialism. But for the sake of peace, neutrality shows its expansive side, its function as a warrant. It defines certain material conditions as neutrally peaceful. In the literature, it serves as a mechanism to sanction certain policies. The historicity of neutrality-as-security helps in assuring the contingency of its content, thus becoming a descriptive and hermeneutical tool. And the abstract notion of matter that lies behind neutrality-as-peace forces its content to be a contingent function of language; it operates by redefining the conditions of fair dialogue. In both

cases, then, when it comes to decide at the level of the particular case, —to concretely determine what counts as violence or oppression, for example—, neutrality arbitrates between competing meanings.

When neutrality determines what is reasonable and what is unreasonable, or sets the rules for the equal treatment of all reasons in an intersubjective space of communication, it operates on an acknowledged failure of the modern project of reason. It affirms pluralism for the sake of pluralism itself: competition between multiple incomplete and flawed views might better avoid tragedy than one colossal metaphysical mistake. Also, it acknowledges the fact that some differences cannot be reconciled at the level of theory, and that people are willing to kill and die for their beliefs. Thus, defining violence (and so on) demands a plurality of interpretations as much as responding to it demands a plurality of means.

neutrality works here as a justification while rejecting at the same time the very notion of grounding. It imports the authority of reason into its role as arbiter and admits weaker forms of reasoning, like reasonability, thereby demanding the same diminishing of all standpoints, the same acknowledgment of failure from all perspectives, and the same acceptance of means —i.e., the same reasonability.

Just like the value of neutrality-as-security is established through the absence of violence and the value of neutrality-as-peace is established by the presence of means, especially material means, so their counterparts, liberty and freedom, are defined in the terms of absence and presence. Liberty is the absence of an impediment whereas freedom is the presence of means. The

Berlinian, and classically liberal, distinction between negative freedom and positive freedom (that I here call liberty/freedom),⁷ is bridged by an inherent paradox embodied by neutrality. In order to arbitrate between what must be excluded for the sake of security and liberty, and what can be included, for the sake of peace and freedom, neutrality presents itself as a *view from nowhere*, (to borrow Thomas Nagel's expression). The paradox is intended: if it's from nowhere is not a view, and if it's a view its vantage point must be somewhere. Both metaphors, however, refer to the mind. A view is meant to have an ambiguous meaning, somewhere between seeing and thinking, and "from nowhere" expresses the idea that place or position is a necessary condition of knowledge. The problem of reason (its subjectivism and its longing to escape from it) is at center stage and thus the basic part of our political discussion is a discussion about reason. *The view from nowhere*—the expression, not the book—shows that neutrality arbitrates dialogues by forcing us to raise questions about the limits of reason as much as it prohibits us from reaching conclusions about it in the public domain. The by-products of reason are picked up as tools and then sharpened according to need—the sharpening is what the liberal theorists quarrel over—thus retaining its force without appealing to its necessity. Intersubjectivity and reasonability are such tools, the offspring of philosophical reason, which failed in its traditional aims; they are meant to retain the possibility of saying 'this is a view from somewhere' in an abstract way.

The consequence of this paradox is not only that neutrality harbors metaphysical commitments, but also that it conceals them. This is why I say it is not relevant for this project to decide whether it is a good or a bad thing that neutralist liberalism is metaphysically committed. It may be the case that there is no way around metaphysics, or that it is actually good that our politics are

⁷ I am referring here to Berlin's classical distinction in his 1969 essay: Isaiah Berlin, 'Two Concepts of Liberty', in *Four Essays on Liberty*, (London: Oxford University Press. New ed. in Berlin 2002.)

metaphysically loaded. What I do see as pernicious is that the neutral state falsely claims to favor no metaphysically loaded conceptions of human nature. The reason is that if it is true that a regime in which the state endorses a comprehensive view of the good and of human nature is oppressive, then the neutral liberal state is oppressive on its own terms. It does not make a difference that we can say that a neutralist state is somewhat stable or secure because peace in the sense of social order does not necessarily contradict oppression.

But even if it is not the case that endorsing certain comprehensive views is oppressive, I still find it pernicious that the liberal state excludes certain political views from the public sphere on metaphysical grounds. It is easy to agree on the convenience of excluding totalitarian or fanatical political ideologies, even if the reasons for doing so are theoretically shaky. Nevertheless, if my main argument is correct, then along with these fanatical views, other political positions are also excluded. I don't intend to judge which views are more pernicious to exclude, but merely to point out that several views that stem from, say, communitarian theories, gender studies, or critical race theory, are (implicitly) excluded from the public sphere because they are inconsistent with the idea of that individual at the heart of liberalism. At the very least, these views are compelled to submit to modification when entering the public domain, so as to adapt to the limits of public reason that neutrality enforces, and are, thus, tamed. I believe it is dangerous to do so, especially when one claims to uphold the political paradigm that avoids doing so.

By means of neutrality, liberalism normalizes and indeed trivializes a certain anthropology with metaphysical undertones, but I suggest that, by the same token, it also normalizes other, non-metaphysical positions that are in dispute in the background culture. Since liberal neutralism is a

theory of state power, and since it is rather successful in concealing its substantive content, it is ideal to protect and conceal what might be arbitrary in the status quo itself. In this way, national identities and cultures, chauvinist and sexist social hierarchies, and discriminations against minorities of all kinds —to name some well-known controversies— can be secured in their purported supremacy by cloaking themselves in neutrality. I submit that this status-quo bias is derived from the way in which neutralist authors justify their views on what it means to be a citizen, which, I claim, includes a certain understanding of human nature and the use of reason that neutrality enforces.

3. Dissertation outline

This dissertation is divided into five chapters.

a.

The first chapter — “Some consequences of an impossibility”— deals with certain disagreements between neutralist liberals. The main goal is to show how the meaning of the principle of state neutrality remains far from settled and that the reason for this disagreement has to do with the acknowledgement that a strong account of neutrality is practically impossible.

If neutrality is applied to everything the state does, it becomes too demanding of a constraint, and leads toward libertarianism. To solve this, some neutralists favor an interpretation of neutrality, which demands only that the state take into account actual controversies, so that if consensus is reached, nothing prevents the state from promoting certain ideas of the good. I call

this relative neutrality. Other neutralists favor a more absolute stance, so that the state must be neutral in an absolute way, not only regarding actual controversies, but also all potential ones; but then, to overcome the excessive restraint that such a stance demands, they circumscribe this demand to include only “constitutional essentials.” While these essentials inform the system, they allow for less fundamental levels of government to be somewhat non-neutral. Finally, what almost all neutralist authors argue is that neutrality should be applied only to the process of justification of state action, and not to the effects of that action, because doing so would actually force the state to intervene in a non-neutral manner. The price they are willing to pay for disregarding the effects of state action is the disappearance of certain forms of life, a disappearance that they consider unavoidable. The acceptability of such a loss, however, is predicated upon the assumption that these lost forms of life are too dependent upon closed systems to survive in the liberal system, which they see as fundamentally open.

In sum, total neutrality is impossible because, on the one hand, it forces the state to be either too weak or too strong, and, on the other, because it would make the state responsible for the diminution or disappearance of less liberal communities and ideologies. This explains why neutralist authors favor the interpretation of neutrality that focuses on the reasons given for state action, which turns their theories into a position on reason, communication, and the limits of self-determination. From the consequences of this impossibility, I derive the suspicion that the strategies that these authors employ to overcome it conceal certain metaphysical commitments.

b.

The second chapter — “Metaphysical Carriers”— explains what these authors mean by metaphysics and why they want to avoid it. It also aims to show why the potential metaphysical commitments that liberal neutrality harbors comes into view if regarded from the perspective of anthropology, that is, by moving the following question to center stage: do they presuppose an idea of human nature? If so, which kind? From that, it identifies the three notions mentioned above —intersubjectivity, reasonableness, and autonomy— as the main carriers of metaphysical content into the liberal system, and it shows the work they do in securing the idea of neutrality. It treats them as pieces in a system —from the outside, as it were.

Each of the three remaining chapters delves into these notions in detail —from the inside— with the aim of unveiling the metaphysical content that each carries.

c.

Chapter three deals with the notion of intersubjective communication in Habermas, and tries to account for the specific conception of rationality that is presupposed in intersubjectivity. In particular, it focuses on the ideal presuppositions present in Habermas’s theory of communicative action and how they inform his political theory. Methodologically, this chapter assesses and utilizes a set of critiques directed at Habermas over the years. It begins by interrogating the project of detranscendentalizing reason, and from there, it questions the notions of validity, universality, and “better argument” that inform Habermas’s understanding of political dialogue. The main conclusion is that the escape that intersubjectivity is meant to represent from the

solipsism of the modern subject is posited as a premise, thereby grounding an understanding of the powers of the subject that fail to be truly postmetaphysical.

d.

Chapter four aims at showing the idealizations that Rawls's notion of reasonability assumes. The main thread that the chapter follows explains that reasonability depends upon an account of the moral powers of the person that already formally contains the substantive content that Rawls's theory of justice claims to deduce impartially. Rawls assumes that we can know what is reasonable because we all share a similar set of moral capacities that inform a common minimal understanding of freedom and equality. In order to defuse the potential metaphysical content of both the moral capacities and the ideas of freedom and equality that they assume, Rawls abides by a separation of the sphere of the political and the *existential* and claims that it safeguards the integrity of our private comprehensive views. My claim, on the one hand, is that this separation does not prevent Rawls's comprehensive doctrine from spreading over and replacing other comprehensive views. On the other hand, I further claim that the separation itself presupposes that we have the kind of relationship with our beliefs that would make it possible for us to adapt them to the requirements of reasonability, when entering the political sphere. The kind of subject that can disengage itself from its beliefs, and negotiate them according to substantive notions of freedom and equality, is a reverberating echo of the idealist subject as put forth by the modern philosophical project. Like the previous chapter, this one also relies on, and brings together, many of the critiques that have been directed at Rawls in the literature in order to show the different aspects of his theory that make these presuppositions.

e.

Chapter five deals with autonomy in several authors. There is a key distinction between chapters three and four, on one side, and chapter five, on the other. While intersubjectivity and reasonableness are explicitly used to secure neutrality, autonomy is more philosophically problematic, due to its traditional connection with metaphysical thinking. For this reason, chapter five deals with the disagreements that neutralist authors have with respect to autonomy, and what these disagreements may imply. Generally speaking, it tries to show that the arguments that try to separate autonomy from a metaphysical understanding of human nature ultimately fail. Since they fail, my aim is to show precisely what kind of metaphysical content their idea of autonomy requires. For this reason, part of this chapter is devoted to a brief account of Kant's idea of autonomy, and another part to analyzing Levinas's unveiling of the idea of spiritual freedom present in the tradition of liberalism (in light of the "philosophy of Hitlerism."). But the main theme of the chapter concerns the idea that neutrality and autonomy are mutually dependent concepts in two important senses. First, in the sense that autonomy is the precise good that neutrality is designed to protect; that is, it is the *raison d'être* of liberal neutrality. Second, in the sense that for neutrality to make sense, the citizens that are ruled by the neutral state must be treated as self-originating, self-authenticating, and value-conferring subjects; and, for that reason, they must be free in the strong, metaphysical sense in which modern idealist thought took them to be. At the core of liberal neutralism, I claim, there is the positing of a *second* nature that is supposed to escape the determinisms of nature, history, social conditions, and bodies.

In my critique of autonomy in these authors, I attempt to tie together the multiple angles of criticism at neutralist liberalism I put forth throughout the dissertation. The fact that many views need to be considered and analyzed explains why this chapter is the longest.

f.

This dissertation begins by asking what it means to believe in freedom and ends by wondering to what extent that belief might be oppressive.

Chapter I

Some Consequences of an Impossibility

The question is this: if, in a just constitutional regime, some conceptions will die out and others only barely survive, does this by itself imply that its political conception of justice fails to be neutral between them? Given the connotations of "neutral," perhaps it does fail, and this is a difficulty with that term. But the important question surely is whether the political conception is arbitrarily biased against these views, or better, whether it is just or unjust to the persons whose conceptions they are, or might be.

There is no social world without loss: that is, no social world that does not exclude some ways of life that realize in special ways certain fundamental values.

John Rawls¹

1. The Principle of state neutrality

The principle of state neutrality is perhaps the most widely espoused ideal of liberal thinking. It states, generally, that in a pluralist society the state must “maintain a position of neutrality towards those large-scale moral, philosophical, or religious frameworks typically referred to in the literature as conceptions of the good, which claim the adherence of citizens.”² In other, more abstract, but also widely-used words, politics should be independent from metaphysics. Bruce

¹ John Rawls, *Justice as Fairness*, (Cambridge: Cambridge University Press, 2001), 154

² Ian Jennings, *Against state neutrality: Raz, Rawls, and Philosophical Perfectionism*, (Saarbrücken, Germany: Südwestdeutscher Verlag für Hochschulschriften GmbH & Co. KG, 2009), 16

Ackerman claims that the ‘essence of liberalism [is] to deny people the right to declare that their particular metaphysic and epistemology contains the truth, the whole truth, and nothing but the truth.’³ In *Political Liberalism* Rawls claims that he does not rely on any metaphysical doctrine of the person and he repeatedly calls his own political conception of justice ‘freestanding.’⁴

It is important to recognize that there are various formulations of political neutrality, and their differences are not merely formal, but point to differences in political content. Indeed, as Steve Wall and George Koslo assert, “there is no single conception of state neutrality that commands assent among proponents of the idea. (...) Accordingly, defenses of state neutrality are in part efforts to articulate the concept in a satisfactory way and in part efforts to explain why the concept so articulated should play a prominent role in political philosophy.”⁵ In a way, some of the most important debates among self-described liberals, and between liberals and non-liberals, turn on the notion of neutrality. It could be said that there is no neutral definition of neutrality that can summon up consensus among neutralists. This is so because the principle of neutrality is presented first and foremost as a delimitation of the state as a whole as much as a justification for its viability, and ultimately this delimitation is a definition of the state itself. Its function is described in abstract terms after a principle that answers, in an absolute manner, the question of what the state is warranted to do.

2. The Separate Sphere of the Political

³ Bruce Ackerman *Social Justice in the Liberal state* (New Haven, Conn: Yale University Press 1980), p355 One wonders in what sense does Ackerman use the word ‘essence’ in this context, but, on second thought, maybe this lapsus expresses well the problem with neutrality, as will be developed later.

⁴ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), for example: “Political liberalism, then, aims for a political conception of justice as a freestanding view.”, p10.

⁵ Steve Wall and George Koslo, introduction to *Perfectionism and neutrality*, (Oxford: Rowman and Littlefield, 2003), 2.

What all neutralists agree on is the separate character of the state⁶ —this is why it can be neutral. All formulations of political neutrality represent one way or another of endorsing this separate character. The state belongs to the political realm, and within this realm, it belongs to its own sphere, a sphere that is different and distinct from that of other human creations such as religion, morality, metaphysics, cosmology, or, in general, any personal attachments based on preference, desire, interest, reason or sentiment. Thereby, the state becomes the impersonal sphere of politics, where actions must depend on some rationale that is independent from every personal leanings: totally abstract.

Abstraction is a consequence of pluralism, and is seen by some defenders of neutrality, like Charles Larmore, as one of the main and most enduring innovations of modern liberalism. Thus, authors like Locke or Bodin can be interpreted as concluding that “the state must be based on principles that abstract from religious ideals,” and Locke’s political theory is valuable not because of his ideas about natural rights and property, “but rather this general method of abstraction and specialization,” which leads to the “differentiation of personal and political ideals.”⁷ The key of this separation is in understanding that axiomatically or “by its very nature,” liberalism “must be a philosophy of politics, and not a philosophy of man.”⁸

By abstraction, these authors mean the act of separating, or abstracting, the political realm from the other realms of human affairs, specially religion and morality. But they call it abstraction because they see religion and morality as specificities of one’s perspective, and the realm of the political must be able to abstract from those particularities, and become an inclusive genus for

⁶ Charles Larmore for example in his “In Praise of Bureaucracy” in Charles Larmore, *Patterns of Moral Complexity* (Cambridge: Cambirdge University Press, 1987), 40-48

⁷ *Ibid.*, introduction, xiii.

⁸ *Ibid.*, 25.

every (reasonable) specimen or, if one wants to scale up the ladder of the taxonomic rank, an inclusive kingdom for every phylum. In the modern context from which liberalism crops out, the autonomy of the political realm aims at a sort of universality; in the high modern context in which neutrality is consolidated as a central aspect of liberalism, this universality is seen as an abstraction from the concreteness of human views.

At the core, this abstraction underlies a more complex understanding of the limits and assumptions of liberalism. The whole system is predicated upon the possibility of separating the political sphere of collective action from the moral sphere of individual action. There exists the possibility, according to liberalism, of endorsing an idea of justice that is not comprehensively moral, or metaphysical, or even epistemological, such as Rawls' 'Justice as Fairness'. This idea of justice is consensually built without regard to personal moral views. How this consensus comes about, or why it is necessary, are questions that liberalism tries to answer pragmatically, based on ideas of stability and peace. But they are rooted in other, less clearly political notions, notions such as freedom of thought, respect for others, and equal dignity. Liberals construe this as exclusively political, and not, say, anthropological or metaphysical, by applying these notions to the public sphere and nothing else.

In a certain sense, this separation of spheres is axiomatic. Because it is theoretically assumed that this distinction makes sense, it is said that certain actions can be evaluated as being only political or only moral, or metaphysical. We do not need metaphysics to talk about justice, they say. Since it is true that the vocabulary⁹ that liberalism uses is not metaphysical, in order to see if this

⁹ In a Rortian sense, that is, as a toolbox for pragmatic uses.

separation holds, we need to look into the allegedly non-metaphysical concepts and their justifications, and see whether they assume a certain comprehensive landscape.

At this point, it may be useful to offer a definition of what the word metaphysics means in this context. This is will be done more thoroughly in the following chapter, but for the time being, it is relevant to point out that none of the authors that will be discussed bellow offer a detailed definition of what they mean by metaphysics. In general terms, what they mean is that their idea of a political conception of justice is not a comprehensive view of the Good or of human nature, and that it does not impose a moral or a cosmological system so that one may endorse the political architecture resulting from their theories from different, “overlapping” (in Rawls’s term), perspectives. These perspectives can be metaphysical in this sense, that is, they can be comprehensive, offer a moral system, be religious, or appeal to *a priori* notions of the essential constituents of the world and our existence in it, so as to endow them with purpose. At the base of all these characteristics, there beats an act of faith, a leap, on some ultimate meaning of the human experience, incompatible with others.

Our work, however, begins with concrete political concepts, such as neutrality, assessing whether in certain cases or moments, they fail to contain themselves within the political sphere.

The first problem with liberals’ falling back on abstraction, of course, is that the state must always be conducted by humans -- and not by gods or beasts, Aristotle’s idiots. Its actions, its laws, its principles must be shaped and carried out by humans; humans that are diverse in their respective attachments. Actions, laws and principles are inherently human, which makes the state a mirror of humanness, a Hobbesian super-person. But when neutralists like Rawls talk about “overlapping consensus of reasonable comprehensive views,” to justify “matters of basic justice,”

they do not mean that every person's view is superimposed on this Hobbesian super-person; they mean that a political conception can be accepted by different persons as being neutral, fair, and acceptable from their view.

By endorsing the very idea of neutrality, liberalism defines the state, above all, as an abstract actor, an impersonal person. As will be shown, the differences between liberals in their definition of neutrality are differences that revolve around a separate notion of action of the state that must be abstracted from personhood.

In dealing with conceptions of the good, liberals design strategies to keep the state neutral, hence the separateness, but these same strategies show that the stuff the state's actions are applied to is the same kind of stuff that is affected by conceptions of the good (moral, metaphysical, religious, cosmological). Here is where the spheres collide. Such intertwining cannot be avoided, thus different formulations of state neutrality preserve a degree of abstraction and separateness at the expense of a degree of intertwining. This is the in-house debate that often occupies neutralists: what can be preserved, what must be given up, when to intervene, and when not. The other positions that lie at each end of the spectrum of liberalism, on the other hand, argue either against these attempts at separating the right from the good along the lines of the public and the private (like Raz and other perfectionists do) or against all attempts at accepting some degree of intervention (like libertarians do).

This dissertation, however, is going to focus on the neutralist position, since the other two positions negate the separation of right and good either by embracing ideas of the good as state policy or by quite erasing the state.

3. Right vs. Good

At the heart of the principle of neutrality lies a specific understanding of what it means for us to believe in some idea of the Good. It entails a *preconception* of the Good, as it were. Since neutrality arises together with modernity, it takes on modernity's presuppositions. Thus, it's only logical that its understanding of the Good assumes that there is no universally valid content, but rather, Good has perspectival nature, which makes it plural —goods rather than Good—, as if the Good could only be understood metaphysically as being of a scattered, indefinite nature. Indeed, for political neutrality to be meaningful it is necessary to endorse a distinction between Right and Good —and, moreover, a priority of Right over ideas of the Good, in Rawls's phrase¹⁰— based on a fragmented conception of the Good and a conception of Right as a force that unifies the fragmented society, which reflects the fragmented Good. Under this paradigm, the Good is no longer identified as a Platonic ideal, the highest principle in which all ideas and paradigms converge, and from which all things stem. The Good becomes merely a personal pursuit, voluntarily accepted (in principle), and privately reasoned (individually or not). In contrast, the Right is understood generally as that which one is obliged to carry out; if necessary, coerced by the state.

In liberalism, this distinction is predicated upon the awareness of the existence of the *Other* as a being capable, generally speaking, of personal deliberation, and also upon a recognition that the Other's pursuit of the good is legitimate (at least, within certain limits and most of the time). From this awareness, there outcrop concepts such as Rawls's equal treatment or Larmore's equal respect, which connect freedom with equality and posit a distinction between respecting beliefs

¹⁰ John Rawls, "The Priority of Right and Ideas of the Good," *Philosophy & Public Affairs* Vol 17, No. 4 (Autumn, 1988), 251-276

and respecting persons: “Respect for persons is not just a belief that others have a capacity for developing beliefs justifiable within their own perspective and it cannot be simply deduced (even by ‘analysis of concepts’) from that fact; it is an obligation to treat others in a certain way because of that fact.”¹¹ Indeed, liberalism reifies the “Other” as something existing in general, the content of the identity of which must be treated as inconsequential for political purposes. Within this paradigm, a competition between different pursuits of the Good emerges. The Rawlsian notion of right (which will be further examined in chapter IV) is distinguished from the Good because it establishes the limits to any individual pursuit of the Good. This is the reason for the public nature of right. The idea of the Good is private in the sense that it is self-centered,¹² whereas right is public in the sense that it establishes “fair” relationships between individuals and groups holding particular ideas of the Good. Right and the Good limit one another, and that limit constitutes the border between the sphere of the political under the rule of neutrality and other spheres, which may be metaphysical. Right is neutral politically with respect to the conceptions of the Good, which are non-neutral and thus, operate outside the sphere of the political.

However, it can already be seen that when neutralists argue that the state should remain neutral with respect to competing notions of the Good, they do not prohibit the state from embodying all forms of political morality, but rather only those which fall outside the contemporary philosophical conception of right that liberalism construes. This is a crucial point for two reasons.

First, because it shows from the outset that political neutrality is not neutral in the absolute sense: it includes some moral values, even if weak, or as Larmore puts it ‘moral commitments,’ (among which he cites the desire for permanent civil peace and continuing rational conversations).

¹¹ Larmore, *op.cit.*, 64

¹² John Rawls, *Justice as Fairness* (Cambridge: Harvard University Press, 2001), 140-145

Moreover, there exists an overarching good that the state protects, a good that explains the very existence of the state-as-neutral, namely the articulation, in the most compatible way, of the plurality of goods present in a given society. This overarching Good of all goods —a Good defined as the coexistence of diverging goods¹³— stems from the conviction that is difficult in the extreme to ignore the metaphysical tenor of our relationship with the world: it acknowledges a certain nature of the Good —a concept of the Good with specific content— and the competition between different conceptions of the Good as a potential for conflict. It also assumes that different particular goods are compatible, at least partially and through certain abstractions. It denies, to this extent, their absolute incommensurability. It is important to note that a particular good may or may not be compatible with other particular good. The incompatibility of certain goods that arises entails the necessity of conceiving a hierarchy of goods, with the highest good on top. The good of compatibility of conceptions of the good creates such hierarchy and works as a decider when conceptions of the good, or the particular goods that emerge for them, are in dispute.

All this begs the question: neutralism is a discourse on the nature of the goods and of the Good, a concept it uses to articulate the right kind of society. To some extent, this already explains why neutralism is a metaphysical doctrine; but it does so from the lowest of standards, and it still remains to be shown which content that doctrine contains, which is the higher standard this investigation has set for itself.

¹³ There is a *logical* point here. The neutrality of state-neutralism is never, and can never be, neutral on the question of whether the state should be neutral. The non-neutral position or “assumption” here seems to be that the state’s role is to protect the Good and this abstract or non-specific Good seems itself to be the ideal of neutrality —this will be seen more clearly as we develop the argument.

Right, therefore, is the limit placed on those goods that we pursue, a limit that is based not on some universal precept of morality, but on the empirically asserted existence of other people's pursuits. There can be no other intrinsic evil than the destruction of the possibility of pluralism of pursuits. Consider the example of rape, that Nozick uses: on neutralist's own terms, rape cannot be an excluded behavior for its intrinsic evil,¹⁴ or its sinful nature, to put it in religious terms; rather, it must be excluded because it inhibits an individual's legitimate pursuit of the good which is made possible by the presence of the limits imposed by right: the pursuit that is made possible by personal integrity, sexual autonomy, freedom of movement, and so forth. Consider the opposite case: liberal-minded people like to say that gays should be able to marry because it doesn't hurt anybody (i.e., doesn't infringe on anybody's rights). This is not the same as saying that homosexual love is sacred, or that it should be protected as an institution.¹⁵ Conservatives then make the argument that same-sex marriage is harmful to society (to the impressionable children, say); to which liberals reply that there is no positive conception of the societal good other than the idea that people should do whatever they want as long as they don't infringe on the rights of others. These variations of the no-harm principle have a vanishing point in the idea that metaphysical views, and the pursuits they engender, should not be curtailed—or harmed—either. At the very least, the notion of right has that moral content defined in relation to the competition of goods.

This leads us to the second reason why the good/right distinction is crucial and fundamental for neutralists: the principle of state neutrality not only includes this understanding and defense of the Good that articulates goods, it also functions as a limit. It determines what is acceptable in

¹⁴ History of sanctioning rape, for example marital rape up until the 70s or 80s: look it up.

¹⁵ However, the notion of what it means to be treated equally emerges here.

the public sphere, which conceptions of the Good can partake in the process. This implies that ultimately neutrality also determines which goods can be purged, “neutralized,” and then turned into rights, and which cannot. Then, from the perspective of those rights, it determines again which goods can be privately pursued.

The debates and disagreements between neutralists revolve around this tension. They agree on the distinction, and they agree that the principle of state neutrality is the final arbiter on the question of what belongs to the domain of right and what to the domain of the Good. What they don’t agree on is what precisely should belong to each domain label.

Disagreements among neutralist can be explained through three dichotomies: the first revolves around different understandings of what it means for goods to *compete*; the second has to do with the range of the principle of state neutrality: to which public *actions* should it be applied?; the third is about scope: what does it mean for the state not to *favor* a particular conception of the good?¹⁶

4. First Dichotomy: Relative or Absolute neutrality

The first dichotomy divides those who, like Charles Larmore, believe that the state should be neutral only among those conceptions of the good that are actually competing in the public sphere, from those who, like Rawls or Nozick, believe that the state should abstain from endorsing any conception of the good when it cannot be justified by impartial means, regardless of the presence or absence of a conflict between particular goods in the public sphere. While the former focus on neutrality as a tool to resolve conflicts, the latter aims at a more general structuring of the basis of a liberal democracy. While the former, Larmore in particular, is less

¹⁶ For these distinctions I am partially following Ian Jennings’ book *Against state neutrality*, (GmbHCo, Saarbrücken, Germany, 2011).

hopeful when it comes to the neutrality of neutrality, and thus supplements it with a realist approach to the good of equal respect, the latter believe it is the only warranted theoretical position if we are to avoid metaphysics and thus it is best hope for stability in the long run. Let us unpack these ideas.

4.a. Relative neutrality

The first option accepts the possibility of the state promoting the Good when there is consensus around a particular good that stems from it. According to Larmore, the ideal of political neutrality “demands only that so long as some view about the good life remains disputed, no decision of the state can be justified on the basis of its supposed intrinsic superiority or inferiority.”¹⁷ This version of political neutrality can be labeled ‘relative neutrality,’ in the sense that the state is neutral only when there is competition. This is neutrality in the strict logical sense of “neither A, nor B” —from the latin *ne-utrum*, not one or the other—and it also corresponds with the origin of its modern political use during the European wars of religion in the sixteenth century.¹⁸ This means, of course, that the word neutrality is here used as a *captatio benevolentiae*, as a mere show of good faith with respect to plurality, an ideal to win goodwill for the system, because in truth the state does indeed endorse and promote conceptions of the good, namely those accepted by the people.

What is relevant about relative neutrality is that it is consistent with a recognition that absolute neutrality is impossible, that is, the idea that the state could abstain from promoting conceptions

¹⁷ Larmore, Charles. *Patterns of moral complexity* (Cambridge: Cambridge University Press, 1987) p.47

¹⁸ The first explicit reference of political neutrality in Western international law is the Treaty of Augsburg in 1555, where the Emperor of the Holy Roman Empire declared himself to be neutral among Catholic and Lutheran princes. For a historical-philosophical account of neutrality and the historical background whence it emerged, see Reinhart Koselleck, *Critique and crisis enlightenment and the pathogenesis of modern society*(Cambridge, MA: MIT Press, 2015), and Paolo Prodi, *Una historia de la justicia: de la pluralidad de fueros al dualismo moderno entre conciencia y derecho* (Madrid, Spain: Katz, 2008).

of the good in the political realm. In *Patterns of Moral Complexity*, Larmore accepts not only that political neutrality is, in fact, not absolutely neutral, but even that the usual ways of justifying political neutrality are not neutral either. They do in fact endorse certain conceptions of the good, in particular those associated with liberalism. His purpose is then not so much to avoid all political commitments to the good, but rather to make explicit in what way the state can legitimately promote certain undisputed goods.¹⁹

Larmore's neutrality is 'relative' in two other ways: first, even if it depends on actual conflicts, it has limits: it functions as a check against fanaticism. Not all conflicts are acknowledged or treated from a neutral perspective: Nazis, the usual example in the literature, are excluded from legitimate competition. Nazis believe in the intrinsic goodness of the "arian race" over, say, the "jewish race," and their political program consists in the dominion and extermination of other people's lives precisely because of that superiority.²⁰ They are, in this sense, not willing to build a neutral state. Only those who accept neutrality as an arbiter of disputes, that is, those who would endorse neutrality when in charge of government, can benefit from the freedom and protections that the neutral position of the state provides, when entering a conflict in the public sphere.²¹

This means that neutrality is also relative because it can increase or decrease depending on which 'rivalries' are seen as legitimate. Therefore, neutrality depends also on a notion of legitimacy that is movable, depending on the actual conflicts.

The aim to exclude fanaticism and totalitarianism is in fact common to all neutralists, relative or absolute. The definition of what constitutes fanaticism or totalitarianism is complex, but it is one

¹⁹ Larmore also attempts to offer a neutral justification of neutrality, which we will analyze in the following chapter.

²⁰ I will say more about the example of Nazism on chapter 5

²¹ Unless there is consensus around a particular good, as we will see.

of the functions of the principle of state neutrality: it not only articulates plurality, it also decides which plurality is accepted in the public sphere. The requirement that all legitimate ideologies accept one form or another of the Principle of neutrality is only the formal part; then, each author offers a more or less detailed set of material principles that characterize these excluded ideologies. The latter is in fact a function of the former, and it expresses the content that neutrality has been carrying all along, as we will see further below.

Simultaneously, since neutrality is also a constraint on the state, and sometimes this constraint can be too great for the state to pursue any action, liberals must be willing to consider tradeoffs between the goal of neutrality and other political goals. “If full neutrality makes such a decision impossible, then neutrality should be made more restrictive until a decision becomes possible.”²² The goal that Larmore pursues is not absolute neutrality, but maximum neutrality, thereby turning the principle into a contingent political tool.

Ian Jennings claims that the reason for this relative perspective has to do with the motives for endorsing neutrality. In this case, it has to do with political stability: “if one is moved to support state neutrality solely out of, say, a concern for political stability, then it is indeed unclear why one would hold that the state should refrain from backing monogamous marriage.”²³ Monogamous marriage is clearly an idea of the good life, but it gathers consensus. Relative neutralists defend the claim that the state should refrain from promoting any disputed notions of the good because doing so could endanger political stability, while supporting an undisputed position not only does not endanger stability, it may foster it.

²² Larmore, Charles. *Patterns of moral complexity* (Cambridge: Cambridge University Press, 1987) p.68

²³ Jennings, *Op. Cit.*, 20.

This raises for us the question of what it means for a particular good to be undisputed, or what it means for there to be consensus around a certain good. Both in quantitative and qualitative terms, dispute and consensus are problematic notions, as it is shown by the difficulty in describing the conditions under which consensus is free, or by the objections that claim that consensus might be just a function of the status quo, as Della Rocca has shown in his *The Taming of Philosophy*,²⁴ or by other criticism, such as Walzer's critique of procedural set-ups to establish consensus²⁵ —all of these will be discussed further in chapters 4 and 3 respectively. For the time being, however, in the context of relative neutrality, we must ask: Is it enough for one individual to dispute a certain notion for it to be understood as disputed and without consensus? And beyond the question of how many people are needed to consider a consensus disputed, what kind of disputing is legitimate?

The first question opens the debate between those who believe that consensus implies consent (like Ackerman, who understands legitimacy as the outcome of a readiness to answer any question about the “legitimacy” of a given power, a question that can be asked at any time²⁶), and those who believe in a more passive form of legitimacy, namely, that as long as there is no politically organized opposition, institutions like monogamous marriage or the public funding of museums can be understood as gathering a societal consensus. These positions serve the goal of stability —which is one of the goods that this position of neutrality defends.

²⁴ Michael, “The Taming of Philosophy,” in *Philosophy and its History*, Ed. Mogens Laerke et al., (Oxford: Oxford University Press, 2013)

²⁵ Michael Walzer, “A Critique of Philosophical Conversation,” in Michael Kelly (Ed.) *Hermeneutics and Critical Theory in Ethics and Politics*, (Cambridge: MIT Press, 1990)

²⁶ Bruce Ackerman, *Social Justice in the Liberal state* (New Haven, Conn: Yale University Press, 1980), 26

The second question, on the other hand, invites a more nuanced philosophical debate. A legitimate dispute is either one that falls within the bounds of neutrality itself because it says: “I object to this promotion of the good because I would prefer another good and the state should be neutral among all conceptions of the good,” or one that defends the convenience of promoting another good on the grounds of some idea of progress. Neutralists only accept the first kind; that is, a legitimate disputation is what liberals call a “reasonable” disputation. What precisely reasonable means is a matter for much debate (as we will see below and in particular in Chapter 4), but in general terms a reasonable position is one that its adherents regard as potentially recognized as reasonable by those who disagree with it.

In this context, then, a reasonable position is such because it does not pose a threat to stability. The one and the same position against a given consensus could potentially be a source of instability only if the state failed to be neutral. Neutrality makes it possible for different positions to compete. Of course, it is possible to find common ground, a space for consensus between two divergent ideas of the good, but this must be worked out on neutral ground. A neutral ground, according to Larmore, is where parties retreat when they see that there is reasonable disagreement, and where

[...] each should prescind from the beliefs that the other rejects, (1) in order to construct an argument on the basis of his other beliefs that will convince the other of the truth of the disputed belief or (2) in order to shift to another aspect of the problem, where the possibilities of agreement seem greater.

He calls this “abstracting from a controversial belief” through the “universal norm of rational argument.”²⁷ So when it comes to understanding the meaning of disagreement it is important to note that a legitimate opposition to a given policy must be construed in a way that already

²⁷ Charles Larmore, *Patterns of moral complexity* (Cambridge: Cambridge University Press, 1987), 57

accepts the limits and constraints of neutrality. Any belief that attempts to enter the space of consensus needs to be neutralized, that is, politicized and tamed by the need of equal toleration to other's views. Neutralizing implies accepting the 'universal norm of rational argument', which in Larmore's view—and in many other liberal perspectives— means not only a set of acceptable epistemological axioms, but also a set of moral qualities— in the case of Larmore, the material content is his understanding of "equal respect." These epistemological and moral commitments limit the space of the conversation, exclude beliefs that disagree with the fundamental content of neutrality, and make possible the emergence of consensus. The key word here is 'conversation'. Understood in this way, neutrality is a rule that applies to certain dialogues, and these dialogues are the most important aspect of establishing political legitimacy.

Moreover, the need for disputing and confronting the consensus in order to widen the space of neutrality puts the burden of argument on the dissenting party, who must actively show her disagreement. A cursory consideration of this claim regarding where the burden of argument falls shows that it is problematic due, among other things, to the awesome power of certain public institutions.

Stability, however, privileges the majority over the minority even to the extent of making what is not reasonable but hegemonic appear reasonable, for two reasons: First, the status quo appears more reasonable than what is not the status quo (be it what was before the status quo or what will be in the future). Second, the hegemonic invokes the notion of propaganda, i.e., the idea that the powers that be exert influence over the people's intuitions regarding what is reasonable or realistic, and what is not.²⁸ Even when this conception of neutrality aims at showing respect for

²⁸ Again, this will be examined more deeply in Chapter IV, when discussing Della Rocca's *The Taming of Philosophy*. Specifically, the point is that liberal neutral positions (Rawls) rely on intuitions which are just facts about us to achieve overlapping consensus and reflective equilibrium; but these intuitions are themselves facts about whoever has these intuitions and hence anything but neutral.

dissent, dispute, rivalry, and so on, it does so with regard to the current state of affairs at any given moment. When there is no objection, when there is consensus, an unchallenged public good tends to become part of the cultural background, and this is why it acquires more stability.

By putting the burden of proof on the dissenter, and the limit of neutrality on rivalry, this notion of relative neutrality makes the state non-neutral in absolute terms. The good of stability becomes an end in itself. The trigger that widens neutrality, —namely, dissent—, is very misleading. After all, the absence of an objection, the absence of rivalry, may not be the sign of the presence of a consensus, but of the prevention of possible disputes and the solidification of a set of goods as neutral ground. In this way, the neutral state is essentially conservative.

Admittedly, a certain good is not the same as a comprehensive view of the good —as Rawls calls it—, or a conception of the good, which implies an overarching perspective.²⁹ But the presence of a certain good promoted by the state, even if it is understood only as that particular good and not as an instance of a comprehensive view, when it prevents the emergence of a rival good, it also prevents the emergence of the conception of the good from which that good can be said to stem. It works as a preemptive falsification. And thereby, its neutral ground can be seen as a comprehensive view. At the same time, however, even when a state explicitly promotes only a particular good, and not any comprehensive view of the good, this particular good necessarily has implications for a comprehensive conception of the good. In short, it is hard, if not impossible, to remain neutral regarding any comprehensive conception of the good while being non-neutral regarding any particular good.

4.b Absolute neutrality

²⁹ See John Rawls, “Justice as Fairness: Political not Metaphysical”, *Philosophy and Public Affairs*, Vol. 14, No. 3. pp. 223-251 (Summer, 1985). This paper will be examined in chapters II and IV.

This flaw in the relative understanding of neutrality is at the root of the other position in the dichotomy described above: the belief that the state should abstain from any endorsement of the good, when it cannot be justified without appealing to a certain conception of the good, regardless of whether or not there is any actual conflict between particular goods in the public sphere.

If the motive for endorsing relative neutrality is ultimately political stability, the reason that justifies a more absolute commitment to neutrality is a defense of individual autonomy and diversity, as Jennings believes³⁰. Diversity and autonomy mirror each other. Diversity is legitimate protected because autonomy inevitably creates different worldviews, it is how the state contemplates the existence of autonomous agents. From the perspective of the state, diversity can be seen as a good to be protected, but since diversity is an open-ended concept, any endorsement whatsoever of one particular good endangers not only the actual presence of other pursuits, but also the potential for new, unimagined ones. If diversity is good then, its increase is an increase of this good, and it would be logical to think that the fewer (institutional) prescriptions there are, the more chances there could be for the good of diversity to increase.

However, when governments adopt a *laissez-faire* stance, stronger dominant life-styles and moral values can become oppressive (resulting in spirals of silence³¹). This should lend support to the idea that minority positions should be protected by the state, bringing the principle of state neutrality back to a relative position, i.e., that of finding equilibrium between socially strong and weak views of the good (which also leads us to one of the other dichotomies that will be discussed

³⁰ Jennings, *Op. Cit.*, 20

³¹ I referring here to the the spiral of silence theory as described by Elisabeth Noelle-Neumann, which explains how the fear of being excluded due to one's opinions or one's belonging to a minority may lead to silence one's views in in the public sphere. See Noelle-Neumann, Elisabeth (1974), "The spiral of silence: a theory of public opinion," *Journal of Communication*, (2): 43–51

below: neutrality of aim vs neutrality of effect). This is where autonomy comes into play. Since autonomy itself is metaphysical in the sense that it begins with the self as a metaphysical notion, liberal neutralists try and thread more carefully than usual what they mean by autonomous, or self-legislating citizens. Even if personal autonomy is not seen as a form of moral solipsism, and to the extent that it includes notions of rationality, which, at least in contemporary times, are inseparable from dialogical forms of relation, it is difficult to argue that it does not imply the idea that the self is the origin of behavior, a *primum movens*. Neutralists try and explain autonomy as a political concept, meaning self-legislating, as opposed to heteronomy, being subject to external laws. Political autonomy means not being involuntarily subject to any external laws made by the state. Even a perfectionist such as Raz distinguishes it from personal autonomy. Personal autonomy would be more metaphysical in the sense that the self is the agent origin of personal behavior, as opposed to material or psychological forces as described by Darwinian materialists or Skinnerian behaviorists, for example. In short, liberals believe that autonomy, as opposed to heteronomy, does not necessarily imply a strong metaphysical account of free will, while freedom of the will can be consistent with a lack of political autonomy. While Chapter V will be devoted to this topic, I want to suggest here that this distinction is self-serving and infused with a dialectical contradiction that makes it useless for the project of relieving liberalism of metaphysical commitments. The idea that one can voluntarily accept a system of laws not only depends on free will, but also has the effect of shaping all public institutions, and the public sphere itself, with the kind of commitments proper of a strong understanding of autonomy.

If the flaw of the relative neutrality doctrine was that it hindered the emergence of unimagined notions of the good because it solidified the state of affairs, the flaw that even neutralist authors see in a total commitment to absolute neutrality is that, theoretically, it scarcely allows the state to

do anything, with the obvious exception of protecting the individual's sphere of action. This would be a liberalism that tends toward libertarianism.

This position invites the question of which prohibitions we uphold in the context of interpersonal relations (i.e., from the rules of private associations to rape —if it can be understood as a 'personal relation'), and which prohibitions we uphold in the context of political discourse (i.e., it excludes all forms of communal policies). Even if this problem is not peculiar to this version of neutrality, but rather common to all of them, it is seen more clearly when autonomy is the vanishing point.

Moreover, a commitment to autonomy introduces the idea of respect for people as ends in themselves, which if asserted thus, is a metaphysical claim. Robert Nozick favors a pure commitment to absolute, libertarian neutrality in order to protect the individual. The inviolability of the rights of the individual is one of the core moral principles around which he builds his ideal of a minimal state. From his perspective, any damage done to an individual for the sake of the greater good treats people as means to benefit others, without sufficient warrant:

Why not hold that some persons have to bear some costs that benefit other persons more, for the sake of the overall social good? But there is no *social entity* with a good that undergoes some sacrifice for its own good. There are only individual people, different individual people, with their own individual lives. Using one of these people for the benefit of others, uses him and benefits others. Nothing more. What happens is that something is done to him for the sake of others. Talk to an overall social group covers this up. (Intentionally?) To use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has. *He* does not get some overbalancing good from his sacrifice, and no one is entitled to force this upon him —least of all a state or government that claims his allegiance (as other individuals do not) and that therefore scrupulously must be neutral between its citizens.³²

³² Robert Nozick, *Anarchy, state, and Utopia* (USA: Basic Books, 1974), 71

Nozick's is one of the most consistent formulations of the consequences of endorsing neutrality because non-intervention is the best way to avoid being responsible for favoring any good. But his theory shows that even in this model, the formulation of neutrality ultimately refers to an idea of the person, in this case, of the person as a separate individual, whose moral responsibility does not imply the need for political coercive interventions. This is an account of human life and ultimately of human nature. To the extent that this account of human nature is meant to be universal and independent from contingency, the connection between assuming an anthropology and a metaphysical view (even if it is only of the *second nature* of man³³) is in need of justification. Specifically, this anthropology harbors a conception of freedom as ego-centered and a radical theory of complete individual autonomy. Yet the question of applicability remains open. What does it mean for the Nozickian state to be neutral? After all, Nozick's minimal state allows for the enforcement of certain rules and prohibitions aimed at protecting the physical security of the inviolable person. And such rules and prohibitions must have a certain form if they are to be neutral. So even in his case, the solution will come with the combination of the other dichotomies that divide the interpretation of neutrality in liberal authors, as we will explore shortly.

And yet Nozick is not a liberal, but a libertarian. For liberals, the key aspect of any endorsement of absolute neutrality is how it can be maintained without eliminating the state altogether or without turning the competition of goods into some version of the Hobbesian state of nature.

The significance of the flaws of absolute neutrality explains why it would be inconceivable that authors like Rawls and his followers would endorse such a version of the principle. Absolute neutrality without further qualification would demand only that the state secure certain

³³ Which in turn relates to an idea of *first nature*, which implies a broader metaphysics, even if it is naturalism. Markus Gabriel, *Transcendental Ontology Essays in German Idealism*, (New York: Continuum International Publishing Group, 2011), Ch 1, *passim*

individual liberties and personal security but not much else. As Rawls says, “these guaranteed liberties taken alone are properly criticized as purely formal. By themselves they are an imposed form of liberalism, indeed not liberalism at all but libertarianism.”³⁴ Rawls’s and Dworkin’s ideas of liberalism include, to an extent, some material principles. Libertarianism “does not combine liberty and equality in the way liberalism does; it lacks the criterion of reciprocity and allows excessive social economic inequalities as judged by that criterion.”³⁵

The flaws of a purely absolute commitment to neutrality explain the specific way in which theorists like Larmore develop their position on relative neutrality. Larmore contends that some neutralists justify their commitment to impartiality by appealing to certain ideas of human nature. Autonomy and experimentation are the ones at play here, he says, and together with an account of moral skepticism, they form the classical triad of reasons to favor a liberal state. According to Larmore, at the end of the day, since autonomy, experimentation and skepticism form the vanishing point of this justification of neutrality, those who endorse it, end up justifying the intervention of the state by invoking “some view of human flourishing than can best be promoted if the government maintains a neutral posture toward the variety of human aspirations.”³⁶ This would be a concealed form of perfectionism.

4.c The Problem of Practicality

But things are not that simple. Although Jennings claims that what distinguishes a relative from an absolute commitment to state neutrality is that in the former stability is at stake, while in the latter autonomy is at stake, in fact, all authors cite both some form of stability and autonomy as

³⁴ John Rawls, Introduction to *Political Liberalism* (New York: Columbia University Press, 1993), lvi

³⁵ *Ibid.*

³⁶ Larmore, Charles. *Patterns of moral complexity* (Cambridge: Cambridge University Press, 1987), 51

the forces driving their adherence to state neutrality. To sharply divide them along these lines (stability/autonomy) would be oversimplifying their positions. Rawls is the clearest example. In the introduction to *Political Liberalism*, he goes to a great length explaining that the main difference between the latter and his earlier *A Theory of Justice* is due to the problem of stability: “To understand the nature and extent of the differences, one must see them as arising from trying to resolve a serious problem internal to justice as fairness, namely from the fact that the account of stability in part III of *Theory* is not consistent with the view as a whole.”³⁷ In *Theory*, liberalism is presented as a comprehensive doctrine in the tradition of Kant, Mill and others, which harbors metaphysical commitments, and thus it is not neutral. His aim in *Political Liberalism*, as the adjective in the title suggests, is to present a ‘political conception’ that is ‘freestanding’ and can gather an ‘overlapping consensus’ of ‘reasonable comprehensive doctrines’ otherwise ‘irreconcilable.’³⁸ In *Political Liberalism*,

I start with the idea of the domain of the political together with the idea of political conception of justice, using as an example the conception of justice as fairness. These ideas and their distinction from comprehensive doctrines of all kinds, are, I believe, among the most crucial ideas of liberalism.³⁹

As we will see, one of the main features of a political conception as opposed to a comprehensive doctrine, is that the former applies only to the domain of the political and, within it, only to the constitutional essentials of a democratic regime,⁴⁰ whereas the latter “applies to all subjects and its virtues cover all parts of life.”⁴¹ Because it is explicitly not comprehensive in this sense, Rawls

³⁷ John Rawls, Introduction to *Political Liberalism*, (New York: Columbia University Press, 1993), xvi

³⁸ “The ambiguity of *Theory* is now removed and justice as fairness is presented from the outset as a political conception of justice.” John Rawls, *Ibid.*, xvii

³⁹ *Ibid.*

⁴⁰ “A political conception tries to elaborate a reasonable conception for the basic structure alone and involves, so far as possible, no wider commitment to any other doctrine.” John Rawls, *Ibid.*, 13

⁴¹ *Ibid.*, xvi, footnote 4.

considers his political liberalism to be ‘freestanding’ and ‘impartial.’ It is also stable because it is able to gather overlapping consensus from various (reasonable) comprehensive views. But his idea of overlapping consensus is not, as Larmore’s relative neutrality can be seen, an attempt at finding a good that is agreed upon by the content of diverse comprehensive views. Rawls tries to articulate a freestanding conception in the sense that it must be completely independent from metaphysical or comprehensive views. A conception that can be seen as right on its own terms. So that “while we want a political conception to have a justification by reference to one or more comprehensive doctrines, it is neither presented as, nor as derived from, such a doctrine applied to the basic structure of society, as if this structure were simply another subject to which that doctrine applied.”⁴² This is an instance of the separateness of the political realm and the state that we have seen above, and while it carries the echo of Kantian moral and political theory, it rejects its metaphysics of reason.

This distinction between the political and the metaphysical is also applied to the notion of autonomy —and to all other notions relevant to the discussion. While I will develop the effects of this distinction in Rawls’s system later on (in Chapter II, IV and V), for the current discussion it is enough to point out that for Rawls, the significant notion of autonomy in this context is political autonomy, a narrower version of the general metaphysical autonomy that manages to avoid the metaphysical aspect of it. Everything is separated in order to affirm the freestandingness of his view. Moral autonomy in its comprehensive content must be left outside the political realm and this way pluralism is preserved even to the extent of protecting views that reject autonomy. This, of course, applies to the possibility of agreeing to let the state pursue a certain good. The separate character of the political, given the fact of (reasonable) pluralism, makes it impossible

⁴² *Ibid.*, 12

for the state to favor any idea of the good outside the realm of rights. Perfectionist policies are excluded from his system to the extent that “the principles of justice do not permit subsidizing universities and institutes, or opera and theater, on the grounds that these institutions are intrinsically valuable.”⁴³

5. Second Dichotomy: Narrow or Comprehensive neutrality

Since an absolutely neutral position entails a libertarian position, other dichotomies come into play. The first way to deal with the impossibility of making an unmitigated commitment to neutrality is to tune its range: how pervasively does the principle reach inside the state. Is the principle of state neutrality to be applied to all government institutions and actions or just a select few? An unmitigated commitment to neutrality would be impossible to apply if one expects the principle to govern all government institutions and activities, without leaning towards libertarianism. This is why some neutralists, especially if they are absolute neutralists, argue that the principle of state neutrality should only be applied to what Rawls calls ‘constitutional essentials:’

- a) fundamental principles that specify the general structure of government and the political process: the powers of the legislature, executive and the judiciary; the scope of majority rule.
- b) equal basic rights and liberties of citizenship that legislative majorities are to respect: such as the right to vote and to participate in politics, liberty of conscience, freedom of thought and of association, as well as the protections of the rule of law.⁴⁴

As we will see, whether these essentials already presuppose some conception of the good that goes beyond what’s warranted remains a matter of debate but in this understanding of the

⁴³ John Rawls, *A Theory of justice* (Revised edition) (Oxford: Oxford University Press, 1999), 291.

⁴⁴ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), .227

principle of neutrality, that Jennings calls ‘narrow’, these essentials are seen as the rules of the game, which everyone must follow as they pursue and promote their own conceptions of the good.

From this perspective, an absolute commitment to neutrality becomes plausible. These constitutional essentials do not need to take into account whether certain conceptions of the good stand in competition with one another. On the contrary, they must strive for a neutral stance that is able to abstain from privileging any actual or possible conception of the good. This helps us understand why for authors like Rawls or Brian Barry there exists a correlation between their absolute commitment and their narrow understanding of its application.⁴⁵ So, conversely, there exists a correlation between a *comprehensive* understanding of the principle and a relative commitment to neutrality, as is the case with Larmore. If neutrality is understood as relative to a competition between different conceptions of the good, then it can be applied to all activity of the state, even when it comes to peripheral legislation and procedure. One might argue that it could be possible to have a comprehensive and absolute position, in so far as it is understood as a regulative ideal. However, turning the principle into a horizon would mean to accept practically the lowering of both standards.

The explanation for these correlations is impossibility. When Rawls asks himself why he should commit only to a narrow application of the limits of state action —through what he calls ‘public reason’—, he answers: “my aim is to consider first the strongest case where the political questions concern the most fundamental matters. If we should not honor the limits of public reason here, it

⁴⁵ Brian Barry, *Justice as Impartiality* (Oxford: Oxford University Press, 1995)

would seem we need not honor them anywhere. Should they hold here, we can proceed to other cases.”⁴⁶

But even when all these qualifications and restrictions are in place, Rawls admits that this position is imperfect on both ends. On the one hand, “I grant that it is usually highly desirable to settle political questions by invoking the values of public reason,” which the narrowness of his commitment to neutrality does not demand, and on the other, “Yet this may not always be so.” Jennings calls this passage ‘rather obscure,’⁴⁷ because Rawls seems to admit that his freestanding political conception, even when it is applied to constitutional essentials, and especially when it is applied to what is beyond those matters of basic justice of society, might not be as freestanding, as merely political-not-metaphysical or as non-comprehensive as he would desire it to be. The basic structure that must be justified neutrally might not hold. But, as Jennings says, “The circumstances under which it would not hold remain unclear.”⁴⁸

What both comprehensive and narrow understandings have in common is that they allow for more local or less general public institutions and actions to promote certain conceptions of the good. In the case of comprehensivists, government officials can promote the good without damaging the principle of neutrality as long as there is consensus and no rivalry. This can and does occur, for example, in regards to local policy questions or in shaping public school curriculum. This is also possible at the highest level of the state for the same reasons. For those who favor the narrow position, however, as long as certain constitutional essentials respect the

⁴⁶ Rawls, *Op. Cit.*, 197.

⁴⁷ Jennings, *Op. Cit.*, 19, footnote 51

⁴⁸ *Ibid.*, 20.

principle of neutrality and peripheral legislation respects the limits set by the constitution, government officials can promote conceptions of the good that are democratically agreed upon.

All these differences and similarities revolve around the notion of coercion. Conceptions of coercion that regard every public action as imposed upon citizens—for even incentives and promotions are funded through taxation, which is compulsory—tend to favor a more comprehensive application of neutrality, so as to cover every act of coercion. On the other hand, if we conceive of coercion in the more limited sense, namely, as the threat of punishment, as in criminal Law, generally a narrower application of state neutrality would suffice because only the cases where *real* coercion is present require the highest level of restraint.

The problem of whether genuine neutrality is possible is bound to emerge again and again, that is why even if it is seen as a regulative ideal it implies impossibility. This speaks to the nature of neutrality itself. The tension between neutrality and coercion, between neutrality and the Good, and between neutrality and the free use of reason become more and more problematic the more one reflects on the possible ways the principle can be introduced in the regular business of the state.

6. Third dichotomy: aim or effect.

6.a Aim

This is even more clearly seen once we introduce the third and last dichotomy, which concerns the scope of application of the principle of neutrality: what does it mean for the state to not *favor* any particular conception of the good? The classical debate over this question is between those

who interpret neutrality as being a question of aim and those (almost nobody) who see it as a matter of consequences.

neutrality of aim implies that the state should not harbor or act out of any intention to helping or hindering either side in a conflict over goods, be the conflict actual (relative) or merely potential (absolute), or, be it over the basic structure of society (narrow) or rather over each and every state action (comprehensive).

How precisely to achieve the aim of neutrality poses a difficult problem. What does it mean “to favor”? How can we know when the *aim* of the state is neutral? Any analysis of an individual’s intentions is already difficult, more so when it comes to collectivities, individuals that rule institutions. Consider the debate around the interpretation of the law, in particular a Constitution, between those who favor the analysis of the intentions of the crafters and those who favor the intentions of those who have to interpret it. How can a norm of this kind—that the intention of the state must be neutral—be enforced? Intentions are, after all, a mental state, which shifts the question yet again into the domain of reason and into the field of expectations. The reason for this vague standard is double: theoretically, it is difficult to offer an understanding of *favoring* without appealing to a certain idea of the good. We favor someone when we offer them a good or make it easy for them to achieve that good, as we understand easiness or advantage. Practically, the focus remains on intentions because one assumes that any action by the state has effects that tend to favor certain conceptions of the good life over others.

Jennings likes to use the metaphor of the sports referee here. A neutral referee applies the rules of the game fairly, yet one team eventually wins and the other loses. “The influence of, for example, particular religions or art forms may wax or wane. If neutrality is a matter of aim, we are not

entitled to conclude from these ‘victories‘ and ‘defeats‘ that the state lacks neutrality.”⁴⁹ Will Kymlicka offers the example of official national languages: promoting the strongest language in a given society threatens minor languages, and yet, a neutral reason might be found to do so. Is a neutral justification enough to prevent the state, or some party in charge, from using the power warranted by neutrality with the aim of favoring a given group?

There are cases that perhaps more clearly reflect this problem, namely, that neutrality of aim seems to be insufficient to clarify when the state is neutral. Consider the two sides of the coin. Nozick uses this example:⁵⁰a law that bans rape may be seen as targeting men,⁵¹ and a position must be found to protect rape victims and show neutrality simultaneously. The actual controversies around sexual assault, namely, balancing victims’ rights and defendant’s rights, speak of the tension within neutrality between what it means to protect a certain status quo and the appearance of impartiality expressed in public reason and justification. Racial profiling is the other side of the coin: is the fact that some marginalized or historically oppressed minorities are more numerous in law-breaking statistics proof enough that there are no biases or outright racism, that is, non-neutral aims, in those policies?

These doubts persist with Nozick’s position, even if it is tamed thus: “There is an *independent* reason for prohibiting rape: people have a right to control their own bodies, to choose their sexual partners, and to be secure against physical force and its threat.” The fact that a given prohibition has different implications for different persons “is no reason to condemn it as non-neutral,

⁴⁹ Jennings, *Op Cit*, 32

⁵⁰ Robert Nozick, *Anarchy, state, and Utopia* (USA: Basic Books, 1974), 273

⁵¹ Of course, it may be seen in this way, but it may also not be seen in this way – it would state that rape is forbidden neutral of gender

provided it was instituted or continues for (something like) the reasons which justify it, and not in order to yield differential benefits.”⁵²

The existence of an independent reason might constitute an instance of neutrality, but it does not preclude the possibility that a non-neutral aim can be seen as operating underneath. The problem with ascertaining the aim is solved via the idea of justification. “The neutrality of justification interpretation understands neutrality to be a rule about what kinds of reasons the state may legitimately offer —over and above the normal requirements of reasonable plausibility — in justifying the legislation or policy it puts forward.”⁵³ It prevents the state from promoting any policy that enhances a particular (and/or controversial) conception of the good on the ground of its intrinsic superiority. According to this rule, if a policy cannot be defended in terms of a reason that does not favor a conception of the good, the policy is clearly non-neutral. It is worth noting here, even if it will be developed in following chapters, that the structure of neutrality as justification is that of an action (state’s) being grounded in reasons independent from subjective preferences or interests. That is one way to express the structure of the belief that underwrites modern reason.

6.b Justification

Bruce Ackerman argues that the legitimacy of a power structure depends on its ability to answer the question of legitimacy with reasons that satisfy these two conditions of neutrality:

⁵² *Ibid.*

⁵³ Jennings, *Op. Cit.*, 43

No reason is a good reason if it requires the power holder to assert:

a) that his conception of the good is better than that asserted by any of his fellow citizens,
or

b) that, regardless of his conception of the good, he is intrinsically superior to one or more of his fellow citizens.⁵⁴

By appealing to justification, neutralists shift the focus of neutrality from the will (aim) to the sphere of language or, as Rawls puts it, “public reason.” Wall and Klosko define it as follows: “The state should not aim to do anything to promote any particular conception of the good, or give greater assistance to those who pursue it, unless a plausible neutral justification can be given for the state’s action.”⁵⁵ Ackerman calls this justification “a conversation,” and uses the expression “power talk”. This is relevant because conversations, reasons, or language pertain to the realm of inter-subjective relationships. neutrality of justification turns the debate about the power of the state into a debate about the conditions of rationality and dialogue, as can be seen in Habermas and his successors.

It is important to point out that the demand is not so much that certain conversations take place explicitly, only that some policy is justifiable in terms of these conditions. This interpretation of neutrality, which aims at preventing the state from interfering between different conceptions of the good, and which intends to be devoid of any metaphysical commitments, depends on a certain notion of rationality and the subjects who use it. How this rationality and these subjects will remain neutral is the key to the whole theoretical edifice of these strands of liberalism. Once we turn the debate about neutrality into a debate about what criteria justify state action, the arguments must revolve around notions such as public reason, reasonability and permissible

⁵⁴ Bruce Ackerman, *Social justice in the liberal state* (New Haven: Yale University Press, 1980) ,11

⁵⁵ Steven Wall and George Klosko, *Op. Cit.*, 8.

dialogue, which means that the constraints placed on the state are also the constraints on the speech acts of the individuals who participate in the political world. To say, with Habermas, that “only the ideologically neutral exercise” of constitutional governments “can ensure that different communities of belief can coexist” in an egalitarian and tolerant way implies that while at the level of the “informal conflict of opinions within the political public arena” religious views must be admitted out of “epistemic” respect, at the same time, “all laws, all judicial decisions, and all decrees and directives” must be justifiable “in secular terms.”⁵⁶

The limits of the state become limits of the individual’s reasons. This is why all neutralists, most notably Rawls, necessarily invoke a notion of the ‘person’ in order for their political conceptions to hold. Although Rawls claims that his is a political conception of the person, as opposed to a metaphysical or “comprehensive” conception, the essential thing here is to determine whether this distinction is anything more than an arbitrary, *ad hoc* solution.

Once this limit is in place, a further question must be raised: to what extent are the notions of reasonability, public reason, or justification flexible or broad enough to avoid the endorsement of a particular comprehensive idea of the person and its rational powers? How can it not be an expression of historical background, —hence an imposition of majority opinion and the status quo, (as Michael Della Rocca suggests)?⁵⁷ Sometimes, as Gaus puts it, the implicit method here appears to be “to select a conception that seems intuitively plausible in the sense that we can imagine that the principle could be followed by a state that does most of what our correct liberal governments do.”⁵⁸

⁵⁶ Jürgen Habermas, *Introduction to Between Naturalism and Religion*, trans. Ciaran Cronin (Cambridge, UK: Polity Press, 2008), 3

⁵⁷ Michael Della Rocca, *Op. Cit.*, 178.

⁵⁸ Gerald F. Gaus, “Liberal neutrality: a Compelling and Radical Principle”, in *Perfectionism and neutrality* Ed., Steven Wall and George Klosko, (Oxford: Rowman and Littlefield, 2003), 138

At the other extreme, if it is not sufficiently broad and flexible, how is it possible to prevent unfair discriminatory policies masqueraded as independent, reasonable justifications? These questions can be answered through concepts adjacent to the principle of neutrality that involve ideas of moral motivation. To endorse neutrality of justification —the position taken by the majority of neutrality theorists— requires a careful balancing act along the borders of all the qualities of a comprehensive view: morality, anthropology, rationality. Often, the discussion of whether they manage to transcend metaphysics or epistemology becomes a semantic debate on the meaning of these labels themselves. We will return to address these debates in Chapter II.

6.c Effects

What neutrality of justification leaves aside is the question of the effects. After all, a policy satisfying the conditions of neutrality of justification does not guarantee neutral consequences, even if the consequences are not entirely attributed to a given policy as a cause. In other words: even if we assume that the law prioritizes rights over certain conceptions of the good —this distinction being rather problematic—a legitimately enforced law can still have the effect of promoting certain conceptions of the good over others.

On the other side of this third dichotomy, there exists a possible interpretation of neutrality that focuses on neither aims nor justifications, but on the consequences of the actions, or principles, of the state. In Wall and Klosko's definition, "The state should not do anything that has the effect — whether intended or not— of promoting any particular conception of the good, or of providing greater assistance to those who pursue it."⁵⁹ The main problem with upholding neutrality of

⁵⁹ Steven Wall and George Klosko, *Op. Cit.*, 8. It is a description, not an endorsement.

effect is that it leaves liberals with nothing but “impractical policies,”⁶⁰ especially when it is backed by absolute or comprehensive commitments to neutrality. This would have the effect of completely paralyzing the state.⁶¹

This paralyzing effect is due, in part, to the fact that neutrality is a device for restraining the state, not a description of the world or of a given society. There is no “independent describable condition of society to be called ‘neutral’ that the ideal political neutrality requires a government to maintain.”⁶² If there were a state in which every citizen is equally able to pursue and promote her own ideal of the good, as Raz’s perfectionism so desire, then we would not be talking about neutrality of effect, but either about the Nozickian ideal of non-intervention and the minimal state or a perfectionist view aimed at promoting an explicit ideal of autonomy, which presupposes a comprehensive conception of the good. Rawls rejects such an ideal because, he contends, not all ideals of the good are permissible, but only those that stay within the limits of his political conception of justice. Since this limit implies coercion, liberals must confront the challenge of justifying it. Hence his defense of neutrality of justification.

6.d Consequences of Disregarding Effects

What neutrality of effect prohibits is anything that would tip the balance of existing conceptions of the good, but it does not require the state to preserve that balance. If the balance shifts without any state influence, the state is not responsible. This is what Nagel calls negative responsibility, as opposed to the positive responsibility of actively producing or maintaining

⁶⁰ *Ibid.*

⁶¹ See Jennings, *Op. Cit.*, 45-47

⁶² Larmore, *Op. Cit.*, 44

neutrality in society.⁶³ Nevertheless, Nagel, Larmore, and Rawls reject even this weak interpretation of neutrality of effect. The fact that state action can be justified neutrally does not prevent it from causing certain conceptions of the good to lose influence; as Larmore notes, “gains entail losses.”⁶⁴

Against this position, those defending neutrality of effects may contend that a given action on the part of the state has an unfavorable effect and thus the neutrality of effect critique

cannot be defeated simply by establishing that a plausible neutral justification can be given for the state’s action. Perhaps, those who have been disadvantaged by the non-neutral effects of the state’s action are entitled to compensation. If so, then this is something that is captured by neutrality of effect, but not by neutrality of justification. A state might grant a group special rights or exemptions if doing so would help it prosper and if the group’s ability to survive had been set back by prior neutrally justified state action.⁶⁵

But this is only for exceptional cases, which are very hard, if not impossible, to delimitate. Nagel and Rawls offer reflections on the idea that a certain degree or kind of loss is an ‘unavoidable’ effect of the neutral state. As long as the discourse that justifies the policies that have those effects follows the rules of neutrality, the effects in question are warranted. Nagel says:

A state might adopt policies for other reasons which have the effect of making it easier for one conception to be realized than another, thus leading to growth in adherence the one as opposed to the other. (...) Hence it will not *be neutral in effect* among conceptions of the good, though it is based on impartiality among those conceptions, and avoids appealing to any of them to justify the use of coercive power.⁶⁶

This game of gaining and losing adherence is not only unavoidable because of liberalism. Rawls believes that it is the condition intrinsic to all social systems. Thus, the fact that his position on

⁶³ Thomas Nagel, *Equality and Partiality* (Oxford: Oxford University Press, 1991), 167

⁶⁴ Charles Larmore, *Op. Cit.*, 44

⁶⁵ Steven Wall and George Klosko, *Op. Cit.*, 8-9. Wall and Klosko do not endorse these definitions, they are only describing them.

⁶⁶ Thomas Nagel, *Op. Cit.*, 166.

justification implies a certain loss is an imperfection that comes from outside the theory and, if anything, that the theory tames:

We may indeed lament the limited space, as it were, of social worlds, and of ours in particular; and we may regret some of the inevitable effects of our culture and social structure. (...) There is no social world without loss; that is, no social world that does not exclude some ways of life that realize in special ways certain fundamental values .⁶⁷

We see a consistent response when these authors reflect on what causes the loss of certain fundamental values. On the one hand, fanaticism dies out because it is not allowed to influence the public sphere, or because it clashes with the force of democratic values —they claim. But that is not much of a loss. Now, of course, this depends on the idea of fanaticism or the idea of the unreasonable that one holds, but even if it were the case that the boundaries of fanaticism or unreasonability can merely be expressions of some higher commitments, commitments that remain hidden behind the strength of these and other words,⁶⁸ as I claim, this is generally uncontroversial. More problematic is the fact that some groups that can be deemed as reasonable according to the same definitions that these authors manage become weakened when different versions of state neutrality are in place.

Neutralists argue that the main reason for this is that these groups' comprehensive views of the good and their values depend on a certain degree of cohesion and cultural closeness, whereas liberal state promotes openness. It is the cost of freedom, they think. Since that openness is necessary for the liberal state to work correctly, citizens must be educated in it. The paradox they see is that openness is necessary for different views to coexist, but that same openness, when

⁶⁷ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 197

⁶⁸ A case in point might be when Rawls comments on the difference between rational and reasonable, and says that a mere self-centered rationality lacks moral motivation and thus "Rational agents approach being *psychopathic* when their interests are solely in benefits to themselves." John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 51 (emphasis added).

imposed as a political value, erode the necessary cohesion that different beliefs and social practices require to survive. Rawls argues that the education of children in “religious sects” that oppose the culture of the modern world “should also encourage the political virtues so that they want to honor the fair terms of social cooperation in their relations with the rest of society.”⁶⁹ Educating children in this way is educating them morally. Habermas says that the coexistence of various beliefs under the neutral state can be effective “only when their civic interactions are not based on a mere *modus vivendi*,” but “founded in conviction.”⁷⁰ This is a common challenge Habermas shares with Rawls: overcoming the danger of a *modus vivendi* liberalism. They fear it because they assume that neutrality may be too stringent. For it to function, according to Habermas, the democratic state must be “sustained by a legally unenforceable form of solidarity among citizens who respect each other as free and equal members of the political community.”⁷¹ They must believe with conviction in the liberal understanding of freedom and equality.

Rawls concedes that educating them in the political conception of liberalism may lead them to be educated in the tradition of comprehensive (i.e., metaphysical) liberalism, to which he can only answer that they must also be taught to understand the difference between a political and a comprehensive conception of justice or of the person.⁷² This is the original axiom of liberalism that we saw above, now applied to the narrative of openness versus closedness as the conflict that explains away the overlapping of the political with the existential spheres of discourse and action.

⁶⁹ *Ibid.*, 199.

⁷⁰ Habermas, *Op. Cit.*, 3

⁷¹ *Ibid.*

⁷² Rawls, *Op. Cit.*, 199-200

Nagel goes a step further when he acknowledges that some critics “doubt” that the claims of impartiality “are made in good faith” because the consequences of placing restraints on state power, as per the demands of “liberal toleration”, “appear with suspicious frequency to favor precisely the controversial moral conceptions that liberals usually hold;” so that this position “really disguise[s] a campaign to put the state behind secular, individualistic, and libertine morality – against religion and in favor of sex, roughly.”⁷³ Again, the narrative reinforces the idea that the cost of political freedom is the endangering of less “open” forms of life.

Larmore makes a similar point, yet in more explicit terms, in regards to minority cultures: “ways of life that depend upon close and exclusive bonds of language and culture –the French in Canada or the Bretons in France—may lose, within a liberal society also tolerating quite different and more open ways of life, some of the authority and cohesion that they would have if they formed complete societies unto themselves. Ideals of the good life that are open in principle to all, whatever one’s cultural inheritance (e.g., professional ideals) may well prove stronger under a liberal state.”⁷⁴ The “unavoidable” weakness of minorities, be them cultural, racial, sexual, etc, under the liberal state presents a particular challenge because, in principle, what makes them different is not so much how closed they are vis-à-vis majoritarian positions. Only their size seems to matter. In the same way that relative neutrality put the burden of proof over the shoulders of the dissenter, and thus it presents a conservative bias, neutrality of justification, when it comes to being indifferent to effects on minority groups, presents a bias in favor of majorities and hegemonic discourses and life-styles. Those discourses that are closer to the foundations of liberalism, a liberalism understood as a comprehensive view and not as the doctrine of neutrality

⁷³ Nagel, *Op. Cit.*, 156.

⁷⁴ Charles Larmore, *Patterns of Moral Complexity* (Cambridge: Cambridge University Press, 1987), 67

(as it happens with Rawls's *A Theory of Justice*), enjoy a greater chance to fulfill their goals and grow in number of adherents. The consequences of disregarding effects show that the presence of the principle of neutrality not only may embody a set of metaphysical commitments linked to a particular understanding of human nature, as I will try and show later, but also that it creates the conditions, the incentives, for that understanding of human nature to flourish: it shapes citizens after that take on human nature which, as I will show, is strongly dependent on ideas of spiritual freedom proper to the modern noumenal self.

7. An Alternative Narrative.

There is an alternative explanation to the narrative on the clashes between ways of life and the loss of authority and cohesion on the part of certain traditional communities. Perhaps it is not a question of more open values against more closed values, open societies versus closed groups, but rather, a clash between two ideals of community, which are equally closed or open. If the ideals that support liberal neutrality were as metaphysically charged as the values that Rawls sees in “religious sects”, or if the boundaries of the moral motivation of neutrality were as anthropologically determined as Nagel's non-secular critics are supposed to be, or if the English speakers in neutral Canada or the French-speakers in neutral France require as much non-neutral cohesion as the Quebecois or the Bretons do, then we would certainly be dealing with something else. Indeed, “goods are not all fully compatible” as Larmore says, but, also, the fact that “gains generally entail losses” —as he adds—would certainly have a meaning that has nothing to do with neutrality or freedom. Rather the force of these positions would be due to the brute fact of being the position of a majority. If this were true, then the impossibility of neutrality that we have been pointing to all along would suggest a different kind of debate that

would diverge sharply from the discussion about neutrality's practicality and priority that we have looked at in this chapter.

A possible explanation of how some views may become stronger than others has to do with the limiting character of neutrality. That some views become weaker is not always lamented from a liberal, democratic standpoint. When legislating from a non-liberal perspective—Buddhism, for example—would be seen as a form of coercion that would limit the rights we have come to regard as fundamental, exclusion from a ruling position makes sense from a neutralist position. That is the whole point of neutrality: to preserve our inalienable right to conduct our lives as we see fit. However, as we have seen, neutrality is not only a constraint applied to the state; it is also a constraint on the kind of reasons that can be employed to advance policies, so that some ideologies are prevented from being used to guide political action. I call them ideologies because of their political implications, but in truth we are talking about any view of the world, be it comprehensive or not. Some religious, communitarian, Marxist, feminist and individualistic views, seen as radical or unreasonable for their rejection of the implicit metaphysical ground of liberal neutrality, are all excluded once the conditions for entering the public sphere are established.

There is more to it. Liberalism also posits the state as a necessity; it asserts or assumes that the existence of the state is preferable to pure anarchism, or even to mere instability. In this respect, the goal of neutrality is not only to protect those fundamental rights but also to justify different degrees of coercion in order to protect those rights, precisely because without stability there are no rights. However, liberalism does not only reject anarchism for the sake of protection; it also rejects the libertarian minimal state out of a concern for justice and equality. The combination of

these two goals —freedom and equality— is the core of the riddle. It is hard to justify coercion in order to achieve these two goals, and the discussion among neutralists revolves around how to decide which formulation of neutrality is most effective on both accounts: how to advance what liberals have come to see as social justice while preserving individual rights.

While the method explicitly adopted for to solve this riddle is consensus, not everyone is invited to this consensus, and those who are invited are not always invited in equal ways. Assuming that everyone holds a belief system, a comprehensive view of the good, the strategy we have seen consists in admitting only ‘reasonable’ views to the consensus. Liberal neutralists seek for a political conception of justice that can be accepted out of the diversity of reasonable views. This is what Rawls calls “overlapping consensus.”⁷⁵ The problem is solved by introducing the need for acquiescence because once one accepts the reasonability of a certain notion of justice, his or her individual rights are protected. There is a link between accepting the limits of reasonability proper of a certain conception of justice and seeing one’s rights protected. However, since laws are created out of the dialogue that the diverse reasonable actors are supposed to have, or may ideally have, the way in which they may present their ideas for political action must be potentially capable of achieving some degree of consensus. Their reasonable comprehensive views must be translated into reasonable political conceptions. In the process, some things are lost; some comprehensive views are not reasonable enough and some political conceptions are too nakedly metaphysical to be admitted into the debate. The notion of reasonability becomes crucial: there must be enough material content in it so as to make liberal policies possible, but not too much so as not to impose a liberal comprehensive view. What would be the right amount of material content for liberal neutrality? If the principle of neutrality is articulated in its purest form, it

⁷⁵ John Rawls, *Justice as Fairness. A Restatement*. (Cambridge: Cambridge University Press, 2001), 32.

certainly “precludes most contemporary legislation” and if its too impure, it excludes various non-liberal views of the world from the public sphere.⁷⁶ There will always be compromises and hence losses, in one direction or another.

I suggest we classify liberal neutralists according to which compromises they are willing to accept as ‘unavoidable,’ and on which core principles they are unwilling to compromise. In between these two ends of the spectrum, there lies the plurality of views that articulate compromises that are acceptable but not unavoidable and principles that are acceptable but negotiable. We can understand their respective stances on neutrality by looking at the way they navigate this overarching spectrum. Concepts like ‘impartiality,’ ‘independent reason,’ ‘justification,’ ‘basic structure’ and so on, are used in formal ways to include and exclude certain material contents. However, there are consequences to doing so; liberal thinkers are willing to accept such losses as “unavoidable” only if the gains are worth it. The decline of a religion, the disappearance of a minority language, or the obliteration of non-modern ways of life are often deemed acceptable, while abandoning egalitarian policies or ‘fair terms of cooperation’ (Rawls’s term) is not. Because such losses are “acceptable” if there is a consensus, the role of the majority is key—as it should be in a democracy—just as much as who is given a voice in processes of consensus-building. What is rational? What is unreasonable? What is fanatical? What is unimportant? This is where the game is played. To what degree there exists a set of incentives to adopt the liberal comprehensive view built into this architecture?

This is why the question of whether neutrality is ultimately impossible is so important. We have seen throughout this chapter that the different positions on neutrality are each shaped by the

⁷⁶ Gerald F. Gaus, *Op. Cit.*, 160.

need to solve the problem of impossibility (meaning, the combination of liberalism's twofold goal: freedom and equality). For example, an absolute neutralist cannot uphold comprehensive neutrality and a comprehensive neutralist cannot but endorse a form of relative neutrality. This is also the reason why neutrality of effect is defended by none of the major authors who defend other versions of neutrality, even though the effects that they deem 'unavoidable' do not weaken their own beliefs. Only some religious people, cultural minorities, groups that feel discriminated against due to gender, race, and sexual orientation, or fringe communities are discontent about it.

I object to the narrative of openness versus closedness because I believe that some allegedly independent reasons are not as independent as they seem. An independent reason for, say, promoting a majority language or lifestyle is often the expression of cohesion of a majority around the societal backdrop of some mainstream culture. Now, if this is true, or at least 'reasonable,' it must mean that behind the political conceptions of liberal neutralists, behind the concepts of public reason, impartiality, justice as fairness, and so on, there are hidden comprehensive doctrines, and the attempt at neutralizing them, however formidable it may be (Rawls's system being the paradigmatic example), is failing. If this is true, neutrality is not only non-neutral but the vanishing point of an oppressive metaphysical doctrine. We must then look for the concepts that might be masquerading behind these doctrines beyond the debates that neutralists are consumed with. I think the best way to identify which concepts can help us test whether this is true—or at least 'reasonable'—is to examine the correlations that the different dichotomies that I have been describing create in the theories of the authors I have surveyed.

8. Impossibilities and further debates

Let us see if the differences we have surveyed point to the possibility of another, maybe more abstract debate.

FIGURE 1

Authors	On competition		On range		On coercion		On favoring	
	Relative	Absolute	Narrow	Comprehensive	Including taxes	Not including taxes	Justification	Effect
Rawls		X	X			X	X	
Larmore	X			X	X		X	
Barry	X			X		X	X	
Nagel							X	
Waldron		X		X		X	X	
Ackerman		X	X				X	
Dworkin	X			X			X	
Gaus		X		X	X		X	

In figure 1 above, I have systematized the three dichotomies that differentiate liberal neutralists and their positions on the topic of coercion, as well as the question of how coercion is defined. The first column divides those who believe that neutrality only applied to actual competition of views (relative) and those who believe that neutrality applied to any and all conceptions of the good (absolute), including potential conceptions that are not currently on the scene, as it were. The second separates those who believe that neutrality should only apply to constitutional essentials from those who believe that it should apply to all legislation. The third expresses the view on coercion: it distinguishes between those who believe that all government programs involve coercion—including incentives—because they are all funded by taxpayers, which are

compulsory, and those who believe that only those policies that carry the threat of violence are coercive are. Finally, the fourth expresses the fact that none of these authors support neutrality of effect, and all of them understand neutrality as a principle that applied to the reasons used to justify state action.

Let us analyze the correlations. The first thing we should notice is that there is a combination that is excluded: none of these authors believes that the principle of neutrality should have the following set of qualifications: 1) relative application (i.e., only to actual conflicts over the good); 2) narrow scope (covering only constitutional essentials and basic matters of justice); 3) negative conception of liberty as freedom from state coercion (i.e., policies carry the threat of violence); and 4) specific and strict focus on the process of justification rather than the content of what is being justified. This combination would be the weakest version of state neutrality. It must be said that it would also be the easiest to apply. At the same time, it would also allow for a great latitude in state action and, practically speaking, would almost serve as a warrant to implement classical perfectionist policies, such as policies that ensure the flourishing of personal autonomy and the availability of material options.

What really matters, however, are the other positions. Why is Rawls inclined to an absolute commitment to neutrality but only to the narrow scope of application, namely, to constitutional essentials, the basic structure of society, and matters of basic justice? His reason for doing so, as we have seen, is practical. Rawls's overlapping consensus is not an average of different comprehensive ideas of the good. Rather, it is a consensus around a separate notion —a “political conception”— that is ‘freestanding,’ meaning that is not metaphysically loaded. Then, citizens agree on the reasonability of this political conception from the perspective of their

comprehensive views. This consensus around a political conception informs the basic constitutional framework of a given society and demands of public officials the use of ‘public reason,’ a category in Rawls’s system that embodies reasonability paradigmatically. The rest of the legislation and state action must be constitutional, but it can favor certain goods if there is sufficient reasonable consensus that does not disrespect the fundamental rights and procedures established in the constitution.⁷⁷ If the same requirements —neutrality— were to be imposed on every state action, the state would be much more constrained than it is, and Rawls’s liberalism would become too formal not to be leaning toward libertarianism. This problem is in what makes an absolute and comprehensive commitment to neutrality impossible or impractical.

The same can be said of the opposite position. Larmore’s position is relative and comprehensive. In his conception, neutrality is watered down in the first dichotomy yet strengthened in the second. The state does not need to be neutral regarding all conceptions of the good life, but only with respect to those that are actually in dispute. “Where everyone agrees about some element of human flourishing, the liberal should have no reason to deny it a role in shaping political principles.”⁷⁸ Since this relative meaning of neutrality is just a way to solve actual conflicts and allows for some ideas of the good to be upheld, it can be applied to all legislation and state action. If his comprehensive commitment were to be applied in an absolute way, this would also be impractical. It is worth noting that Larmore acknowledges the non-neutrality of this position, insofar as he is content with the highest possible formulation of neutrality, assuming that a full commitment to it would make it “too empty to generate any substantive political principles.”⁷⁹ In

⁷⁷ This is highly schematic. In truth, some of these constitutional essentials include basic goods and Rawls’s notion of reasonability is laden with material principles, in particular the difference principle, as we will see in Chapter IV.

⁷⁸ Charles Larmore, *Op. Cit.*, 67

⁷⁹ *Ibid.*

addition, since “basic liberties and distribution” are also a goal of liberals (in addition to the goal of neutrality), we should be willing to suspend neutrality up to a certain point for the sake of those other goals. Otherwise, it would indeed be impossible or impractical. There is a similar pattern of compromises between dichotomies in Ackerman, Barry, and Dworkin. Barry, for example, is comprehensive when he says: “It would surely be leaving judges with too much discretion simply to state that principle in the constitution and let the courts decide on its application. Rather, the principle would form the basis for specific rules, prohibiting the state from interfering in religious worship, guaranteeing freedom of speech, and so on.” And he is relative when he says: “But, so long as there was sufficient agreement among the members of a society about the implications of social justice, these might be embodied in specific substantive rules and incorporated in the constitution.”⁸⁰

Waldron’s case is quite different. He is willing to accept a comprehensive and absolute commitment to neutrality, but only because he thinks that coercion properly understood refers only to those norms enforced through the threat of violence. Incentives, promotions, encouragements, education may be influential but they are not coercive. They can be excluded on this account. Otherwise, if every state action was considered an example of coercion, neutrality would become impossible or impractical.

Gaus is consistent in the extreme: his neutrality is absolute, comprehensive, and includes taxes and everything they fund as a form of coercion. But then he concludes that neutrality “precludes most contemporary legislation.”⁸¹ This is precisely what the other authors deem impossible or impractical. He believes it isn’t, and he admits it’s radical. But even in his case, neutrality is

⁸⁰ Brian Barry, *Op. Cit.*, 96

⁸¹ Gerald F. Gaus, *Op. Cit.* 160

understood as a principle of justification because the conflicts around the plurality of conceptions of the good is based on the kind of moral motivation that is sustained with a process of reasoning—of giving reasons—and he believes that the principle of neutrality itself is the outcome of a commitment to the freedom of thought, the freedom of giving reasons for our actions and preferences, not so much about consequences or effects.

What is relevant about these correlations is that these authors seem to agree on the difficulty of applying neutrality in its purest, strictest form. Neutrality can only be accepted if it's qualified in its applicability. Rawls calls the principle of neutrality “unfortunate”, “misleading,” with “altogether impractical principles” in “some of its connotations.”⁸² Thus, Rawls uses the “the familiar idea of neutrality” with precaution and “as a stage piece, as it were” to point out that when ‘justice as fairness’ excludes non-permissible conceptions of the Good (“those permitted by the principles of justice”) and aims at fostering “political virtues,” it is consistent with a formulation of neutrality of aim or justification. Rawls admits that ‘justice as fairness’ is not procedurally neutral, in the sense of setting up a procedure that is morally neutral. Its principles of justice are “substantive and express more than procedural values, and so do its political conceptions of society and the person,”⁸³ but nevertheless it is neutral in aim, because it forbids the state “to do anything intended” to favor “any particular comprehensive doctrine.”⁸⁴ Indeed, “admitting” the “virtues of civility and tolerance, of reasonableness and a sense of fairness,” into a political conception of justice “does not lead to the perfectionist state of a comprehensive doctrine.”⁸⁵ ‘Justice as fairness’ is neutral because it allows for ideas of the good to be “freely

⁸² John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 191.

⁸³ *Ibid.*, 192

⁸⁴ *Ibid.*, 193

⁸⁵ *Ibid.*, 194

introduced as needed to complement the political conception of justice” if they are translated into a “reasonable political conception of justice.” In this way, the presence of neutrality forces ideas of the good to be turned political and thus, in the assumed sequitur, non-metaphysical: “political virtues must be distinguished from the virtues that characterize ways of life belonging to comprehensive and philosophical doctrines.”⁸⁶ The distinction allows the state to “strengthen the forms of thought and feeling that sustain fair social cooperation” while being “very different from the state’s advancing a particular comprehensive doctrine.”⁸⁷ The question is then if these virtues, that presuppose “fair” cooperation, and that are grounded in a form of reasonableness that includes the conviction that citizens are, and ought to be, “free and equal”, are really disconnected from the comprehensive doctrine that turned them over the political realm. More fundamental is the question if stated thus: does the distinction between the political and the metaphysical make sense? Maybe it could, but the device that turns that possibility into a normative precept, namely neutrality, depends upon the establishment of virtues that are warranted because they are said to follow that precept by being “political.” On the one hand, then, the theoretical architecture begs the question: it tries to define fairness by appealing to fair attitudes, attitudes promoted by the system that is “fair.” On the other, it is circular: the separation of the political and metaphysical is axiomatically stated as a brute fact, but it is only workable as a practical distinction if citizens embody certain substantive virtues. These virtues are neutral because they accept that neutrality separates the realm of politics from that of metaphysics and personal morality. The separation depends on the neutrality of the virtues, and the virtues are neutral because they dependent on the separation. The strategies and differences that liberal neutralists express are ways of escaping this circle and answering substantively the

⁸⁶ *Ibid.*,195

⁸⁷ *Ibid.*,195

begged question (i.e., what is justice). The exceptions they allow to neutrality (its relative, narrow, non-coercive qualities, and its focus on justification), are the gate from where they enter in the public sphere the virtues that substantiate the theory. Notions of reasonableness, intersubjectivity (consensus), and autonomy, each with its correspondent chapter in this dissertation, are the *neutralized* forms of the substance of liberalism. Some authors, like Larmore or Nagel, are more ready to admit that these notions may carry some metaphysical weight, whereas others, such as Rawls, firmly believe that the separation of the political and metaphysical spheres holds water.

We must look into these notions (intersubjectivity, reasonableness, and autonomy), because the strategies and differences among neutralist theorists also show the limits of the principle of neutrality. In other words, the principle of state neutrality that these liberal authors defend is not, strictly speaking, neutral. It is just as good as it gets. After all, “the motive for higher-order impartiality is far more obscure than the motive for wanting everyone to have a good life.”⁸⁸ In every case, neutrality’s impossibility means that the state cannot promote, enforce, or apply certain goods that are considered too fundamental to be excluded without becoming non-neutral.

When the good is introduced in the system (be it an equality-based kind of good, or a good based on fostering autonomy and self-realization) it constitutes an exception of neutrality for the sake of ‘practicality’ and ‘basic justice.’ One possible interpretation is that these goods are not derived from neutrality; they are present precisely when neutrality is suspended. It is true that these goods are presented as basic, and they are sincerely thought to be the minimum. In this sense, the search for these minimal goods is shaped by the presence of neutrality: formal and material freedoms constitute a balance in these systems.

⁸⁸ Thomas Nagel, *Equality and Partiality* (Oxford: Oxford University Press, 1991), 156

The question we must ask under this interpretation is where these goods come from, if not from neutrality. What kind of justification is used to make them unexpendable. The danger, of course, is that they simply respond to the needs of a comprehensive, metaphysically-loaded, view of human nature and human community: what I have been calling ‘the alternative explanation.’ In this case, any appeal to neutrality, far from being a deterrent or a constraint, is in fact the mechanism to impose certain views of the good to non-liberals or to some members of the liberal family with other priorities. Which leads us to the second interpretation: maybe they are derived from neutrality. The very belief that there can exist a neutral sphere predicated upon the distinction between political and metaphysical realms, can already be seen as a stance on our relationship with our beliefs and how flexible is the relationship between our self and its comprehensive view. The notions of intersubjectivity, reasonableness, and autonomy would then constitute the constituents of the self that neutralist liberalism requires.

This dissertation considers both perspectives throughout because the dialectical contradiction (in a Hegelian sense) between neutrality and the substantive content of the idea of human nature that it ordains is twofold: on a reading of the texts of these authors that buys into its premisses, these notions seem to be affirmed as brute facts, and thus it can be said that they come from somewhere else. This is due to the reluctance to engage in metaphysical debate (Rawls’s “method of avoidance,” that will be analyzed in the following chapter). However, on a reading of these theories that tries to unveil the metaphysical commitments present in the three substantive notions mentioned above, it becomes plausible to see how the very idea of neutrality might be seen to demand those commitments and not others.

Our task, then, is to analyze the conceptual devices that are used to make neutrality consistent with the material goods that are introduced for the sake of practicality. Since what all of these authors have in common is the endorsement of the neutrality of justification, it seems clear to me that the development of the notions of reason and collective dialogue is the most clear candidate, together with autonomy. The epistemology that is assumed or claimed by these philosophers is the door to the meaning of the other concepts at play. And then, the bridge is the connection between this epistemology and the way it justifies moral standpoints. This is the bridge that is crossed by autonomy and intersubjective consensus, together with secondary notions such as moral motivation, reciprocity, fairness, objectivity, and other expressions used in the works of these authors, in order to make them consistent with neutrality.

I argue that some of the debates among neutralists and between neutralists, libertarians and perfectionists are mistaken. The question is not whether the state should promote ideas of the good (as perfectionists argue) instead of striving to be neutral, or disappear almost altogether (as libertarians claim). The relevant question is whether such a neutral, non-metaphysical view-from-nowhere is possible and practicable. If it is not, it does not necessarily mean that the grounds it protects might be the wrong political conception of justice, nor does it mean that it is not the most desirable. It just means that it is not neutral and that this should be acknowledged. If we did, maybe the problem of its effects would not be that unavoidable.

This chapter begins by asking what it means to *believe* in the liberty of the liberals. The philosophical form of this belief is the topic of the following one. Following the first interpretation, Chapter II begins by asking what does it mean to present metaphysical commitments within the logic of this debate. Following the second, it continues by arguing in

detail why the three notions of intersubjectivity, reasonableness, and autonomy, are what I call “metaphysical carriers.” In between, it tries and shows how these notions secure the building that neutrality coronates. Chapters III, IV, and V are each devoted to analyzing from the inside one of these notions, and their co-dependency with the principle of state neutrality.

Chapter II

Metaphysical Carriers

*Neque enim quaero intelligere ut credam,
sed credo ut intelligam*

Anselm of Canterbury, *Proslogion*, 1.

Introduction

Liberal authors have theoretical and practical reasons to avoid metaphysics. When they frame their theories within the limits of “postmetaphysical thinking” they do so because they don’t find any reason to see the ideas that metaphysics deals as anything but unscientific beliefs. The vocabulary of metaphysics has ceased to be useful in dealing with the problems they want to deal with politically. The practical reason for this obsolescence has to do with the connection between tolerance and stability. A political system that enforces a metaphysical position is set to clash with the plurality of metaphysical positions present in post-enlightenment societies, especially if they are like our present ones in the West, —i.e., somewhat democratic. This clash is relevant because a metaphysical position with respect to the real nature of the universe, (be it reductive naturalism, Christian theology, idealism, or any other sort), is seen to be the explanation for our understanding of our nature, our place, and the meaning of our life (or lack thereof). From that idea of human nature there come, it is assumed, our moral priorities. Religious, philosophical, cosmological views are seen to inform our views on justice by presupposing an anthropology. Since these views appear to be very important to the people who believe them, and history shows that we are ready to die and kill for them, it is more dangerous to try and suppress them than come up with a way to co-

exist peacefully. Toleration is not only a tool to secure stability, however, it is also an idea of fairness. It is fair to let everyone live according to their beliefs. That is why the beliefs admitted in the public domain are the ones willing to accept that they cannot impose their beliefs onto others.

The theoretical reasons to embrace post-metaphysics are also related to the practical reasons not to propose a metaphysical doctrine for the state. The impossibility to achieve a satisfactory level of scientific status for metaphysics fuels a form of skepticism that deems impossible to assert a *real truth* about the ultimate nature of the world, of human nature, and of justice. Since there is no theoretical way to decide which doctrine is superior to others, no doctrine may rule over others practically. This skepticism demands that the state operates skeptically: without endorsing any particular metaphysical view and creating the widest possible space for diverse doctrines to coexist and dialogue.

Liberal neutralists often conflate metaphysics with ideas of the good, comprehensive views of the world or the good life, even sometimes personal morality —as opposed to political morality, as we shall see. The reason is the relationship between belief and action. Believing in a certain nature of the world and the person determines, at least partially, our reasons for acting in one way or another both individually and collectively. But they also state that there is a way to organize public affairs around a conception of justice that will permit both a consensus between diverging metaphysical, cosmological, and moral beliefs, and an equal opportunity for everyone to pursue the goals in life that are determined by those same beliefs, as long as they are “reasonable.” What they offer is a division of spheres: political and non-political. In the political sphere there are certain constraints derived from the idea of neutrality. Such constraints include treating others under the

assumptions that rational conversation can potentially lead to agreement (i.e., intersubjectivity), that we can formulate our beliefs in such way as to expect that others will find them acceptable (i.e., reasonableness) and that our beliefs are both our own and affect our actions in a self-legislating way (i.e., autonomy). The argument that liberal neutralists want to make is that these assumptions, by virtue of being political and allowing for a wide range of beliefs to exist on equal foot, do not presuppose a comprehensive understanding of the good or a metaphysical position on being *qua* being, or a definite stance on the *Real* or the *Truth* of what exists, or a final end of life and the world. If they can offer a political theory that will not need to be grounded in anything of that sort, and that can be accessible and accepted from diverging views of that same sort, then their theory will be consistent with neutrality. At least, so long as the system that it designs prevents those who will not honor such restraint from ruling oppressively and impose laws out of their perceived intrinsic superiority.

The fundamental question is whether that political conception of the person that is assumed only in the political realm is actually just an instance of a broader anthropology that presupposes ideals of the good life and a metaphysical position on man and the world. Liberalism does have a metaphysical position, as Rawls himself recognizes¹. Markus Gabriel has identified *liberal naturalism* as the belief that human nature constitutes a *second nature* that escapes, through freedom, some of the determining causes of nature.² *Second nature* is the reason for the universal dignity that grounds an equal respect for everyone. That one form or another other of the idea of freedom as part of human nature is what grounds liberalism is, to an extent, uncontroversial. The relevant aspect is to

¹ Rawls, John, introduction to *Political Liberalism* (New York: Columbia University Press, 1993), xvii

² Markus Gabriel, *Transcendental Ontology Essays in German Idealism*, (New York: Continuum International Publishing Group, 2011) 3-6, and in particular the chapter "The Ontology of Knowledge." This insight will be developed at the end of chapter V. 6

know whether in the attempt at turning that anthropology into a neutralist political conception, liberal authors are successful in avoiding the metaphysical commitments that such an anthropology requires. I claim they do not. I want to show that the idea of neutrality, as it is described in these theories, is dependent on that anthropology and its metaphysical commitments. In order to do so, I must first show that while certain notions of liberal theory (like the three mentioned above), are embraced because they are deemed to be consistent with neutrality, and thus non-metaphysical, they are consistent precisely because they carry the metaphysical weight that the idea of neutrality presupposes. The attempt at ‘neutralizing’ the metaphysical aspect of liberalism only manages to hide it, but not to cancel it. This is why I proceed by analyzing first the problem with metaphysical positions, comprehensive views and morality in the political realm that justify the need for neutrality, and then move on to analyzing the role that the notions of intersubjectivity, reasonability and autonomy play in securing it. This chapter contemplates these notions as pieces in the system, from the outside, as it were. The three following chapters will be devoted to analyze each of these notions from the inside, so as to show the extent of their metaphysical commitments.

1. What do they mean by metaphysical?

1.a) Post-metaphysics

The organization of human beings living together in a society has historically featured the presence of ideas, principles, and norms undergirded by metaphysical commitments. From rational ideals of justice to divinely granted authority, from notions of the good to hereditary moral virtues, from “indisputable” truths about human nature to rational systems, the presence of the metaphysical tradition has been pervasive.

However, it is difficult to say precisely what we mean by metaphysical. Even in times when its vocabulary was common currency, the fuzzy borders of the discipline of metaphysics made it problematic and contentious —Kant, in a memorable turn of phrase, referred to it as the “battlefield” of “endless controversies” that’s “called metaphysics.”. This remains the case today, when the prevalence of “post-metaphysical” thought, especially in political and social theory, is only contestable from an eccentric, marginal position. Once one has overcome the temptation of saying that every theory is, in a way, metaphysical —because every position can be construed as entailing some metaphysical affirmation or negation— and of saying that nothing is, in truth, really metaphysical —because all metaphysical doctrines are nothing but contingent, historically tightened expressions of psychological or cultural fictions—, the borders of what can be called a metaphysical thought become a philosophical problem in itself. Sometimes metaphysics is presented as a collection of academic problems bound only by the fact that they are called ‘metaphysical’ (mereology, modality, causality, set theory, phenomenology, and so forth); this sense of the word “metaphysical” has to do with identifying its place in a taxonomy of academic disciplines. None of these perspectives concern us, as liberal thinkers think of it as meaning a commitment to certain theory of human nature and the good, even if this position is ultimately related to these more academic stances on metaphysics that postmetaphysical theorists try to avoid. That avoidance narrows the critique it can sustain from the affirmation or the negation of metaphysics in itself, as we will see.

Nevertheless, it seems to be of utmost importance for contemporary political and social theorists to be able to claim, and protest, that they don’t presuppose any metaphysical commitments. Indeed, the critique of traditional metaphysics is a trend that begins with modernity, the optimism

surrounding scientific progress, and the backlash against scholasticism. There exists a thread that ties Kant to Habermas and Rawls in that respect.³ Howard Williams has shown this thread very consistently, beginning with Kant's critique of the "super-sensible realm," continuing with Rawls' "political, not metaphysical" conception of justice, and concluding with Habermas' "post-metaphysical age."⁴ Kant actually does not reject the super-sensible. He is very interested in it, and even believes that the super-sensible – despite the fact that it cannot be known – can be spoken about universally. That's the sense in which Kant is not a post-metaphysical thinker. His position, e.g. on autonomy, actually needs a metaphysics. One of his most important achievements is to undermine all attempts to reduce the super-sensible (or certain noumena) to the sensible. Even if we understand the Kantian super-sensible as a product of reason (regulative, not constitutive), it is not for him a mere artifact or fiction. Kant ultimately rejects metaphysics as a theoretical science (object of knowledge), but not as a practically necessary field open to rational inquiry (again, in the practical sense). I think it is safe to say that he would reject the postmetaphysical pretensions of Habermas and Rawls. Still the kind of critique that Kant initiates is the point of departure of these liberal post-metaphysical thinkers. Habermas and Rawls, and other Kantian liberals, want to accept Kant's attack on theoretical, speculative metaphysics while doing without his metaphysics of reason that is entailed in his practical philosophy. Their political projects aim at salvaging the political intuition of the Kantian enlightenment without ending up trapped in the metaphysical aspects of it that appear to have fallen from grace since the second half of the twentieth century.

³ Even if Kant mocks those who pretend that they don't have a metaphysics as indifferentists.

⁴ Howard Williams, "Kant, Rawls, Habermas and the Metaphysics of Justice", *Kantian Review* Vol. 3. pp 1 - 17 (March 1999), 1

For Habermas, post-metaphysical thought tries to find a space ‘between naturalism and religion.’⁵ The “exercise” of a “secular governmental authority” that can “only” be “ideologically neutral”⁶ is predicated upon the possibility of giving a response to “the spread of naturalistic world-views and the growing political influence of religious orthodoxies.”⁷ Habermas is trying to go beyond naturalistic, positivistic reductionism while simultaneously keeping religions in check.

Fallibilistic, but non-defeatist postmetaphysical thought differentiates itself from both sides by reflecting on its own limits —and on its inherent tendency to overstep these limits. It is as wary of naturalistic synthesis founded on science as it is of revealed truths.⁸

Habermas restates the Kantian perspective by identifying the tension of his position as between this reflection on the limits of postmetaphysical thought and the tendency to overstep these limits. They find their balance in the giving and receiving of reasons. Not just any kind of reasons, and not just any kind of ‘giving’ or ‘receiving,’ as we will see.⁹ For now, what is telling about this post-metaphysical space is that it purports to be *post* because it can give up hard forms of truth, be they scientific or religious. It opens up a different set of standards based on a giving-and-taking of social reasons. Ultimately, it aims at having an impact on both naturalism and religion when they interact as world-views in the public sphere.

1.b) Comprehensive Views

When it comes to politics, metaphysics becomes an analogy of religious thought, theology, personal morality, cosmology, an abstract conception of the good, sometimes “ideology” or, in sum, a

⁵ Habermas, Jürgen, *Between Naturalism and Religion*, trans. Ciaran Cronin (Cambridge, UK: Polity Press, 2008)

⁶ *Ibid.*, 2

⁷ *Ibid.*, 1

⁸ *Ibid.*, 6

⁹ In Chapter III.1

comprehensive view of the world and the good life. This last expression is telling, and widely accepted, especially after Rawls' started using it and refining it for decades.¹⁰ "Comprehensive" is meant to capture the idea of what is broadest, of totality, of the whole, of the entirety of what exists,¹¹ but in a way that already implies an interconnectedness, a correspondence between different parts of a given universe. This interconnectedness explains why even though political theorists do not usually deal with this kind of comprehensive ontology, they do have to deal with conflicting ideas of human nature that most of the time assume some sort of overarching ontology.

This correspondence within the 'whole' is coextensive with 'meaning'. A comprehensive view of something is a grasping of the broad, general, regulatory meaning of that something, —even if that grasping is a negation of meaning, as it so happens with some fatalistic cosmologies.¹² This is the same totalizing perspective of the "mythical" way of understanding the world, the 'antithesis' of the modern, rationalizing view.¹³ Following Lévi-Strauss's anthropology of "archaic" peoples, Habermas points out that "myth constructs a gigantic mirror-effect, where the reciprocal image of man and the world is reflected ad infinitum[...]. By analogy, the whole world makes sense, everything is significant."¹⁴ Kołakowski, similarly, attributes to both religious and ideological belief systems the common quality of purporting "to impose an a priori meaning on all aspects of

¹⁰ See Brian Barry, *Justice as Impartiality* (Oxford: Oxford University Press, 1995), 20-23, and Chap. 5, for the uses of this expression and their debt to Rawls.

¹¹ Not exhaustively but conceptually: what can be included in the set 'being'

¹² After all, 'to comprehend' and 'comprehensive' both derive from the Latin *comprehendere*, meaning to take together, to unite, include, or seize, and -by extension-, to 'take in the mind', or perceive. The Latin prefix *com* expresses 'completely' and the lexema *prehendere* means to take hold of, or seize something. Of all the dictionaries where this explanation is found, the most appropriate to quote here is: Ernest Klein, *A Comprehensive Etymological Dictionary of the English Language*, (New York: Elsevier Publishing Company, 1971)

¹³ While the point of metaphysics is that it purports to be like a science, or rational knowledge, the point of postmetaphysical thinking is precisely to leave metaphysics in the realm of beliefs in practical reasons for the individual moral sphere, or voluntary associations, but not the state.

¹⁴ Jürgen Habermas, *The Theory of Communicative Action*, Vol.1: *Reason and The Rationalization of Society*, trans. Thomas McCarty, (Boston: Beacon Press 1984), 46

human life and on all contingent events [...] and no imaginable, let alone real, facts could refute the established doctrine.”¹⁵ Most of the times, this comprehensiveness is pregnant with a teleology: it is the purpose that shapes the meaning. Other times, comprehensive views can be said to reject a *telos*. neutrality aims at being agnostic with respect to a *telos*; an intrinsic purpose inscribed in nature or in life would also be comprehensive, and its negation would be as well.

This might be the post-metaphysical way of expressing agreement with Heidegger in that “every metaphysical question always encompasses the whole range of metaphysical problems. Each question is itself always the whole.”¹⁶ So a comprehensive view is not just any view, but the most inclusive one, expressing the most general meaning possible.¹⁷ The all-encompassing view could lack attention to specifics, but express the general or fundamental principles of a given theory of the *real*, and thus be comprehensive by assuming consistency.

However comprehensive it may be, it is nevertheless still a view. If the comprehensiveness assembled the fragmentary nature of perception and, arguably, of the universe (or being, or existence itself) the perspectival quality conveyed in the word ‘view’ sends it crashing back to its subjective, fragmented source. A view expresses a singular point whence that comprehensiveness is seized. A view is ambiguously positioned somewhere between thinking and seeing (between theory and perception), it can be rational or it can be irrational, it can be *a priori* or it can be *a posteriori*. But

¹⁵ Leszek Kolakowski, “Why and Ideology is Always Right,” in *Modernity on Endless Trial*, (Chicago: The University of Chicago Press, 1990), 232

¹⁶ Martin Heidegger, “What is Metaphysics?” in *Pathmarks*. Edited by William McNeill (Cambridge: Cambridge University Press, 1998), paragraph 3

¹⁷ An “understanding [of] how things in the broadest possible sense of the term hang together in the broadest possible sense of the term”, to quote Sellars’ famous words about philosophical inquiry. Wilfrid Sellars, “Philosophy and the Scientific Image of Man” in *Empiricism and the Philosophy of Mind*, (London: Routledge & Kegan Paul Ltd, 1963), 1

a view is personal, subjective, and, in principle, unshared—even if later it can be communicated—, because nobody can, so to speak, inhabit your eyes, to follow the allegoric theme. Of course, we don't have to inhabit each other's perception if we have the same perceptive faculty. This is, for example, the basis of Kant's claim that transcendental idealism is compatible with empirical realism. We (none of us) perceive or know the thing in itself, but we do have the same mental equipment and therefore can reject perspectivism. Nietzsche rejects this empirical realism with a deeper psychologism that entails perspectivism. But by using the expression “comprehensive view,” liberal neutralists stress the perspectival, subjective, and fragmented nature of our beliefs. Even if those beliefs are those of Kant, are somehow shared, or include some idea of universality and objectivity, they are still your view. Seeing, of course, as a metaphor of knowing, implies a sort of representational, picturesque schema,¹⁸—simultaneously eidetic (form) and enlightened (light)—. A “view” might emphasize also the cognitive a priori dispositions for knowledge, à la Kant. And in a post-Nietzschean world, a *view* is always and only a perspective; it is the quality of a thought that makes it mine.¹⁹ And precisely because it may not be shared *a priori*, “postmetaphysical” authors find it dispensable in public discourse. That explains why Thomas Nagel, in his attempt at salvaging objectivity for political theory, and neutrality for the state, uses the paradox ‘the view from nowhere’²⁰. And it is also the reason why social theorists who assert the sociality of reason use such a perspectival metaphor—a view—as the beginning of the conversation.

¹⁸ Not to mention the etymology of “theory.”

¹⁹ I am paraphrasing Leszek Kolakowski, *Metaphysical Horror*, revised edition (Chicago and London: The University of Chicago Press and Penguin, 2001), 26

²⁰ Thomas Nagel, *The View From Nowhere*. (Oxford: Oxford University Press, 1986)

In sum, a comprehensive view is a particular mental perspective on the whole-as-meaningful, sometimes expressing a reciprocally teleological relationship with one's understanding.²¹ A comprehensive view on the whole of what? Sometimes, authors use it with 'of the world' when they mean a cosmology—a natural order—or 'of the good life'—a moral order—or, simply, 'of the good'—when sometimes they refer to both: the good as the cause of the natural order.²² Returning to Heidegger's remark, however, it is not necessary that one express their view comprehensively in order for it to be an expression of a comprehensive view. An assertion of a single aspect of our moral, social or political life can be a sample of the whole of one's comprehensive view, even if only implicitly. Indeed, for one to endorse a comprehensive view, it is enough to presuppose a framework where our vocabularies make sense as a whole—and not as self-consciously contingent language-games à la Rorty—or to pose a moral landscape that both binds and excludes peoples and discourses. Specially, it is enough to assume an anthropology, to assert an idea of the person as such, able to explain axiomatically the behaviors and preferences that will later be deduced as empirical validity-proofs of truth or justice or induced as value-conferring selves.

1.c) Metaphysical Commitments

In order to see from which part of the whole this comprehensiveness can be ascertained in an allegedly post-metaphysical world, Williams' characterization of Kant's three basic conceptions of metaphysics is particularly useful:

²¹ (A similar tension to that of a comprehensive view is seen in the other widely used expression: 'conception of the good', which emphasizes the rational making of the content of the good, and emphasizes also the moral aspect of it, not excluding the idea of rational advantage as 'good'.)

²² See Klosco et al., *Op Cit*, for different uses.

First, he understands metaphysics as claiming knowledge of the super-sensible (through pure reason). This is what he subjects to thoroughgoing criticism in the *Critique of Pure Reason*. Second, he conceives of metaphysics positively, corresponding to his own project, as outlining the fundamental, a priori features of cognitive and moral awareness. The transcendental logic and the *Critique of Practical Reason* might be seen as examples of this. Third, Kant conceives of metaphysics in a heuristic sense as accounting for the proper regulative role of reason in terms of our knowledge and its genuinely constitutive role in terms of our practice (or moral life). The discussion of the ideals of reason and the *Metaphysics of Morals* might be regarded as instances of this.²³

It is enough for one to endorse the second or the third senses of Kant's conceptions of metaphysics in order for one's notion of justice, theory of democracy, anthropology, or sociological outline to be said to have metaphysical commitments, and thus for one to affirm a comprehensive view of the world, of the good life, or of the Good, in liberal neutralists terms. When it comes to analyzing the conceptions and ideas that are tacitly used or assumed by liberal neutralists, it is important to pay attention to the way in which those conditions for cognitive and moral awareness, as well as the regulative use of reason, might be at play in their justifications, even if their post-metaphysical stances reject *prima facie* the first understanding of metaphysics, already rejected by Kant himself.

A case can be made to separate metaphysics from the idea of comprehensiveness. Metaphysics as a science might be always comprehensive, but one can present certain metaphysical beliefs as not being comprehensive. For example, "metaphysics" sometimes describes supernatural beliefs (e.g., ghosts) that are not comprehensive. Similarly, naturalistic theories are comprehensive (or aim to be), but are held to be non-metaphysical. Liberal neutralists, especially those who rely heavily on the distinction between the political and the non-political to articulate their "political" conception of justice as something different from a "metaphysical" idea of the good, such as Rawls, most often conflate comprehensiveness and metaphysics. The reason is twofold. On the one hand, as

²³ Howard Williams, *Op. Cit.*, 2

explained above, when it comes to give an account of how human affairs should be organized, the presence of one aspect of a metaphysically implied system might be enough to create an oppressive regime. Interconnectedness explains this. But more importantly, if metaphysics is comprehensive, it seems that by offering an idea of the person that is only political, and only operates in the public sphere, liberal neutralists are avoiding the trap of comprehensiveness. They don't claim a truth about human nature in general, but only a political fiction to ordain public affairs. They are not comprehensive, hence they are not metaphysical—they seem to suggest. This possibility of turning one's comprehensive view into something merely political is what allows individuals and groups to partake in the public sphere. They need to translate their comprehensive view into a political view, and to do so, they must assume in the public sphere the features of the idea of the person that liberalism has managed to place at the center of the system, just for "political" purposes. Of course, the question is not only whether this distinction holds, but whether this distinction is an instance of a comprehensive view.²⁴

A comprehensive view as a basis for a political system is deemed dangerous because it assumes a truth based on a belief that is dogmatic, not open to criticism under any shared standard, and therefore necessarily sectarian and exclusive. The creed of our time is that when it comes to deciding how the general or basic structure of the ideal political system, it becomes all-important not to be sectarian or exclusive. Theoretically, this means that one must find a way to justify a political system without recourse to any theory that can be deemed metaphysical, and to do that, the "system" must refrain from interfering in spheres of human action that may be governed by

²⁴ Corresponding to the two interpretations of the critique of neutrality explained in Chapter I.7

comprehensive conceptions of the Good. neutrality means that this system must not be comprehensive.

However, political theorists devote much less time and ink to clarifying where a given argument starts to become metaphysical, to the extent that, as we shall see, the preferred method to solve this conundrum is called “the method of avoidance.”²⁵ The purpose is to be able to justify the political system without even entering into metaphysical debates or using its vocabulary. If a theory of justice can be described in its entirety without having to clarify a position for or against metaphysical doctrines then it can be said to be freestanding because it does not need to presuppose a comprehensive view of the real or a comprehensive view of human nature to hold.

It may be imaginable that a theory is presuppositionless; “freestanding,” it is rather unimaginable for a state, which is a being of some kind, to be presuppositionless or positionless. The method of avoidance of metaphysics assumes that connection: it is the freestandingness of the theory what allows for the state to be neutral. In not needing recourse to metaphysical parlance, or a theory of the whole, liberal neutralists claim to show that the state that their theories propose is available and acceptable to all reasonable metaphysical positions. The reason may be that any attempt at describing what is and what is not metaphysical will necessarily involve some metaphysical statement, or maybe, as Rorty claims, because all methodological debates about presuppositionless theories ultimately fall prey to circularity or question-begging.

²⁵ John Rawls, “Justice as Fairness: Political not Metaphysical”, *Philosophy and Public Affairs*, Vol. 14, No. 3. pp. 223-251 (Summer, 1985), footnote 22.

Every philosophical rebel has tried to be ‘presuppositionless,’ but none has succeeded. (...) To know what method to adopt, one must already have arrived at some metaphysical and some epistemological conclusions. If one attempts to defend these conclusions by the use of one’s chosen method, one is open to a charge of circularity. If one does not so defend them, maintaining that given these conclusions, the need to adopt the chosen method follows, one is open to the charge that the chosen method is inadequate, for it cannot be used to establish the crucial metaphysical and epistemological theses which are in dispute. (...) Every philosophical revolutionary is open to the charge of circularity or to the charge of having begged the question.²⁶

If one of liberalism’s main claims is that it is legitimate precisely because it avoids metaphysical commitments —and this is why the state ought to be neutral— there must be a way to show that that is the case. Given the apparent difficulty, and the common thread of presuppositionless rebels, what is the relevance of proving that there are metaphysical commitments in liberalism when it would apparently suffice with a classical critique of liberalism as an ideology or with an elegant demonstration of circularity and biased presuppositions?

Rorty’s quote assumes that these rebels are willing to “defend these [metaphysical] conclusions” and therefore the difference would lie with whether one does so with the help of the new method or not. But this is not what neutralist authors do. They deny the need for those metaphysical conclusions for their theories and affirm a principle of the state that has, as its sole content, the absence of those conclusions. Their systems and methods appear self-contained, silent about the conclusions that their methods imply and deaf to suspicions about metaphysical assumptions. This is why so many of their claims appear to be mere declamations or rhetorical maneuvers rather than robust arguments. When Rawls says in his “Political, not Metaphysical,” that a conception of the person from a political point of view “need not involve, so I believe, questions of metaphysical

²⁶ Richard Rorty, “Metaphilosophical Difficulties of Linguistic Philosophy” Introduction to *The Linguistic Turn, Essays on Philosophical Method*, Ed. Richard Rorty, (Chicago: The University of Chicago Press, 1967, 1992), 1-2

doctrine of the nature of the self,”²⁷ he is saying that he’s not going to get entangled in those arguments. And yet, the rhetorical strategy to argue for it remains an article of faith (“so I believe”). As Groff suggests, “the underlying thought is plausible enough: that one can ask certain kinds of questions without having asked, or implicitly answered, other kinds of questions.”²⁸ Unless, of course, the questions *are* the same type or species.²⁹

Neutralist authors, and in particular Rawls, claim to have developed a theory of the state that can accommodate any *reasonable* metaphysical view. So it is not so much that they adopt an anti-metaphysical stance, which would be easily attacked, but, rather, that they try and avoid referencing metaphysics while leaving open the door for an ‘overlapping consensus.’ This strategy is vulnerable to a critique that points out that everything is just a linguistic trick. If we should think of being “non-metaphysical” as a philosophical achievement, that is, through a rigorous methodology aimed at overcoming metaphysical presuppositions (Hegel, e.g.), then it is clear that that status is *not* achieved by merely avoiding any mention of metaphysics, or, rather what is commonly (i.e., “consensus”) deemed metaphysical. Neutralists are vulnerable to the charge that they clean up their language, omitting reference to metaphysics, shift the burden of proof to the critics, who detect metaphysical presuppositions in their theories, and then they simply ignore these critics or, scarcely better, dismiss them cursorily. In any case, the absence of an explicit anti-metaphysical claim is one of the reasons why a classical ideology critique is problematic, and must be refined.

²⁷ John Rawls, “Justice as Fairness: Political not Metaphysical”, *Philosophy and Public Affairs*, Vol. 14, No. 3. pp. 223-251 (Summer, 1985), 230-231

²⁸ Ruth Groff, *Ontology Revisited: Metaphysics in Social and Political Philosophy*. (London: Routledge, 2012), 1, and, generally, Chapter 1: “The myth of metaphysical neutrality”

²⁹ Let’s go back for a moment to Habermas’s point about steering a course between the Scylla of naturalism and the Charybdis of religion: the unexamined belief in a certain conception of the person (e.g, reasonable, etc.) is a non-neutral position on the central debates of our age about the person (e.g., Darwinian and Christian positions).

2. Ideology Critiques and Critique of Metaphysics

Ideology critiques are a matter of much debate and this is not the place to solve it. However, a couple of features of ideology critique may be useful for showing the relevance of a metaphysical critique of neutrality. In a chapter entitled *Why an Ideology is always Right*, Kołakowski draws a distinction between religion and ideology. In religious thought, faith “is interpreted as, and effectively is, a condition of a world-view within which empirical facts appear as so many manifestations of the hidden spiritual meaning.”³⁰ Faith is explicitly affirmed as the main basis for the hierarchical ordering of facts and interpretations. On the other hand, ideologies “are supposed to possess absolute truth and be testable at the same time.”³¹ This is key: ideologies *pretend* to be testable, but “they live on bad faith...they need a special psychological technique that prevents believers from seeing facts, or shapes a peculiar form of double consciousness.”³² This is a very extreme and pejorative view of ideology, but it captures something essential: the game ideologists and critics of ideology play is that of testing claims, that is, the game of social reason. This game can be played internally or externally, but it always accepts the very premise that religious thought negates, namely, the existence and possibilities of social reason for emancipatory reasons.

In *The Idea of Critical Theory*, Raymond Geuss says that if an ideology is described simply as a program—a set of ideas—it is uncontroversial to claim that “liberals do have an ideology—they have a general view of society and how it works, and more important, a general view about how it

³⁰ Leszek Kołakowski, “Why and Ideology is Always Right,” in *Modernity on Endless Trial*, (Chicago: The University of Chicago Press, 1990), 234

³¹ *Ibid.*

³² *Ibid.*

ought to work”³³. Whereas, Critical Theory, as “a reflective theory which gives agents a kind of knowledge inherently productive of enlightenment and emancipation,”³⁴ it is a research program “initiated by the observation that agents in the society are deluded about themselves, their position, their society, or their interests.”³⁵ However, if we are to see ideology in the “pejorative sense” and we ask in what sense a form of consciousness can be ideologically false, we can make use of the three kinds of questions that Geuss considers typical of Critical Theory, categorized a) in virtue of “some *epistemic* properties of the beliefs which are its constituents; b) in virtue of “its *functional* properties,” namely, dominion, repression, etc; and c) in virtue of its *genetic* properties, as in, for example, Nietzsche’s account of the origin of Christianity out of hate for noble, aristocratic values.³⁶

A critique of neutrality targeting its metaphysical commitments is not effective along these lines for two main reasons. First, because, according to (a), what is questionable about neutrality is not the presence of a belief, the epistemic value of which has been mistaken, as is the case with, say, positivist critiques targeting the unverifiability of some pseudo-empirical claim. Rather, it is the critique of the purported absence of a belief, an absence that turns out to hide a belief. As per (b), we still don’t know whether the prevalence of the principle of state neutrality is cause for some form of oppression: to assume so, would be to assume neutrality’s *raison d’être* and necessity: that metaphysical commitments are, in principle, oppressive. We cannot be guided by c) either because although it is illuminating to have a genealogy of the principle of state neutrality, and some references to it must be done, our aim is to explore the historical question that neutrality answers,

³³ Raymond Geuss, *The Idea of a Critical Theory: Habermas and the Frankfurt School*, (Cambridge: Cambridge University Press, 1981), 12, footnote 24.

³⁴ *Ibid.*, 2

³⁵ *Ibid.*, 12

³⁶ *Ibid.*, 13 and ss

not to undermine it by means of a contingent or allegedly vicious social origin, say, the self-interest of the bourgeoisie.

The second reason is general: this kind of critique would ultimately be external. For us, it would imply that we have to show that the world has a certain outlook in which neutrality is just a case of bad consciousness or a self-interested and ignoble lie. We would be buying into neutrality to the extent that we would be buying into the need for a neutral perspective from which things are objectively true—or intersubjectively true—and not a deception. This would force us to describe or accept a standard of rational discussion: to articulate a notion of enlightenment, and thus to create a sphere of what is debatable and how. This lays bare the key aspect of neutrality: its being neutral vis-à-vis comprehensive views is what allows the principle to draw the limits of the sphere of discussion; it tells us what is acceptable and what is not acceptable as a reason. It creates the conditions for consensus as much as the conditions for exclusion. The only way to show that the exclusion might be not neutral is by taking seriously the avoidance of metaphysics in the very same terms that neutralist authors put forth.

3. Pragmatism and Social Reason

In his *The Reconstructing of Social and Political Theory*, Richard J. Bernstein addresses the question of whether the so-called empirical theories of social scientists escape ideological bias—for example, when they talk about objectivity or factuality. He contends that...

an ideology is not any set of moral, social, or political beliefs and attitudes that informs and shapes an individual's (or a class's) interpretation of the world and his behavior. I reject this

excessively liberal conception of ideology because of its hidden relativism and nihilism. It suggests—and many think this is entailed by the concept of ideology—that all ‘belief-systems’ ultimately have the same epistemological status, that all are equally unjustifiable. According to this view, when we criticize a given ideology, we are supposing another ideological stance that we do not question. This permissive use of the concept of ideology obscures the fact that ideologies are based upon beliefs and interpretations which purport to be true or valid. These beliefs and interpretations are consequently subject to rational criticism.³⁷

In short, what Bernstein claims — he is explicit about it: he is not a relativist, he says— is that not all beliefs are the same. There are moral, social, and political beliefs and attitudes whose epistemological status is superior to that of others because they are held to be true or valid in a way that submits them to rational criticism. Since beliefs are at question they can be subjected to criticism and hence not all of them are on a par. This also allows their epistemological status to be secured, even if they are disproven and rejected at a later time. This what makes a belief or a set of beliefs an ideology. Presumably, then, there exist other beliefs that are similarly influenced by bias, but they are not ideological in that they are not open to rational criticism. This is the difference that we saw clarified above in the quote from Kołakowski on religion and ideology. Conversely, Bernstein also claims in this passage that we can criticize a given ideology without falling prey to another ideology. We must understand how ideology “both *reflects* and *distorts* the historical material conditions of social life, and grasp what factors influence and sustain the acceptance of an ideology”³⁸. Of course, this opens up the same question about the ideological bias of this perspective and the notions of rationality and objectivity that they might imply: do we need a standard of rationality or objectivity in order to be able to criticize without being subject to ideological bias? Bernstein thinks that we do not. Bernstein is not *strictu sensu* a neutralist, since he

³⁷ Richard J. Bernstein, *The Reconstructing of Social and Political Theory*, (USA: University of Pennsylvania Press, 1976), 108

³⁸ *Ibid.*, 108

contends, along with Charles Taylor, that theoretical frameworks, necessary for social science, are never value-neutral, and that even the “connection between the factual base and the valuation is built, as it were, into the conceptual structure” of social theories.³⁹ According to the quote, he is not a relativist either and believes in normativity. He also believes that theory and facts are value/norm laden. But does this mean that he is not a neutralist? When neutralist authors explain their systems, they also reject relativism, at least in two senses: they do not accept sectarian ideologies as valid for political life, and they place at centerstage political conceptions of justice that, by being freestanding, they can gather consensus from a variety of comprehensive views. Neutralists also believe in normativity. The role that reasonableness plays in filtering and reshaping the conceptions that can take part in ruling the state informs an implicate normativity. They also believe that theory and facts are value/norm laden, that is why they deny that one conception possesses the right facts and the right theory for government. neutrality is the recognition of the tendency towards bias, in good faith, that we all have. There must be a space between what the government can do and what we believe is the good thing to do; that space is neutrality. Can Bernstein’s position be described as viewing neutrality as a value? A liberal neutralist claims not to be a relativist nor a nihilist with respect to values. The account of social reason that justifies the prevalence of some values over others is both descriptive, in that it considers plurality and intersubjectivity a fact, and normative, in that the right kind of consensus that emerges from the actuality of social reason must be conducted following certain norms of communicative action (Habermas), rational conversation (Larmore), and constraints on power positions, such as the ones represented in Rawls’s ‘original position.’

³⁹ *Ibid.*, 109, (quoting Charles Taylor)

If a biased ideology is equated with Marxian ‘false consciousness’, the contrary notion of ‘true consciousness’ must somehow be demarcated, but that seems untenable given the impossibility of value-neutrality. Bernstein argues that “the *achievement* of ‘true consciousness’ is a regulative ideal of the critique of ideology” and “since every ideology is based on beliefs and interpretations that make the claim to validity, we can examine these claims to validity and show their falsity. We can show the falsity of an ideology without claiming that we have achieved a final, absolute, ‘true’ understanding of social and political reality.”⁴⁰ This “regulative ideal” position is very rich in possibilities. It could even be taken as a position in defense of the principle of state neutrality, since it assumes that the full application of neutrality is not desirable because it is at risk of falling prey to ideological bias, and be the cause and outcome of false consciousness. Trying to apply it fully might defeat the purpose of neutrality itself. A most neutral position implies not taking a fanatical stance on one’s beliefs in order to enter the public sphere, and this position must be applied to itself. The apparent paradox is that being fully neutral implies not being fully neutral. Since neutrality is a device of restraint, its absence in the political architecture it regulates is the presence of substantive content. The paradox of neutrality as a regulative ideal is this: the uncertainty and critical perspective that justifies it, together with a rejection of relativism, nihilism, and radical skepticism, leaves room for the *right* kind of substantive content (i.e., liberal content). It is the practical application of the standard that the regulative ideal *qua ideal* rejects theoretically. But unlike Kant, without offering an explicit metaphysics of reason to back it up.

Moreover, criticizing something as ideologically biased or as ‘false consciousness,’ even if it does not need a positive account of ‘true consciousness’ or a standard of objectivity, demands an account of

⁴⁰*Ibid.*

what counts as a critique. The question of what is acceptable as a critique runs parallel to that of what is acceptable in the public sphere. The demarcation of what is acceptable and what is not is what makes neutrality an operative principle and also what makes criticism valid or invalid even when no absolute standard of truth is given. Bernstein addresses this problem from a pragmatist perspective:

The attempt to specify a single univocal set of criteria as a basis for distinguishing what is genuinely objective from what is not, has been one of the most obsessive and futile preoccupations of modern thinkers since Descartes. But we are not there obliged –as it is often claimed—to retreat to skepticism, self-defeating relativism, or irrational subjectivism. The lesson to be learned, rather, is how difficult and complex it is to articulate the standards of objectivity relevant to different domains of inquiry, and the ways in which these standards are themselves open to criticism. Furthermore –*and this is perhaps the most important point*—the very standards of objectivity and rationality themselves depend on the existence of communities of inquirers who are able, willing, and committed to engage in argumentation.⁴¹

Bernstein’s “most important point” can also be construed as a metaphysical claim: the *existence* of some group, and the unique characteristics of that group, are what ground objectivity. This is very similar to having recourse to reasonability. There appears to be a hidden presupposition regarding the nature of the person and her capacities, let alone interests, for rational inquiry.⁴² This is relevant because the general kind of rational criticism that is applied to the biases and ungrounded assumptions of a given ideology has to be applied here to the specific case of neutrality. The possibility and impossibility of neutrality as an epistemological position lies at the very core of all ideology critiques. neutrality is predicated upon the same principles that ideology critiques are predicated upon—it wants to articulate the ability, willingness, and commitment to engage in argumentation—, and thus it forces any attempt at laying bare its biases to account for this awareness.

⁴¹ *Ibid.*, 111

⁴² Not just “person”, but group, society, language, etc

What we are after is whatever is not open to criticism. In order to see whether liberalism holds its central promise, we shall want to see whether there are moral, social or political beliefs that are not open to criticism, so that their axiomatic nature affirm a certain outlook of the (moral, social or political) world as such. Of course, everything is open to rational critique, but what Bernstein is calling here 'ideology' is a specific form of belief that is explicitly open to critique. Secondly, in the content of this concept of critique there lies part of what is in need of examination. The notion of the "community of inquirers," common in pragmatism since Pierce,⁴³ already assumes the validity of intersubjective consensus as a substitute for classical and modern metaphysical foundations. I would fall prey to circularity if I tried to prove that intersubjective consensus is somehow ideologically biased using the standard of rationality that "depends on the existence of communities of inquirers," that is, the same type of community that neutrality tries to articulate and delimit. It is this standard itself that is subjected to scrutiny.

Bernstein's understanding of ideology clarifies something that is at the core of this critique: In what sense must a critique of neutrality be a critique of metaphysics? In the sense that the ideology of neutralist authors consists of being 'freestanding' with respect to metaphysical commitments, and their principle of state neutrality embodies institutionally that avoidance of commitment. Precisely for this reason, only a critique able to show those metaphysical commitments will be able to show the moral, political and social consequences the principle of neutrality produces.

⁴³ I have explored this notion in Jordi Graupera, "Dependencies between Logic and Community. Philosophical Implications of Peirce's Categories for Praxis" *Ramon Llull Journal of Applied Ethics*, Issue 2, 179-194 (Barcelona, 2011)

In principle, neutrality does not have any other content but the emptiness of its perspective. For that reason, all systems built with reference to a non- or post-metaphysical neutrality must provide an account of social reason as their only source of truth and value. At the same time, social reason is the only tool left at our disposal in a post-metaphysical world for criticizing neutrality. Neutrality highlights the contingency of the outcomes of social reason, and social reason seems to demand some sort of horizon—such as neutrality—for its actuality to legitimately claim to embody enough diversity, freedom of thought, tolerance, and ultimately sufficient exchange of perspectives for any valuable knowledge to arise. Neutrality is the fundamental presupposition in these accounts of social reason—at least until they explicitly establish the rules of rationality that assess the legitimacy of a critique under a given paradigm.

4. Kołakowski's and Rorty's external critique

While Rorty believes that metaphysics is just another “vocabulary” that can be and must be left behind,⁴⁴ Kołakowski believes that even if the answers to metaphysical questions are always too problematic to hold water, the questions are eternal and unavoidable, even for those who claim to avoid them.⁴⁵ For Kołakowski, “metaphysics—as a search for a self-rooted Being—has survived, relegated to life in a kind of philosophical demi-monde,” even if “its language has been largely de-legalized.”⁴⁶ Kołakowski also distinguishes between myth and metaphysical theories: “Myths are not translatable into some non-mythical language that is supposedly better at conveying the

⁴⁴ Richard Rorty, “Cultural politics and the question of the existence of God” in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 3-26.

⁴⁵ Leszek Kołakowski, *Metaphysical Horror*, revised edition (Chicago and London: The University of Chicago Press and Penguin, 2001), sections 1-9.

⁴⁶ *Ibid.*, 80

genuine content to us. (...) If myths had metaphysical equivalents, they would be dispensable.”⁴⁷ For him, part of the problem of metaphysical answers is their inability to come to terms with the limits of rationalization—problems like infinite regress, arbitrariness, etc. Myths express some of the things that unavoidably lead us to metaphysical questions, even when we try and translate them into non-metaphysical theories.⁴⁸ The unavoidability of such questions, the limitations of all answers—both metaphysical and anti-metaphysical (skeptical, relativistic, nominalistic)—, and the presence of the ineffable that art and religion express, indicate that the attempt to separate those spheres of thought, belief, and life from the public sphere of politics, arbitrated by neutrality, is not only impossible but contradictory.

Most strikingly, both Kołakowski and Rorty reject neutrality for similar reasons: because neutrality is the precondition for metaphysics, and not the way to avoid it.⁴⁹ For Kołakowski, the very idea that people look for meaning, that we are natural hermeneuticians or that we construct and interpret comprehensive views is not ontologically neutral.⁵⁰ It presupposes a reciprocal relationship between the mind as a revealer of meaning and as a container/creator, “itself endowed with meaning.”⁵¹ The existence of diverse perspectives demands, for them to be communicable and understood, some recourse that transcends self-reference paradoxes or paralyzing skepticism. The attempt to account for all these perspectives as perspectives and nothing else, to categorize them according to this abstract formality, i.e., their capacity for meaning, is not

⁴⁷ *Ibid.*, 95

⁴⁸ *Ibid.*, 95-96

⁴⁹ Rorty’s argument, however, can be seen as rejecting neutrality only because he is very worried about neutrality. Since he cannot accept any metaphysics – because metaphysics will be not liberal, like a religion imposed on politics – he abandons neutrality and asserts cultural non-neutrality

⁵⁰ Kołakowski, *Op. Cit.*, 120 and ss.

⁵¹ *Ibid.*, 124

neutral either. Especially when it comes to articulating them, as neutralist liberal authors attempt to do, it is precisely the ambition to be neutral that makes their descriptions and understanding metaphysical: “there is no angle that opens up all perspectives simultaneously, unless it coincides with the divine eye.”⁵²

From the other side of the mirror, Rorty believes that most of our philosophical talk, including the question of whether or not to talk “about the existence of immaterial and infinite beings” should be “turned over to cultural politics,”⁵³ where our conversations will be hierarchized according to a set of humble political goals, —not to be “cruel” to one another, mostly.

There is no problem in giving either ‘what it is like to be conscious’ or ‘God, a being without parts or passions’ a place in a language game. We know how the trick is done, and we have had lots of experience watching both games being played. *But in neither case is there any point in raising questions about existence, because there is no neutral logical space within which discussion can proceed between people inclined to deny and people inclined to affirm the existence of the relevant entity.* Metaphysical questions like ‘does God exist?’ and ‘is the spatiotemporal world real?’ are undiscussable because there is no list of ‘neutral’ canonical designators by reference to which they might be answered.⁵⁴

Behind this claim there lies the idea that there is no such thing as a neutral space, perspective, or universe of meanings within which we can discuss questions of truth, existence, or reason definitely. In other words, any space to discuss them isn’t neutral. There is only a sort of conversation in which meanings change according to interest, perspective, contingencies. The only meaningful questions we might pose are those that don’t bother about ultimate meanings. Rather, we ask about what is best for us to discuss in order for our responsibility to ourselves and to others to be best accomplished, so we can fulfill our goals and —presumably— be less cruel.

⁵² *Ibid.*, 109

⁵³ Rorty, *Op. Cit.*, 19

⁵⁴ Rorty, *Op. Cit.*, 20 (my italics)

Politically, this entails a severe critique of these liberal attempts to organize a conversation among different groups or individuals according to a set of limits imposing neutrality on the public sphere, under the label of reasonability or other equivalent moral-epistemic values. Neutralist liberals agree with Rorty that these controversial metaphysical questions are unresolved because we do not possess a set of epistemic principles that can tell us when and how certain fundamental positions are true, so everyone arrives at a different conclusion about it, and that is the reason why the state must endorse pluralism as an inescapable fact of human communities. Where Rorty and these liberal authors diverge is in how to reintroduce neutrality in the political sphere to arbitrate these positions. Rorty is adamant in his opposition to Rawls's uses of reasonability and Habermas's endorsement of the possibility of a 'better argument' that will always win a theoretical debate.⁵⁵ He sees these positions as remnants of a universalistic, metaphysical tradition, and urges us to "peel apart Enlightenment liberalism from Enlightenment rationalism,"⁵⁶ which he thinks is still pervasively present in the work of these neutralist liberals; indeed it attempts to enjoy the fruits of metaphysics without assuming the appropriate responsibilities.⁵⁷

Just as Kołakowski saw that the attempt at opening up all perspectives is the mark of a God's-eye view, Rorty believes that "we should stop trying to put our discursive practices within a larger context, one which forms the background of all possible social practices and which contains a list of

⁵⁵ In particular in Richard Rorty, "Justice as larger loyalty", in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 42-55

⁵⁶ *Ibid.*, 55.

⁵⁷ Richard Rorty, "The Priority of Democracy to Philosophy." In Rorty, *Objectivity, Relativism, and Truth* Philosophical Papers Vol 1 (Cambridge: Cambridge University Press, 1991).

‘neutral’ canonical designators that delimit the range of the existent once and for all.”⁵⁸ What is the larger context he’s referring to? In the case of liberal systems, by means of the principle of toleration, constitutions are meant to create a stage for this background. It surely is a larger context with respect to each discursive practice. But it is in the theories of neutralist authors that this attempt becomes clearer. Rawls himself speaks of the “background culture” from which the canonical designators are derived in order to build a system of legitimacies for political discourses. Time and again, neutrality tells us what can be included and what must be excluded from the political conversation and public justifications. Liberal theorists create (or claim to identify) these neutral canonical designators—reasonableness, consensus, autonomy—that delimit the range of the politically relevant. Rorty says that “if there were such a context, it would of course be the proper object of study of an expert culture charged with determining the future direction of the Conversation of Humankind.”⁵⁹ This is exactly what neutrality is supposed to do and what authors such as Rawls believe that philosophers should be doing. As Wolterstorff critically (and rather ironically) describes it: “Let the political theorists among us analyze the political culture—the political ‘mind’—of our liberal democracies, with the aim of identifying the fundamental organizing ideas of that culture.”⁶⁰ For his part, Rawls says that theorists must “elaborate” or “unfold”⁶¹ those ideas into principles of justice, principles “specifying the fair terms of social cooperation between citizens regarded as free and equal⁶²” as the apparently *largest possible* context.

⁵⁸ Richard Rorty, “Justice as larger loyalty”, in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 24.

⁵⁹ *Ibid.*

⁶⁰ Nicholas Wolterstorff, “The Role of Religion in Decision and Discussion of Political Issues” in Robert Audi and Nicholas Wolterstorff, *Religion in the Public Square: The Place of Religious Convictions in Political Debate*, (New York: Rowman & Littlefield Publishers, Inc, 1997), 92.

⁶¹ Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 27 as quoted by Wolterstorff, *Op Cit.*, 92

⁶² *Ibid.*, 4, as quoted in *Idem*, 92.

For Rorty, unsurprisingly, “there is no such context,”⁶³ because, precisely, “‘Ontology’ is not the name of an expert culture, and we should stop imagining that such expert culture would be desirable.”⁶⁴

The external perspective that Kołakowski and Rorty occupy points to the work that neutrality does in securing metaphysical theories, what I called “the second interpretation” of the critique of neutrality. Rather than avoiding metaphysics or failing in leaving some metaphysical content out, neutrality would be securing grounds for a metaphysical theory, or, in other words, neutrality would be a metaphysical idea. From these two external perspectives, Kołakowski’s fideism and Rorty’s relativism, the reason why we cannot have metaphysics is precisely because a neutral standpoint is not possible, and thus the affirmation of such a standpoint in itself already ordains human priorities according to a metaphysical agenda. In analyzing the “canonical designators”, the notions that support the context that neutrality create, we will try and show that this is partially right, but without the need for radical relativism or fideism.

5. World-views in Habermas

Habermas draws a contrast between mythical and modern ways of understanding the world. He thinks that what “irritates us members of a modern lifeworld is that in a mythically interpreted world we cannot...make certain differentiations that are fundamental to our understanding of the world” because the mythical understanding shows a “peculiar *confusion between nature and culture*,” as

⁶³ Richard Rorty, “Cultural politics and the question of the existence of God” in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 24

⁶⁴ *Ibid.*

Durkheim or Lévi-Strauss also believed.⁶⁵ On the other hand, we, modern rationalists, “make the correct conceptual separations between casual connections of nature and normative orders of society” which also allows us to differentiate “between *language and world*; that is, between speech as the medium of communication and that about which understanding can be reached in linguistic communication.”⁶⁶

The intersubjective space, teeming with formal world-concepts, is what makes it possible for language and world to be linked through world-views that are open to criticism. “We proceed from those *formal presuppositions of subjectivity* that are necessary if we are to be able to refer to something in the one objective world, identical for all observers, or to something in our intersubjectively shared social world.”⁶⁷ These world-views are seen as culturally, historically, i.e., contingently informed, and this is why they are open to criticism and revision:

Validity claims are in principle open to criticism because they are based on formal world-concepts. They presuppose a world that is identical for *all possible* observers, or a world intersubjectively shared *by members*, and they do so in an abstract form freed of all specific content. Such claims call for the rational response of a partner in communication. (...) The contents of a linguistic worldview have to be detached from the assumed world-order itself. Only then we form the concept of cultural tradition, of a temporalized culture, whereby we become aware that interpretations vary in relation to natural and social reality, that beliefs and values vary in relation to the objective and social worlds.⁶⁸

It is this formalization, this abstraction from content, that allows each member to presuppose an identical world for all possible observers or a world shared intersubjectively. It is because this abstract form is freed from all specific content that a given claim can solicit a rational response from

⁶⁵ *The Theory of Communicative Action*, Vol.1: *Reason and The Rationalization of Society*, trans. Thomas McCarty, (Boston: Beacon Press 1984), 48, (italics in the original)

⁶⁶ *Ibid.* (italics in the original)

⁶⁷ *Ibid.*, xx, (italics in the original)

⁶⁸ *Ibid.*

another. This account of rationalization —discursive rationalization— is supposed to do the same work that neutrality does: it offers formal procedural rules in exchange for specific content. But, as we see in the last part of the quote, this comes at a price: all *members* of the intersubjective rational space must detach their linguistic worldview from the assumed world-order, or, in other words, their claims from their beliefs. Once that linguistic abstraction is made, once the post-metaphysical frame is bought into, once the linguistic-pragmatic turn is accepted, we can form the notion of cultural tradition as something historical, temporal, contingent, “whereby we become aware that beliefs and values vary in relation to the objective and social worlds.” To participate in this rational conversation means to make one’s beliefs relative and contingent—in other words, to stop taking them seriously.

In contrast, the mythical space is a space for bare belief, and as such, closed off to criticism, because mythical world-views “prevent us from categorically uncoupling nature and culture, not only through conceptually mixing the objective and social worlds but also through reifying the linguistic worldview. As a result the concept of the world is dogmatically invested with a specific content that is withdrawn from rational discussion and thus from criticism.”⁶⁹ The inferiority of a dogmatical, mythical world of bare belief vis-à-vis the modern world of discursive rationality is based on three characteristics: first, an account of the external, objective world; second an account of subjectivity; and third, an awareness of the problematic dialectic between objective and subjective spheres, a self-reflective attitude that is able to see the account of what’s objective as a culturally bound product of the historical time.

⁶⁹ *Ibid.*, 49

The modern worldview needs an objective standard because:

Only against the background of an objective world, and measured against criticizable claims to truth and efficacy, can beliefs appear as systematically false, action intentions as systematically hopeless, and thoughts and fantasies, as mere imaginings. Only against the background of a normative reality that has become autonomous, and measured against the criticizable claim to normative rightness, can intentions, wishes, attitudes, feelings appear as illegitimate or merely idiosyncratic, as nongeneralizable and merely subjective. To the degree that mythical world-views hold sway over cognition and orientation for action, a clear demarcation of a domain of subjectivity is apparently not possible.⁷⁰

The quality of the subjective is dependent upon and derivative from the objective. Beliefs, actions, and thoughts can appear false, hopeless, or imaginary because the external, objective, natural world, confronts them as a standard they can fail to live up to. In addition to the distinction between objective and subjective, we also have here a distinction between the normative and the subjective. Similarly, autonomous social norms are imposed upon moral states of mind, and thus correct their nongeneralizable, “merely subjective” defects. Such a sharp, Enlightenment distinction between nature and culture, and between objective and subjective, demands a justification without metaphysics, if it has to be consistent with Habermas’s project of detranscendentalization.

Habermas’s justification consists in talking about claims, either to truth and efficacy in the natural world, or to normative rightness in the moral world, and assume their criticizability. “Objective world” and “normative reality” are measured against these criticizable claims, and through them, they are transformed from opposite worlds into complementary ones. The external world “is defined by its being shared with others” and the social world of normativity is “presupposed in common as the totality of all interpersonal relations that are recognized by members as

⁷⁰ *Ibid.*, 51

legitimate.”⁷¹ Even those experiences that belong to the domain of subjectivity (because only the individual has (privileged) access to them) can be shared, via “expressive utterances with a criticizable validity claim” and thus subjective worlds “can also be drawn into public communication.”⁷² All this is very questionable. Since the others are part of the external world that is defined for us by its being shared with the same others, there needs to be a distinction between the nature of the external world and the nature of the others, as hermeneutical partners. An idea of *second nature* must come into play.⁷³ Similarly, the way Habermas talks the legitimation of normativity reads like a tautology: normative is what is recognized by members of a social community as normative. There must be something else, a normativity implied in the description of ‘interpersonal relations.’ Finally, how are utterances about the internal world criticizable? Let’s say this holds for utterances about the external world, by what external standard would we question or criticize the utterance “I feel X” (scared, angry, happy, etc.) In principle, this utterance is a statement about the internal world and does not require the external world for verification. For subjective worlds to “be drawn into public communication,” intersubjectivity must be able to offer a standard from where to judge, and negate solipsism and radical skepticism.

Finally, we conceive of the dogmatic nature of mythical, bare beliefs because via this criticizable space, we create a distance between our beliefs and the world, a distance that is not a Cartesian, solipsistic, unbridgeable abyss,⁷⁴ but rather, a distance of suspicions that can be shared. We understand our world-views as “attached to cultural traditions, constituted by internal interrelations

⁷¹ *Ibid.*, 52

⁷² *Ibid.*

⁷³ Markus Gabriel, *Transcendental Ontology Essays in German Idealism*, (New York: Continuum International Publishing Group, 2011) 3-6, and in particular the chapter “The Ontology of Knowledge.” This insight will be developed at the end of chapter V. 6

⁷⁴ Until God bridges it out.

of meaning, symbolically related to reality, and connected with validity claims —and thus exposed to criticism and open to revision.”⁷⁵ Our claims about the world might be wrong in different ways, and it becomes a matter of metaphysical belief to assert that something is just simply true.

Time and again, the only quality that is not reducible to a higher, more abstract, or relativist principle is precisely the porous border of the subject that is spread over the “transsubjective” space. This is the core of Habermasian anthropology of communication. It becomes uncriticizable because criticizing it in terms of ideological presuppositions à la Bernstein, or in the terms of the presuppositionless rebels described by Rorty above⁷⁶, would be self-defeating. In the act of criticizing, from a contingent perspective we would be practically proving the point itself. It would mean delegitimizing the very act of critique. Modern social reason replaces logical contradiction with a social act of intersubjective critique. However, if the “mythical world-view” gives rise to “an illusion about itself because the mind endows idealities with an existence outside of man and independent of him,”⁷⁷ it is legitimate to ask whether the transsubjective space is one of these idealities. This is a version of our question.

This question connects the apparent philosophy of human nature expressed in Habermas’s theory of communicative action with our project to unveil the metaphysical commitments of neutralist authors. Anthropology is, in this sense, the place-holder for metaphysics. The goal of this critique is to show which concepts underlie this anthropology, and break them down into their tacit metaphysical commitments. Intersubjectivity is definitely one of these concepts that may

⁷⁵ Jürgen Habermas, *Op. Cit.*, 52

⁷⁶ In epigraph 1.c) “Metaphysical commitments.”

⁷⁷ Jürgen Habermas, *Op. Cit.*, 47

carry metaphysical weight, that is why I call it a metaphysical carrier. It is central to all neutralist authors and it works as the cornerstone of their claim to contingent rationality, i.e., non-metaphysical. In the same way that we have obtained the concept of intersubjectivity from Habermas, in the following section, the second concept (i.e., reasonableness) will be obtained from Rawls's attempt at showing that his theory is not metaphysical. The last section will deal in the same manner with the notion of autonomy in diverse neutralist authors. In search for the metaphysical weight they carry, each of these carriers will be respectively analyzed in the last three chapters of the dissertation.

6. Rawls's method of avoidance

6.a. Universal Truth

Arguably, when John Rawls resurrected the liberal conception of justice with hopes of finding a substitute for the utilitarian model, part of his success and influence was due precisely to his method, which he outlines in his 1985 paper "Justice as Fairness: Political not Metaphysical"⁷⁸ in terms that are directly related to my argument:

...It may seem that this conception depends on philosophical claims I should like to avoid, for example, claims to universal truth, or claims about the essential nature and identity of persons. My aim is to explain why it does not. (...) The idea is that in a constitutional democracy the public conception of justice should be, so far as possible, independent of controversial philosophical and religious doctrines. Thus, to formulate such a conception, we apply the principle of toleration to philosophy itself: the public conception of justice is to be political, not metaphysical.⁷⁹

In these four sentences, we in fact see Rawls adopting certain philosophical stances. Rawls makes it explicit that his conception of 'justice as fairness' must remain independent of claims to universal

⁷⁸ John Rawls, "Justice as Fairness: Political not Metaphysical"; *Philosophy and Public Affairs*, Vol. 14, No. 3. (Summer, 1985), pp. 223-251.

⁷⁹ *Ibid.*, 223

truth and any specific philosophical anthropology. What does it mean for one's idea of justice not to be universally true? It must mean either that it is the product of historically contingent situations or that it is the product of a given consensus that does not obtain its validity from the content of what's agreed upon but from the procedure that has led to it. If it is the former, then it seems to be a sort of banal *Zeitgeist*, the content of which is irrelevant and requires no argument. If it is the latter, then the question is whether there exists a procedure that could lend validity to a given or potential consensus, not despite, but precisely because such procedure is simultaneously not dependent on a certain idea of universal rationality and a way to persuade or decide based on a better, or sounder, or truer argument. In this context, what would it mean to say that such a procedure, or its rules, is not universally true? It could mean that they are merely contingently valid, in an historical sense. More deeply, Rawls's position might mean that there is no such thing as universals, or any truth, at least in the political/moral realm, that can be properly so called and recognized as such. In this case, an argument for the very act of communicating ideas of justice—or any idea, for that matter—or even for the positing of the fact of social cooperation would be at the same time desperately needed and yet useless. It would be needed because if we cannot agree on the truth of the existence of the agreement itself at all, we would not know in what sense we could then meet and agree that certain claims are to be regarded as valid principles to correct situations of injustice. A validity claim, as Habermas suggests, is indeed necessary for any coherent understanding of social reason. But an explanation of how communication of ideas of justice works would also be hopelessly useless: if there is no such thing as truth in the socio-political realm, what good does any argument make? Persuasion is what Rorty would answer. But this is not what Rawls means. What Rawls is after is a conception of justice that can articulate the basic structure of a democratic society not because it claims to be the universal truth about justice, but because

people implicitly agree it's the best option, regardless of their beliefs on whether it is actually universally true or not, but still defensible from their own comprehensive views.

This conception of justice cannot simply be '*justice as consensus*' because people might agree on genocidal measures targeting a minority population. Once one begins to ask what must be excluded from potential agreements, something substantial or something procedural must be available to deal with the limits on what is acceptable. In Rawls's case, this would be an idea of fairness that is not metaphysical. Avoiding metaphysical claims means affirming contingency, yet contingent consensus can be horrific even if it isn't metaphysical, as in certain cases of genocidal violence. The question is how to steer the course between evil and irrationality, on the one hand, and dogmatic metaphysics, however rational they are, on the other. If a claim to universal truth is one of the things that can make a conception of justice metaphysical, what procedure for finding consensus on a certain conception of justice is contingent enough to avoid making a claim to universal forms of validity or rationality?

In addition to "universal truth," Rawls "should like to avoid" "[any] claims about the essential nature and identity of persons."⁸⁰ This is no coincidence because to scape the problem of universalism, so that the validity of a given conception of justice does not come from its fundamental truth, liberal neutralists look for it in people who agree upon it. And if the rationality of the conception cannot be claimed of the conception itself, it may be claimed of the people who agree upon it. Faithful to modern vocabulary, let us say that if the grounding is not in the absolute, it must be in the subject. The consequent danger is to turn the subject into an absolute, as some German idealist may think so. At least, if the reason people agree upon a certain

⁸⁰ Rawls, *Op. Cit.*, 222

basic structure of society is not the supposed universal validity of its underlying conception of justice, the reason for agreement must have something do with their natures being expressed in the underlying conception of justice.

What Rawls is acknowledging when he says that to make claims about the essential nature and identity of person would make his proposal metaphysical is that there is a connection between the possibility of accessing to a universal truth and an anthropology that would give an ontological account of what the person is. To propose either a conception of justice that would be fair to the dignity of that person, or an idea of rationality that would justify whatever is agreed upon, demands an equally metaphysical set of concepts. Thus any endorsement of liberal neutrality must avoid any recourse to an anthropology of that sort.

In the second part of the passage, however, there is a change in tone: “The idea is that in a constitutional democracy the public conception of justice should be, so far as possible, independent of controversial philosophical and religious doctrines.” Let us set aside, for the sake of the argument, the qualifying note ‘so far as possible.’⁸¹ This formulation quite obviously is a paraphrasing of neutrality. A question arises: what if the very idea of independence is a controversial philosophical doctrine? As we have seen in the previous chapter,⁸² what *consensus* might mean is directly related to what *competing* conceptions of the good might actually mean.

⁸¹ Admittedly, it is easy to construe this expression as ‘the least controversial public conception of justice’ or ‘the most independent.’ But if Rawls is implying that dependence on controversial philosophical and religious doctrines would make the theory, somehow, metaphysical, then coming up with *the least* metaphysical conception of justice still defeats the purpose of the entire proposal. After all, the point is to make the conception of justice *political, not metaphysical*. The problem with metaphysical commitments, according to liberals, is not so much that one conception of justice is very metaphysical and another one is less, even minimally, metaphysical; the problem is that a certain conception of justice, if it is metaphysically loaded, however slightly, is incommensurable with other metaphysical beliefs on justice. So, let us ignore the qualifying remark and assume that the goal is to build a public conception of justice that is actually “independent from controversial philosophical and religious doctrines.”

⁸² In epigraph 4.a “Relative neutrality”

Rawls was, on our account, an absolute, rather than a relative, neutralist, (see Chapter I, epigraph 4.a) which means that he was in favor not of neutrality regarding *actually* competing conceptions of the good, but rather of neutrality regarding all actual *and* potential conceptions of the good.

6.b Moral Conceptions

Rawls claims that “no general moral conception can provide a publicly recognized basis for a conception of justice in a modern democratic state.”⁸³ Simultaneously, he admits that his “justice as fairness” is a moral conception. He explains away this apparent contradiction by drawing on the separateness and narrowness of the political realm:

While a political conception of justice is, of course, a moral conception, it is a moral conception worked out for a specific kind of subject, namely, for political, social, and economic institutions. (...) It should also be stressed that justice as fairness is not intended as the application of a general moral conception to the basic structure of society, as if this structure were simply another case to which that general moral conception is applied. In this respect justice as fairness differs from traditional moral doctrines, for these are widely regarded as such general conceptions.⁸⁴

Three ideas are presupposed in this passage. 1) there exist different kinds of moral conceptions that are specific to certain institutions, namely, social, economic and political ones; 2) such conceptions are usually grounded in general principles of morality of which this political, social and economical morality would just be a specific instance; 3) ‘justice as fairness’ is a kind of moral doctrine that somehow significantly breaks with traditional morality in this respect. It manages to be a moral conception specific for social, economic, and political institutions but it’s not grounded in general principles of morality. Such moral doctrine is “freestanding.”

⁸³ Rawls, *Op. Cit.*, 225

⁸⁴ *Ibid.*, 224-225

The distinction between a general and non-general moral conception brings the argument back to the question of comprehensiveness, which we discussed in the previous epigraph. The fact of being comprehensive makes some view unwarranted. Examples of “comprehensive ideals” unsuited for a “political conception of justice” are the modern notions of “autonomy” and “individuality.”

As found in Kant and J. S. Mill, these comprehensive ideals, despite their very great importance in liberal thought, are extended too far when presented as the only appropriate foundation for a constitutional regime. So understood, liberalism becomes but another sectarian doctrine.⁸⁵

Combining these three passages, we might conclude that Rawls needs to show that he can ‘work out’ a moral conception for all these institutions without falling into three traps: (a) being comprehensive or appealing to principles of a general moral conception, (b) appealing to a universal truth or universal structure of rationality that would render one truth superior to others universally, and (c) presupposing an essential nature or identity of the person.

The heart of the matter: is it enough to merely apply one’s moral conception to political, economical and social institutions—or to say that one does so—, for it to not be metaphysically and/or comprehensively committed? Rawls admits as much when he says we may be “tempted to think that a metaphysical doctrine of the person is presupposed” in his theory: “While I said that this interpretation is mistaken, it is not enough simply to disavow reliance on metaphysical doctrines, for despite one’s intent they may still be involved.”⁸⁶ But discussing it, could force Rawls to discuss distinctions that belong to a metaphysical discourse, and position his theory accordingly.

⁸⁵ Rawls, *Op. Cit.*, 245-246.

⁸⁶ *Ibid.*, 239

Even the question, “does this set of distinctions—moral, political, metaphysical—hold in the context of discussing justice?”, cannot be answered. Rawls rightly says, “to rebut claims of this nature requires discussing them in detail and showing that they have no foothold.”⁸⁷ But he declines to do so: “I cannot do that here. I can, however, sketch a positive account of the political conception of the person.”⁸⁸ His strategy is to propose a “political” conception of justice without appealing explicitly to metaphysical positions, with the hopes of gathering consensus from a plurality of comprehensive views. But that turns the very distinctions that justify the position as non-metaphysical into uncriticizable preconditions. At the very least, the fact that it cannot be criticized at that fundamental level without forcing a response with metaphysical undertones suggests that the strategy to make it ‘freestanding’ is not successful.

At first glance, it would seem utterly pointless to publish a paper trying to show that one’s conception of justice is not metaphysical, but political and, at the crucial moment, when that very question arises, to shirk from it and yet admit that certain presuppositions might be involuntarily entailed. Instead of answering this question, Rawls insists on the very political/metaphysical distinction that is under examination, thereby using a maneuver in his argument that simply begs the question. But there must be an explanation. This last passage contains a long footnote.⁸⁹ In order to examine it in detail, I have divided it into three parts

Part of the difficulty is that there is no accepted understanding of what a metaphysical doctrine is. One might say, as Paul Hoffman has suggested to me, that to develop a political conception of justice without presupposing, or explicitly using, a metaphysical doctrine, for example, some particular metaphysical conception of the person, is already to presuppose a

⁸⁷ *Ibid.*, 240

⁸⁸ *Ibid.*

⁸⁹ Rawls, *Op Cit.*, 240, footnote 22.

metaphysical thesis: namely, that no particular metaphysical doctrine is required for this purpose.

Certainly, a case can be built to show that a hypothetical conception of justice, thought and articulated without recourse to metaphysical presuppositions and without a comprehensive conception of the person on the horizon, is in and of itself a metaphysically loaded position. The point is crucial: with Wittgenstein, the majority of “linguistic” philosophers and all relativists, one can argue that metaphysics is a form of non-sense that we, speaking primates, like to do (for reasons unknown), specially if one knows that for that reason one cannot really understand the critique, since it is metaphysical too. To hold this position may lead one to believe, with Nietzsche, that Truth-with-a-capital-T is “a mobile army of metaphors,” or with Rorty, that “Galileo, Hegel and Yeats” had the common achievement of developing “new vocabularies,” that equipped them “with tools for doing things which could not even have been envisaged before these tools were available.”⁹⁰ Metaphysics, in this view, is just but another vocabulary that has become obsolete in our time, and we must be content with a historically grounded set of metaphors, a vocabulary, that explains our merely contingent conceptions of justice.

The problem with this position is best exemplified by the first and the last sentence of Wittgenstein’s *Tractatus*.⁹¹ The first sentence, much less often quoted than the last, reads: “Perhaps this book will be understood only by someone who has himself already had the thoughts that are expressed in it—or at least similar thoughts.” In other words, maybe, Rorty is right when he says that when philosophers try to spell out the unarticulated aspirations of their age while at the same time drawing on their own resources, “they willy-nilly create, co-create or perhaps articulate new

⁹⁰ Rorty, the contingency of language, p 17, *Contingency, Irony, Solidarity*

⁹¹ I am partially following Kolakowski in this analysis

languages.” Yet, as Kolakowski points out, “more often than not, understanding a new language of philosophy is a matter of spiritual conversion: understanding and believing are indistinguishable” and thus, the first sentence of the *Tractatus* “amounts to a repetition of *credo ut intelligam*.”⁹² If that first sentence is right, there’s no move to a social or political conception of reason-giving. The only person who can ‘understand’ Wittgenstein’s view is someone who thought these things himself – mutual understanding is not dialogical, discursive. So any such view becomes essentially private – nothing that can serve as the basis for what Rawls or Habermas need to do.

Thus, Hoffman’s critique of Rawls is relevant because even if Rawls were to buy into a pragmatic-relativist conception of language and history that would allow him to create a vocabulary for justice that can be independent of metaphysics, the crucial point would not be whether he has managed avoid recourse to the traditional toolbox of metaphors that metaphysics use. Rather, whether what he is trying to avoid, namely exclusive and sectarian belief, is already emerging in the form of a new *credo ut intelligam*. And whether, unlike Rorty, he is guilty of denying its presence or its historicity. And yet Rawls does not adopt this relativist-cum-pragmatist view that metaphysics is just a kind of philosophical “vocabulary”. To the extent that this position can be described as a pejorative view of metaphysics, Rawls cannot make it his without ceasing to be impartial. This is an example of how discussing the non-metaphysical character of his theory would lead Rawls to having to explain his distance or proximity with metaphysical and anti-metaphysical claims. This is why his understanding of ‘freestandingness’ runs parallel to avoiding such discussions. Rawls is trying to articulate and affirm a moral conception of justice that avoids metaphysics by concerning itself

⁹² Kolakowski, *Op. Cit.*, 108

only with political, economical and social institutions, and not by endorsing a relativist theory of language and truth.

Thus, the other problem is the last line of the *Tractatus*, once metaphysical nonsense has been revealed: “Whereof one cannot speak, thereof one must be silent.” In what sense can Rawls speak of moral conceptions, political conceptions of justice, or fairness, while simultaneously rejecting metaphysics in general? In what sense can Rawls separate the notion of the political from the notion of the comprehensive without appealing to any metaphysical thesis? He cannot speak about it, and he cannot pretend it’s a shared political conception. What Rawls must do is show that his method is able to both avoid metaphysics and defend a conception of justice. But to do so, it is not enough to claim that he understands the political realm and the metaphysical realm as separate kingdoms. This is why Hoffman’s critique hits a nerve.

The footnote continues with further potential criticism of Rawls’s method of avoidance:

One might also say that our everyday conception of persons as the basic units of deliberation and responsibility presupposes, or in some way involves, certain metaphysical theses about the nature of persons as moral or political agents.

This is a substantial critique of Rawls’ theory and, in particular, of his ‘original position’ (that he just discussed in the paper prior to the footnote). It targets the anthropology of the theory both regarding the moral powers (“responsibility”) and the account of rationality (“deliberation”). They are connected with the notions of individuality (“basic units”) and autonomy that Rawls himself highlighted as being too metaphysical in the passage quoted above. As we have seen, Rawls builds his ‘justice as fairness’ on an account of social reason. His notions of overlapping consensus, reflective equilibrium, public reason and reasonability, together with the “instrument” of the ‘original position’, depend upon an understanding of the person, of which it is assumed to have

certain moral capacities (as we will see below). As in the case of justice, all these notions make up a vocabulary with strong metaphysical resonances that can be overcome only if Rawls can find a way to explain their meaning without lapsing into any foundationalism, representationalism, transcendental subjectivism, moral realism, or any other –isms that require the toolbox of metaphysics. This is not a minor issue for a theory that claims legitimacy precisely because of its overcoming of these difficulties. However, Rawls’s final response in the footnote is: “Following the method of avoidance, I should not want to deny these claims.”

What is the method of avoidance? Rawls is trying to offer a conception of justice that can be endorsed from different metaphysical and non-metaphysical perspectives. This is what he calls “overlapping consensus,” and it is in this sense that his ‘justice as fairness’ and the state that is informed by it are neutral. The method of avoidance is a rejection of making explicit the metaphysical stance that Rawls, as the intellectual author of the conception, takes in order to build up the conception. Every metaphysical or non-metaphysical position can and will relate ‘justice as fairness’ to their own understanding of metaphysics, of rationality, and of the person. Rawls claims to be only offering a political conception that, if successful, everyone can work out as they wish or need. But what Rawls is also avoiding is the question implicit in both critiques: is this notion of consensus through avoidance presupposing a metaphysical thesis or an anthropology? As the scholastics knew well, an abstract problem is never solved by appealing to a more abstract solution because there another more abstract question always arises⁹³. Rawls is trying to claim that his separation of the political from the metaphysical is not an abstraction in the traditional sense, but a

⁹³ For example in the discussion about infinity in Thomas Aquinas, *Summa Theologiae*, 1, 46,-2, ad 7m. *Contra Gentiles*, I, 13 (ed. Leonine), v. XIII, p. 33bMN.

specific application that works out the horizontal space of shared reasons and not the vertical space of rationalist abstractions. This is why he ends this long footnote as follows:

What should be said is the following. If we look at the presentation of justice as fairness and note how it is set up, and note the ideas and conceptions it uses, no particular metaphysical doctrine about the nature of persons, distinctive and opposed to other metaphysical doctrines, appears among its premises, or seems required by its argument. If metaphysical presuppositions are involved, perhaps they are so general that they would not distinguish between the distinctive metaphysical views—Cartesian, Leibnizian, or Kantian; realist, idealist, or materialist—with which philosophy traditionally has been concerned. In this case, they would not appear to be relevant for the structure and content of a political conception of justice one way or the other.

Technically, we do not abandon the same problem as before, that is, the distance between claiming that no particular metaphysical doctrine appears “among its premises” and not having some implicit metaphysical commitments. But the argument is also weak because it first assumes that there is a level of generality that is so great that includes almost every view, which in itself is a view on how metaphysical theories are related to one another. Secondly, in truth it excludes many metaphysical positions, for example the metaphysical proposition that God is a person who commanded me to do something. Or a teleological metaphysics that claims that a people is chosen by history for a particular mission.

In spite of Rawls’s claim that his potential metaphysical presuppositions are too general to be ascribed to any specific tradition, there are some key ideas and conceptions used for the presentation of justice as fairness that Rawls borrows from what he labels ‘Kantian constructivism’: “...the adjective “Kantian” indicates analogy not identity, that is, resemblance in enough fundamental respects so that the adjective is appropriate.”⁹⁴

These fundamental respects are certain structural features of justice as fairness and elements of its content, such as the distinction between what may be called the Reasonable

⁹⁴ Rawls, *Op. Cit.*, 224 footnote 2

and the Rational, the priority of right, and the role of the conception of the persons as free and equal, and capable of autonomy, and so on.⁹⁵

At first glance, it is the last part of this quote that gives away Rawls's tacit anthropological assumptions: persons as free and equal, autonomous, and so forth. But Rawls is very insistent that these notions are not referred to the person fundamentally, not even with respect to their profound self-image, but only politically. One can believe or not believe in the person's fundamental freedom and equality, and yet nevertheless might accept it on political grounds for the sake of some universally acceptable system. For this method of avoidance to work, the key aspect lies in the distinctions predicated in the first part of the quote: reasonable versus rational and the priority of the right over the good. These distinctions are the irreducible aspects of his "political, not metaphysical" and it is where we will find his hidden metaphysical presuppositions. The former refers to what is allowed to enter the public realm, the second is what is allowed to be imposed from that sphere over the others.

Habermasian intersubjective space and Rawlsian reasonableness and the priority of the right over the good are the building blocks of neutrality that shall be analyzed in chapters III and IV respectively. What's lacking is an account of the moral capacities that are attributed to the individual who is going to reasonably participate in the trans-subjective space where questions of right prevail over questions of good.

7. Autonomy and Moral Powers

⁹⁵ *Ibid.*

The question of the account of the moral powers of the individual in the liberal state is controversial among neutralist authors. The controversy revolves around the ideal of autonomy. Is individual autonomy a precondition of neutrality? As we have seen above, Rawls himself believes that if his kind of neutral liberalism were based on the ideal of autonomy it would become “another sectarian doctrine.”⁹⁶ Brian Barry goes even further in noting the tension between individual autonomy and state neutrality: “It is not simply that neutrality is not entailed by a conception of the good as autonomy. The two are actually incompatible.”⁹⁷ Charles Larmore similarly contends that a common mistake made by liberals (“including, at least in part, John Rawls”) is to include a “commitment to autonomy,” which actually undermines the attempt at securing neutrality.⁹⁸ Perfectionists such as Raz implicitly agree with these statements when they endorse the public promotion of autonomy and reject, for this very reason, the principle of neutrality.

On the other hand, neutralists like Ackerman argue that the principle of neutrality is logically derived from the defense of autonomy, defined as “the capacity to form a rational plan of life.”⁹⁹ neutrality and autonomy are coextensive because “whenever I offend neutrality, I offend this good.” This formulation conflates the idea of autonomy as a capacity and the idea of autonomy as a political good, or as a conception of the good, that must be protected or promoted. This position is explicitly rejected by both Barry and Larmore, but their disagreement stems in part from a different understanding of what is entailed when we the claim that autonomy is the basis for neutrality.

⁹⁶ *Ibid.*, 245-246

⁹⁷ (Barry, *Op Cit.*, 130

⁹⁸ Charles Larmore, preface to *Patterns of Moral Complexity* (Cambridge: Cambridge University Press, 1987), x “..liberals themselves, from Kant and Mill until today (including, at least in part, John Rawls),...have supposed that the ideal of political neutrality ought to include a general commitment to “autonomy””.

⁹⁹Bruce Ackerman, *Social Justice in the Liberal state* (New Haven, Conn: Yale University Press, 1980), 367

Ackerman claims that in order to violate (or “offend”) neutrality, “by definition, I must declare that another citizen is to be subordinated to my purposes in the power struggle rather than recognized as a person whose capacity to form a life plan is no less valuable than my own.”¹⁰⁰ He goes on draw the following conclusion: “So long as it is morally better to respect my moral autonomy than to force me to achieve one of these *other* goods, the path has been laid for a rational commitment to neutrality.”¹⁰¹ Part of the challenge with such arguments is that it would seem that you need to argue for subordination of other people to question it at all. But the question here is whether one can argue against it without a metaphysics, or rather one needs a metaphysics to argue against it.

For Larmore and Barry (and Rawls), the defense of autonomy is the defense of a particular conception of the good that violates neutrality. Barry calls it “a second-order conception of the good” because “it does not specify what the good actually consists in.”¹⁰² Autonomy refers to a twofold quality: a quality of the person who specifies what the good actually is, and a quality of the good itself. Autonomy is the conception according to which “what is of central importance in human life is that people should make up their own minds about how to live and what to think, and that they should be able to express their beliefs freely and act on their conclusions about the best way to live.”¹⁰³ But Autonomy, beyond being a certain conception of the person, also presupposes a broader metaphysics. For example, a certain conception of nature, such as liberal naturalism, according to which we are a *second nature* that allows for freedom to operate.¹⁰⁴ At least, it

¹⁰⁰ *Idem*.

¹⁰¹ *Idem*, 368

¹⁰² Barry, *Op. Cit.*, 132.

¹⁰³ *Ibid*.

¹⁰⁴ Markus Gabriel, *Op. Cit.*, 3-6.

presupposes that some metaphysical conceptions of nature are false. Possibly, also of freedom; of ends, etc... This is where strong metaphysics is coming in the back door.

At first glance, freedom to pursue one's purpose in life would seem to be exactly what the principle of state neutrality is meant to protect. But, what Barry—not to mention Larmore and Rawls—see as problematic is that there might be some goods that people live by that are not derived from their autonomy, but rather from obedience or indoctrination, for example. The idea that all conceptions of the good are derived from a personal choice, and that thus, neutrality actually protects autonomy, is in itself a controversial (and perhaps metaphysical) conception of the person. According to Ackerman, the road from autonomy to neutrality is paved with a “conception of the good that gives central place to autonomous deliberation and denies that it is possible to *force* a person to be good.”¹⁰⁵ Barry rejects this idea, saying that it implicitly appeals “to a Protestant or Kantian account of virtue, which makes it a matter of the will.”¹⁰⁶ This would be just one of many possible anthropological positions. From an Aristotelian alternative view, we can regard moral virtue as a kind of habit, and as Barry points out, “habits can, we know, be acquired in the first place under coercion.”¹⁰⁷ However, the underlying question is whether there is a necessary connection between autonomy and the good, or whether their connection is accidental, contingent upon other conditions. If what neutrality is meant to protect is freedom of choice, or at least, freedom from state coercion on our comprehensive views, the question becomes whether that weak understanding of freedom presupposes the idea of a rational freedom from any idiosyncratic factors (e.g., our biological, psychological, and other circumstances), and that capacity is the

¹⁰⁵ Ackerman, *Op. Cit.*, 11

¹⁰⁶ Barry, *Op. Cit.*, 131

¹⁰⁷ *Idem.*

reason for the protection that neutrality is meant to offer, then autonomy would be both an anthropological presumption and a good to protect with clear metaphysical commitments.

If that were the case, by taking autonomy and neutrality to be coextensive, the demarcation of what's acceptable is reduced because even though it seems that anything could be regarded in principle as good in a second-order way, in truth, a particular first-order good would only be valid “so long as the person who conceived it as good (in a first-order way) had arrived at this conception in a way that satisfied the requirements of autonomy.”¹⁰⁸

However, neutrality is meant to signify that the state does not intervene in the commitments to comprehensive doctrines or conceptions of the good that citizens might have. It is meant to address the fact of plurality. As Larmore points out, “one advantage of making neutrality the primary ideal of liberalism is that it explains what freedom has generally meant for the political liberal.”¹⁰⁹ The controversy lies precisely in the idea of freedom that stems from neutrality¹¹⁰: namely whether we are to assume that this non-intervention entails that people are or ought to be free to choose their one conception of the good. Admitting that it does, implies accepting some sort of autonomy. Denying it amounts to saying that people are either determined to believe what they say they believe, or that they are forced, indoctrinated, or oppressed into believing what they say they believe. If we are to accept that people are victims of spiritual determinism, —that their spiritual beliefs are determined by their circumstances or their destiny—, it is not clear what is

¹⁰⁸ *Idem.*, 129. Also: “A conception of the good as autonomy does not imply that the pursuit of all substantive conceptions of the good is equally valuable. Only those conceptions that have the right origins --those that have come about in ways that meet the criteria for self-determined belief- can form a basis for activity that has value.” *Idem.*, 132

¹⁰⁹ Larmore, *Op. Cit.*, 46.

¹¹⁰ And if that freedom is in truth presupposed, thus neutrality stems for it.

gained in terms of a path to avoid metaphysically loaded anthropological descriptions such as autonomy, nor is it clear from which perspective neutrality is affirmed: from an acceptance of that spiritual fate? Would the theory of neutral liberalism be also the expression of some determined nature or some determined set of circumstances, however contingent? Say: neutral liberalism would be the expression of the social nature of wealthy, white males that have everything to gain if the world is ruled following the rules of their trade (i.e., political philosophy). They would be unfree, in their sense. Thus, if neutrality were only to arbitrate among oppressed groups, and secured some sort of peaceful co-existence, it is not clear in what sense this is meant to be a theory of justice, nor what difference would it be or what is gained as an alternative to a effective tyranny that imposed a set of beliefs for the sake of peace.

Autonomy combines questions of free will with notions of the subject's normative capacities. This is a problem for liberal neutralists because the questions that the idea of freedom raises because they are asked and answered from metaphysical positions. To that extent, they are controversial. And neutrality is supposed to, well, be neutral among them. Nagel says that "this is not a case where there are several possible candidate solutions and we don't know which is correct. It is a case where nothing believable has (to my knowledge) been proposed by anyone in the extensive public discussion of the subject."¹¹¹ The belief in autonomy

presents itself initially as the belief that antecedent circumstances, including the condition of the agent, leave some of the things we will do undetermined: they are determined only by our choices, which are motivationally explicable but not themselves causally determined.¹¹²

¹¹¹ Thomas Nagel, *The View From Nowhere*. (Oxford: Oxford University Press, 1986), 112

¹¹² *Ibid.*, 14

This seems to be partially at odds with the project of post-metaphysical neutrality to the extent that these liberal authors try to pose the question of how to construct the best polity for individuals who do not exist *in abstracto*, but are always physically, linguistically and socially situated.¹¹³ And autonomy connects that project to Kant's transcendental ethics and his metaphysics of morals. As Williams notes, Kant "makes no bones about focusing on the subject separately from the physical, social, and historical context."¹¹⁴ Most likely, any theoretical justification of our feeling free would rely on some intuition of certain moral principles. If we were to find something to support our feeling that our free actions originate with us, this would probably depend on identifying some fundamental features of a moral awareness that transcends all particular cases.¹¹⁵ This would clearly be part of a metaphysics conceived "in a heuristic sense as accounting for the proper regulative role of reason in terms of our knowledge and its genuinely constitutive role in terms of our practice (or moral life)" which is how Williams characterizes Kant's third understanding of metaphysics, quoted above in this chapter.¹¹⁶

While rejecting autonomy and (absolute) free will to justify a liberal neutralist position is understandable for strategic and philosophical reasons, some sort of freedom needs to be asserted for neutrality to make sense. Larmore, who rejects autonomy as the basis for neutrality, justifies the affirmation of the ideal of equal respect as the minimal content that neutrality requires. neutrality

¹¹³ paraphrases of Williams, *Op. Cit.*, 1

¹¹⁴ *Ibid.*, 3

¹¹⁵ Nagel, *Op. Cit.*, 120, and Williams, *Op. Cit.*, 15

¹¹⁶ Epigraph 1.b

explains “what freedom has meant for the political liberal”, and that it “emphasizes the equal freedom that all persons should have to pursue their conception of the good life.”¹¹⁷

Then, what kind of personal freedom is needed for these non-metaphysical conceptions of justice to make sense? In order to see whether neutrality’s protection of free choice is dependent on some metaphysical view, and is no different from the strong understanding of autonomy, both as a capacity and as a good to protect, we must know how it is connected to the notions of intersubjectivity and reasonableness. As usual, the most complete account is Rawls’s (my italics):

Since persons can be full participants in a fair system of social cooperation, *we ascribe* to them the two *moral powers*...: namely, a capacity for *a sense of justice* and a capacity for *a conception of the good*. A sense of justice is the capacity *to understand, to apply, and to act* from the *public* conception of justice which characterizes the fair terms of social cooperation. The capacity for a conception of the good is the capacity *to form, to revise, and rationally to pursue* a conception of one's *rational* advantage, or *good*.¹¹⁸

It is unclear to what extent this description is different from old-school Kantian autonomy, from “our sense that our free actions originate with us,” as Nagel puts it (and rejects it)¹¹⁹, or “the capacity to form a rational plan of life,” as Ackerman describes it (and endorses it).¹²⁰ Kołakowski also comments on this: “if each of us is a source of independent initiative which sets itself in motion and has no sufficient cause beyond itself, then it seems that we, too, perform creative acts *ex nihilo*.”¹²¹ This suggests, in a Kantian fashion, that the mere possibility of autonomy requires metaphysics. Be that as it may (for now), what matters for the purpose of showing neutrality’s

¹¹⁷ Larmore, *Op. Cit.*, 46

¹¹⁸ John Rawls, “Justice as Fairness: Political not Metaphysical”, *Philosophy and Public Affairs*, Vol. 14, No. 3. pp. 223-251 (Summer, 1985), 234, (my italics throughout)

¹¹⁹ Thomas Nagel, *The View From Nowhere*. (Oxford: Oxford University Press, 1986), 117

¹²⁰ Ackerman, *Op. Cit.*, 367

¹²¹ Leszek Kołakowski, *Metaphysical Horror*, revised edition (Chicago and London: The University of Chicago Press and Penguin, 2001), 94-95.

metaphysical commitments is to unveil how this idea of “moral powers” and this idea of “rationally” forming, revising and pursuing one’s rational advantage is part of a common anthropological view assumed in the principle of neutrality. And if so, whether that means that it fosters a determinate kind of human being by means of public power and coercion, based on a comprehensive understanding of what human beings are and ought to be.

Rawls’s quote shows to what extent the entire system is dependent on this idea of moral power and rational revision. Rawls says that the assumption is that persons can participate in a fair system of cooperation. That assumption leads him to “ascribe” to them these moral powers. Cooperation and fairness require moral subjects. Now, of these moral powers, the first, “a capacity for *a sense of justice*” means to understand and act from the kind of conception of justice that Rawls labels ‘public’ (in part, to make it political, not metaphysical), and which “characterizes the fair terms of social cooperation.” We should pay attention to the circularity here: since they can participate in a (non-metaphysical) fair system of cooperation we ascribe to them the capacity to understand and act from what characterizes (as non-metaphysical) fair systems of cooperation.

The second moral power, the capacity to form and revise one’s comprehensive view of the world is what ultimately leads one to actualize the first moral power, the understanding and acting from a fair system of cooperation. This is a delicate argument and it all hangs on the assumption that by ascribing these two moral powers, the powers to endorse simultaneously metaphysical and non-metaphysical goods, in the private and public domains respectively, we are not endorsing any essentialist anthropology. It is a very fine line that neutralists walk, especially those who say to reject autonomy. This is the irreducible core of the overarching argument in favor of neutrality, and thus

it must be examined. And it is where their metaphysical commitments will outcrop, when we reconstruct the entailed anthropology in the following chapters.

In this chapter, we have selected three notions that form the backbone of neutrality because all three are common neutralists liberals and they occupy the border between the metaphysical and the non-metaphysical. There are many more notions that could explain the relationship between neutralist systems and metaphysics, but these three cover a wide range of political and anthropological realms, and they are sufficient for our purposes. Intersubjective consensus explains the space where decisions are said to be made independently of any ideas of universal truth. Reasonability explains the epistemological status of political ideas that are said to avoid any claim to superior rationality as a reason for their political superiority. Autonomy explains the ambivalence between the need for moral powers in their systems and the rejection of essential anthropologies. The following chapters are devoted to them individually. If this chapter has shown the role these notions play in the justification of neutralist systems, in the next chapter we will examine the arguments that these authors use to justify each of these notions in particular, and I will try and show how and why these arguments harbor metaphysical commitments. This is why I have called these three notions metaphysical carriers: like moonshiners, they carry the weight the system hides.

Chapter III

Intersubjectivity's Metaphysical Content in Habermas

“Julie smiles to herself, under her hat. The smile’s is cold. “You know what’s fun, if you want to have fun,” she says, “is to make up explanations. Give people reasons, if they want reasons. Anything you want. Make reasons up. It’ll surprise you—the more improbable the reason, the more satisfied people will be.”
“That’s fun?”
“I guarantee you it’s more fun than twirling with worry over the whole thing.”
David Foster Wallace
“Little Expressionless Animals”¹

The previous chapter showed the central role that three concepts —intersubjectivity, reasonability, and autonomy— play in justifying state neutrality, and how they can be seen as carriers of metaphysical commitments. By carrier I mean something that harbors yet does not explicitly express a metaphysical commitment, and thus it carries that content into the system. This chapter will analyze the inner structure of the first these three concepts —intersubjectivity — as Habermas articulates it, in order to unveil those commitments.

1. Habermas's Intersubjectivity as Detranscendentalization of Reason

Intersubjectivity is the key to discursive democracy; it underwrites the dialogical space that neutralist institutions are meant to arbitrate.² It is because the possibility of intersubjective consensus exists that peaceful pluralist democracies, in the liberal sense, are said to be possible,

¹ David Foster Wallace, “Little Expressionless Animals” in *Girl with Curious Hair* (New York: W. W. Norton, 1989), 33

²I call them “neutralist” and not “neutral” because whether they are neutral or not is what it’s at stake in the argument—I believe they are not—. On the other hand, they are ‘neutralist’ because they embody the ideology that advocates for the possibility and convenience of neutrality.

coherent, and stable. Different views are brought together in agreement because reasons can be exchanged, meanings can be transmitted, common grounds are discovered and recognized by subjects living intersubjectively. Intersubjectivity is also at the center of Habermas's philosophical project, which, in broad terms, can be summarized, in McCarthy's words, as "the shift of the focus of the critique of reason from forms of transcendental subjectivity to forms of communication."³

Because it lies at the core of his life's work, Habermas's understanding of intersubjectivity is rich and impossible to unfold comprehensively in just a few pages, but for the purposes of our argument, several aspects of its meaning must be highlighted in detail. To begin with, the transition from (Kantian) transcendental subjectivity to (post-Hegelian) intersubjective communication has to do with the failures of early modern rationalism and metaphysics, as much as with Habermas' attempt at salvaging from the wreck of modern (post-Hegelian) idealism whatever universalism can be legitimately used for the purposes of morality and justice. So, on the one hand, Philosophy can "no longer refer to the whole of the world" in the sense of a "totalizing knowledge,"⁴ and all attempts at "discovering ultimate foundations, in which the intentions of First Philosophy live on, have broken down."⁵ On the other hand, "we have to face the question whether there is not a formal stock of universal structures that develop, according to their own logics, under the abstract standard of truth, normative rightness, and authenticity."⁶

³Thomas McCarthy, "Kantian Constructivism and Reconstructivism: Rawls and Habermas in Dialogue" *Ethics*, Vol. 105, No. 1, pp. 44-63 (University of Chicago Press: Oct., 1994), 45

⁴ Jürgen Habermas, *The Theory of Communicative Action*, Vol.1: *Reason and The Rationalization of Society*, trans. Thomas McCarty, (Boston: Beacon Press 1984), 1

⁵ *Ibid.*, 2

⁶ *Ibid.*, 180

While we look for these universal structures, he thinks, we should not deny “the pluralism and the incompatibility of historical versions of civilized humanity,” but we should regard this “multiplicity of forms of life” as “limited to *cultural contents*.” This “universalist position” asserts that “every culture must share certain *formal properties* of the modern understanding of the world.”⁷

Such formal properties are embedded in the very act of communicating (speech acts), in the practice of argumentation (discourse), and in the presuppositions that sincere attempts at reaching understanding imply for every conversation (consensus). This philosophical gesture belongs to the tradition of the linguistic turn, however its pragmatist inflection is opposed to the tradition that ended up, as Rorty puts it, in the “reification of language.”⁸ The latter is proper to the early Wittgenstein, Russell, or Frege, who substituted the notions of “language” and “meaning” for the Kantian notion of “experience,” such that the conditions of possibility of experience became the conditions of possibility of meaning. Rorty believes that this was an attempt at finding an *a priori* standpoint from which to do philosophy,⁹ secure from the naturalization of knowledge. But he also thinks that this linguistic turn was “too honest to survive” on these metaphysical terms.¹⁰

⁷ *Ibid.*

⁸ Rorty, Richard, “Wittgenstein, Heidegger, and the reification of language” in *Essays on Heidegger and others: philosophical papers*. (Cambridge, Eng.: Cambridge University Press, 1999), 57

⁹ *Ibid.*, *passim*

¹⁰ Richard Rorty, “Metaphilosophical Difficulties of Linguistic Philosophy” Introduction to *The Linguistic Turn, Essays on Philosophical Method*, Ed. Richard Rorty, (Chicago: The University of Chicago Press, 1967, 1992)

Habermas rather belongs to a more pragmatist version of the linguistic turn, one that accepts, as Davidson puts it, that “we must give up the idea of a clearly defined shared structure which language users master and then apply to cases,”¹¹ and which rejects the idea of turning Philosophy of Language into a new “first philosophy,” in Michael Dummett’s words.¹² Indeed, Habermas’s theory of meaning and communication wants to turn away from the philosophy of consciousness into the realm of “detranscendentalized” reason. This task also belongs, he thinks, to “both the historicist tradition from Dilthey to Heidegger and the pragmatist tradition from Peirce to Dewey (and, in a sense, Wittgenstein).” This is because they understand “the task of ‘situating reason’ as one of detranscendentalizing the subject of knowledge.”¹³ As part of his turn away from the philosophy of consciousness, “Habermas does not picture the social world as an object (or collection of objects) standing over against a plurality of subjects with which it causally interacts. (...) Rather, it is a medium that we inhabit. It is ‘in’ us, in the way we think and feel and act, as much as we are ‘in’ it.”¹⁴ The break with the philosophy of consciousness thus rejects the tension in the subject-object relationship and, with it, most of its philosophical questions, from solipsism to radical skepticism, and also most of its answers, from positivism to the transcendental a priori presuppositions of knowledge.

¹¹ Donald Davidson, “A Nice Derangement of Epitaphs” in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. Ernest LePore (Oxford: Blackwell, 1986), p 446, as quoted by Rorty, Richard, *Wittgenstein, Heidegger, and the reification of Language*, in *Essays on Heidegger and Others. Philosophical papers. p 50*

¹² Rorty, Richard, “Wittgenstein, Heidegger, and the reification of language” in *Essays on Heidegger and others: philosophical papers*. (Cambridge, Eng.: Cambridge University Press, 1999), 51

¹³ Jürgen Habermas, *Between Naturalism and Religion*, trans. Ciaran Cronin (Cambridge, UK: Polity Press, 2008) 25

¹⁴ James Gordon, *Habermas* (Oxford: Oxford University Press, 2005), 31. I call this, horizontal transcendence.

Thus, both natural law theories and Kantian formal morality have been foreclosed for theorists like Habermas. For the purpose of finding universal normative standards, or building a rational morality, or a theory of the social and political world, Habermas turns to this new “situated”, “detranscendentalized,” reason. He focuses on the social aspect of reason, on the relationships between subjects, and looks into the tacit presuppositions of communicative action. This is why the shift is predicated upon a pragmatist theory of meaning, in which the function of speech is “to bring interlocutors to a shared understanding and to establish intersubjective consensus” that “enjoys priority over its function of denoting the way the world is.”¹⁵

Hoy and McCarthy offer a synthesis of the implications of this balance between detranscendentalization and the search for normativity in communication which already shows that Habermas’s position knowingly hangs over the edge of transcendentalism:

This [move] has the effect of relocating the Kantian opposition between the real and the ideal within the domain of social practice. Cooperative interaction is seen to be structured around ideas of reason which are neither fully constitutive in the Platonic sense nor merely regulative in the Kantian sense. As idealizing suppositions we cannot avoid making while engaged in processes of mutual understanding, they are actually effective in ways that point beyond the limits of actual situations. As a result, social-practical ideas of reason are both “immanent” and “transcendent” to practices constitutive of forms of life.¹⁶

For Habermas, intersubjectivity is an empirical fact of our lives, embedded, as it were, in the fabric of our species. Indeed, this is the most salient trait of his anthropology: from our distinctive rearing practices—“No other animal is dependent for so long a period of rearing on the protection of the family and a public culture intersubjectively shared with conspecifics”¹⁷—to the structure of our mind—“For the individual mind is imbued with structure and content by

¹⁵ *Ibid.*

¹⁶ David C. Hoy and Thomas McCarthy, *Critical Theory*, (Cambridge, MA: Blackwell, 1994), 38. cited in Habermas, *Op. Cit.*, 27

¹⁷ *Ibid.*, 13-14

locking into the ‘objective’ mind of the intersubjective interactions of *intrinsically* socialized subjects.”¹⁸ Intersubjectivity is what defines both our reason and our life.

Habermas develops the centrality of intersubjectivity in two spinal, load-bearing theories: the theory of pragmatic meaning and the theory of communicative rationality.¹⁹ These theories are contiguous and interdependent. The pragmatic aspect of meaning highlights the fact that “reaching understanding is the inherent telos of human speech. (...) The concepts of speech and understanding reciprocally interpret one another.”²⁰ The theory of communicative rationality deals with the ideal presuppositions that are present in the act of communicating, arguing, and reaching understanding —as opposed to instrumental rationality and action. These are presuppositions of a certain communicative rationality, an implicit and intuitive *know-how* that the theory wants to make explicit and normative, and from which Habermas (and others, such as Karl Otto Appel) hope to secure formal-pragmatic norms for practical reason.

These two theories are the basis for three research programs throughout Habermas’s work: the program of social theory, the program of discourse ethics, and the program of political theory. The normative universality that Habermas distills through his analysis of the pragmatic presuppositions of communication is the substance underlying these three programs: a sociology, a morality, a politology. They all ultimately rest on the analysis of what happens when a plurality of interlocutors try to reach an understanding. The very idea of democracy is built around this

¹⁸*Ibid.*, 14. The beginning of the quote: “I have in mind the image of a subjectivity that one must imagine as a glove turned inside out in order to reveal the structure of its fabric woven from the strands of intersubjectivity. Inside each individual person, we find a reflection of the external social world.”

¹⁹ Gordon., *Op. Cit. passim*

²⁰ Jürgen Habermas, *The Theory of Communicative Action*, Vol.1: *Reason and The Rationalization of Society*, trans. Thomas McCarty, (Boston: Beacon Press 1984), 287

event: willful agreement. And the norms that explain and must regulate the processes behind a democracy are ultimately nothing but an idealization of the presuppositions in that action. Democracy is, in Habermas' view, the political-legal procedure that makes that agreement, that acquiescence, fair and rational.

The overarching character of Habermas's theory has drawn many criticisms. In their theoretical debate of 1995, Habermas and Rawls accuse one another of being too substantial. Each claims to be more "modest" than the other. I will review what Habermas says of Rawls in the following chapter, but Rawls's criticism of Habermas's work is precisely that it implies a comprehensive view: "His logic is metaphysical in the following sense: it presents an account of what there is — human beings engaged in communicative action in their life-world."²¹ And later in the essay: "His doctrine is one of logic in the broad Hegelian sense: a philosophical analysis of the presuppositions of rational discourse which includes within itself all apparent substantial elements of religious and metaphysical doctrines."²² Rawls's own theory is based on a very sharp distinction between the political realm and the moral/religious/cosmological/comprehensive views of each person or group. His goal is to build a freestanding political conception that avoids such debates. Habermas's theory offers an account of "what there is" in the strong sense, and builds its political theory, the norms of its practical reason, from an anthropology that Rawls sees as comprehensive, even if sublimated by means of a theory of intersubjectivity.

²¹ Rawls, "Reply to Habermas", in *Political Liberalism* (New York: Columbia University Press, 1993), 379

²²*Ibid.*, 382

Rawls, from his liberal perspective, is not the only one to see substantial elements in Habermas' theories. Michael Walzer, Alasdair MacIntyre, Richard Rorty, Karl Otto Appel, Seyla Benhabib, Michael Sandel, and James Gordon —among many others—, all see it in different aspects of them, to the extent that this “substance” can be seen as metaphysically committed. From liberal, communitarian or pragmatist perspectives, from neo-Hegelian, neo-Aristotelian or from a neutralist standpoint from which to serve as arbiter over all other perspectives, they perceive the ever-present risk of smuggling in some non-neutral substantial content. Adding to this debate, I submit that at the heart of Habermas' theory there is a leap of faith in regards to the concept intersubjectivity, that is, at the heart of his anthropology; furthermore I submit that the criticisms of these authors show where such a leap of faith is manifested. Rawls sees it in the connection between consensus and validity, secured by the implicit presuppositions of communicative rationality. Walzer sees it in the “thickness” of the normativity of those same presuppositions, particularly in the way ‘designed’ philosophical conversations presuppose the agreements they claim to reach at the end; Rorty in the “unforced force of the better argument” that wins over in the context of rational discourse and its claim to universality; Gordon in the “*non-sequitur*” connection between understanding and agreement in Habermas theory of pragmatic meaning; Kelly and Benhabib in the principle of universalization that Habermas introduces in his “social practice.” I propose that these criticisms taken together show that at the core of intersubjectivity there is the belief, but not the evidence, that communication *happens*. This seems like a very uncontroversial fact, —otherwise, why would one ever speak or write —, but Habermas's life's work shows us that in order to assert that communication can be the basis for fair social practices with universal value and rational appeal, we must be able to describe our relationship with speech in terms of quasi-transcendental preconditions of rationality. These preconditions support the

idea that whenever we *make an utterance*, our meaning can be effectively understood and agreed upon because we can put ourselves in the perspective of the listener—and vice versa—and that we can do so because we share a certain rational/formal structure of reason-as-communication. The modern isolation of the self-as-consciousness is overcome because we actually inhabit a trans/intersubjective space where reasons mean the consensus they reach, and this trans/intersubjective space that we inhabit is actually the very structure of our rationality—that is, in an intersubjectively constructed space of reciprocal idealism. This may have been called situated reason, social reason, decentered reason, and pragmatic reason, but I submit that it is neither post-metaphysical nor detranscendentalized.

Let us go step by step.

2. Validity Claims and Rawls Critique of Habermas's Comprehensiveness

In this subchapter, I will address the criticism that Rawls addresses to Habermas's political theory. To do so, we begin by showing the role that validity claims play in Habermas's theory of communicative action, and how that influences his view of the political organization of society that Rawls sees as too substantial to be neutral.

In Habermas's detranscendentalized concept of reason, truth and rightness are substituted for "validity claims". In the same manner that the modern notions of truth and rightness refer at bottom to a certain correspondence between mind and world, or word and object—which is the concern of the philosophy of consciousness and what Habermas wants to eliminate—validity claims refer to a correspondence between meaning and understanding among different interlocutors. When Habermas says, "we understand the meaning of a speech act, when we

know what would make it acceptable,”²³ he is embracing a certain rationalist thesis to the extent that meaning depends on validity, and validity implies that the hearer has to bring to mind the reasons that amount to its justification. The essential point here is that, as Gordon phrases it, “reason and validity, not truth, are doing the work.”²⁴ Interlocutors are oriented toward “reaching understanding” (the “inherent telos” of speech, according to Habermas). The speech acts they perform are, implicitly or explicitly, acts of giving reasons. We know this because we understand these speech acts when “we know the kinds of reasons that a speaker could provide in order to convince a hearer that he is entitled in the given circumstances to claim validity for his utterance—in short, when we know *what makes it acceptable*.”²⁵

Amidst the social complexity of giving a receiving reasons, any speech act, (for example: a constative, fact-stating speech act),²⁶ if it is uttered for the purpose of cooperation or agreement, implies the following validity claims: first, it expresses what Habermas calls ‘the inner world’ (an intention to communicate a belief), second, it establishes a communicative relation with a hearer (a social world in which both speaker and hearer know they are exchanging information), and third, it attempts to represent the external world. This triadic structure entails a set of tacit claims: “the claim that the speech act is sincere (non-deceptive), is socially appropriate or right, and is factually true (or more broadly: representationally adequate).”²⁷ In other words, sincere

²³ Habermas, *Op. Cit.*, 297

²⁴ Gordon, *Op. Cit.*, 37

²⁵ Jürgen Habermas, *On the Pragmatics of Communication*, B. Fultner (trans.) (Cambridge, MA: MIT Press, 1998), 232

²⁶ For this summary, I am following Bohman, James and Rehg, William, "Jürgen Habermas", *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2014/entries/habermas/>>.

²⁷ *Ibid.*

speech-acts necessarily make three simultaneous claims: a validity claim to truth, to rightness, and to truthfulness. They are “necessary” because they are assumed to be inherent *ipso facto* in the act of speaking.²⁸

When the hearer does not accept these claims because she does not know what makes them acceptable, and thus no agreement is reached, speaker and hearer may shift to “reflexive communication,” from speech to discourse. Discourse makes explicit and tests the kind of rational justifiability implicit in the failed speech act, to see whether it is acceptably true, correct, and authentic. For this discursive register (which is a moment of argumentative dialogue) to be the basis for understanding after the previous misunderstanding, or agreement after disagreement, and ultimately for acquiring consensus and social order, several conditions —“ideal conditions”— must be present (more on these conditions below). When these conditions are met, then the “unforced force” of the better reasons carries the day, and consensus around coordinated, rational, and fair action is achieved.

For Rawls, when we reach the goal of the theory —a certain social order— too many substantial aspects of what a human being is and ought to be, and how the world and the mind work, and ought to work, are presupposed:

Habermas’s theory of communicative action yields the analytical divide of the ideal discourse situation, which offers an account of the truth and validity of judgements of both theoretical and practical reason. It tries to lay out completely the presuppositions of rational and free discussion as guided by the strongest reasons such that, if all requisite conditions were actually realized and fully honored by all active participants, their rational consensus would serve as a warrant for truth or validity. Alternatively, to claim that a statement of whatever kind is true, or a normative judgment valid, is to claim that

²⁸ Gordon, *Op. Cit.*, *passim*

it could be accepted by participants in a discourse situation to the extent that all required conditions expressed by the ideal obtained.²⁹

This correspondence between the presuppositions for rational and free discussion on one side, and truth and validity on the other, is what Rawls calls a “logic in the broad Hegelian sense,” which he links to “all the substantial elements of metaphysical and religious doctrines.”³⁰ Rawls uses Hegel’s *Sittlichkeit* (Ethical life) to describe this comprehensive quality of Habermas theory (as opposed to *Moralität*). Rawls is here citing *Between Facts and Norms*, where Habermas states that the intersubjective practice of argumentation makes it possible for “Hegel’s concrete universal [*Sittlichkeit*] [to] be sublimated into a communicative structure purified of all substantial elements.”³¹

Sittlichkeit is grounded in what Hegel calls “ethical substance”, which is a particular shape of *Geist* (spirit). Take ‘thou shall not kill or steal,’ for example. You don’t need a philosopher to tell you these are sins. It does not require abstract reflection or, more importantly, rational principles. Individuals living in a given society partake in its “ethical substance” and have immediate consciousness of the moral law (in contrast to official statutes passed by legislative bodies). In other words, *Sittlichkeit* represents the (apparently) natural norms or mores in a given society. Of course, this also means that *Sittlichkeit* harbors the biases, inequities, and power relations in that society too. Hegel is not absolutely opposed to *Moralität*, i.e., the philosopher’s morality based on purely rational principles (e.g., Kant’s categorical imperative).

²⁹ Rawls, *Op. Cit.*, 382

³⁰ *Ibid.*

³¹ *Between Facts and Norms*, FG, 280, as quoted by Rawls, *Reply to Habermas*, in *PL*, p 378.

Moralität is the more abstract and academic term, while *Sittlichkeit* is rooted in social practices and concrete forms of social existence. Hegel analyzes abstract rationality, in part, as an index of and reaction to the disintegration of ethical substance (again, the unwritten rules and norms of social existence). It is worth noting that we know the law, however confusedly, before we grasp it in a thoroughly rational manner through philosophical reflection.

Philosophers have historically played a role in bringing about the disintegration of ethical substance and substituting it for the maxims of the transcendental subject. Kant is a prime example, with his critiques of Enlightenment reason and revealed religion, but in his defense he was witnessing the breakdown of ethical substance (i.e., Christian faith and monarchy) during the crisis of the German Enlightenment at the end of the eighteenth century and the beginning of political modernity (e.g., the French Revolution). Hegel's privileged example of the breakdown of *Sittlichkeit*, under the weight of its suppressed inner tensions, is the drama of Antigone where the equally valid laws of Antigone and Creon come into absolute opposition (previously the family and the state, the divine and the secular, were in harmonious complementarity).³²

Habermas makes use of the concept of *Sittlichkeit* because he wants his theory to avoid transcendentalism but he wants to purify *Sittlichkeit* of its negative side: bias, coercion, etc. Habermas thinks *Sittlichkeit* can be sublimated, so to speak, in the way he salvages universality for moral and political purposes while eschewing any transcendental schema à la

³²Some of Hegel's many discussions of *Sittlichkeit* can be found in the chapter on "Spirit" (pp. 263-289) in Hegel, G. W. F. *Phenomenology of Spirit*. Trans. A. V. Miller. Ed. J. N. Findlay. Oxford: Oxford University Press, 2013; Hegel, Georg Wilhelm Friedrich. *System of Ethical Life* (1802/3) and *First Philosophy of Spirit* (part III of the *System of Speculative Philosophy* 1803/4). Trans. Henry Siltan Harris and Thomas Malcolm Knox. Albany: state U of New York, 1979; and Hegel, Georg Wilhelm Friedrich. *Elements of the Philosophy of Right*. Trans. Allen W. Wood and Hugh Barr Nisbet. Cambridge: Cambridge UP, 2012.

Kant. It is not that we must will our maxims to be universal, but that we must be willing to put ourselves in the shoes of our interlocutor when it comes to argumentation:

The practice of argumentation recommends itself for such jointly practiced, universalized role taking. As the reflexive form of communicative action, it distinguishes itself socio-ontologically, one might say, by a complete reversibility of participant perspectives, which unleashes the higher-level intersubjectivity of the deliberating collective.³³

For Habermas, the intersubjective space created by a reflective attitude that takes into account the other's perspective simultaneously solves the problem of metaphysical universality present in *a priori* transcendentals, and of the particularity present in relativism and historicism. This act of communication infused with moral-empathetic and rational-idealizing attitudes sublimates the concrete universal of the Hegelian tradition³⁴. Rawls uses the Hegelian 'admission' he sees in the quote above as a proof that Habermas goes beyond what his (Rawls's) "political, not metaphysical" rule and his *method of avoidance* warrants. Habermas's consensus theory, Rawls thinks, is built into the substance of metaphysics. The 'admission' is enough for Rawls, and thus he just states the connection with metaphysics without deigning to further argue the point: this is consistent with his general strategy of avoiding metaphysical parlance. Interestingly, in this sense, Rawls clarifies his conception of metaphysics in a footnote, "as being at least a general account of what there is, including fundamental, fully general statements, for example, the

³³ Jurgen Habermas, *Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy*, (Cambridge, MA: MIT Press, 1996) 280, as quoted by Rawls, *Op. Cit.*, 378.

³⁴ A quick word about this technical concept as I understand it: the individual thing (be it person, mineral, or animal) is not itself universal but rather particular. Any spatio-temporal thing is by definition this or that particular thing. This or that particular thing, however, is a bearer of many universal attributes. Insofar as I am Jordi Graupera, I am just a concrete particular guy, but insofar as I am a rational being or mortal, I am concrete instantiation of universals (e.g., rational, mortal, etc.). As I understand it, this idea of sublimating means the process by which the individual suspends his particular position (here and now, with this or that predilection or passion, with this or that identity or idiom) and grasps the universals within him, that is, finds what in himself or herself he or she has in common with others, such as moral principles and material needs, etc. What makes Habermasian sublimation through discourse so metaphysical in Rawls's eyes is that Habermas uses the anthropology described in his theory of communication as a way to universalize certain moral principles that then serve as a basis for social consensus. The issue concerns the tension between the particular interlocutors in a communicative relationship and the universal (inherent in the conceptual terms that make the medium of communication: language).

statements ‘every event has a cause’ and ‘all events occur in space and time,’” which express criteria for demarcating metaphysics rather –and purposely—superficial, in line with his attempt to avoid the subject entirely. To see if Rawls’s critique, which basically states that Habermas’s entire edifice is too comprehensive to qualify as non-metaphysical, holds we must not avoid a metaphysical interrogation, analyzing each component, with the help of other critics, and with a view to the role it plays in his theory.

3. Walzer’s Critique of Habermas’s Thickness

In order to evaluate Walzer’s critique of Habermas, it will be helpful to consider the entire passage from *Between Facts and Norms* that comes just before the quote that Rawls cites as evidence of substantial content. It provides a richer and more complete picture of this sublimation, and it opens up the question for a more thorough analysis, and other critiques, such as those put forth by Walzer and others. We will unpack it further in what follows in this chapter. Habermas’s quote contains five central claims (my numbers):

(1) Whenever we want to convince one another of something, we always already intuitively rely on a practice in which we presume that we sufficiently approximate the ideal conditions of a speech situation specially immunized against repression and inequality. (2) In this speech situation, persons for and against a problematic validity claim thematize the claim and, relieved of the pressures of action and experience, adopt a hypothetical attitude in order to test with reasons, and reasons alone, whether the proponent’s claim stands up. (3) The essential intuition we connect with this practice of argumentation is characterized by the intention of winning the assent of a universal audience to a problematic proposition in a non-coercive but regulated contest for the better arguments based on the best information and reasons. (4) It is easy to see why the discourse principle requires this kind of practice for the justification of norms and value decisions: whether norms and values could find the rationally motivated assent of all those affected can be judged only from the intersubjectively enlarged perspective of the

first-person plural. (5) This perspective integrates the perspectives of each participant's worldview and self-understanding in a manner that is neither coercive nor distorting.³⁵

The first claim is the core of Habermas's discourse theory: there are certain "ideal conditions" that are "always already intuitively" assumed when we enter into an exchange of reasons. Even if they are not completely actualized, they are "sufficiently" actualized. These conditions are the basis from which Habermas builds both the procedure for democratic deliberation – which guarantees that institutions will remain neutral between the "substantive" positions in contest – and the moral content of democratic regimes that aspire to universality.

These "ideal conditions" are "immunized against repression and inequality." It's precisely here that Michael Walzer finds Habermas's project wanting – or, better said, overstepping. In his *Thick and Thin*, Walzer argues against the common – and Kantian³⁶ —picture of morality as a "(thin) set of universal principles adapted (thickly) to these or those historical circumstances."³⁷

This picture suggests "mistakenly" that "men and women everywhere begin with some common idea or principle or set of ideas and principles, which they then work up in many different ways." But, according to Walzer, the "intuition" that "they start thin" and "thicken with age", is "wrong."³⁸ On the contrary: "Morality is thick from the beginning, culturally integrated, fully resonant, and it reveals itself thinly only on special occasions, when moral language is turned to

³⁵ Jürgen Habermas, *Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy*, (Cambridge, MA: The MIT Press, 1996), 228

³⁶ According to Richard Rorty, "Justice as Larger Loyalty" in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 45

³⁷ Michael Walzer, *Thick and thin*. (Notre Dame, IN: Univ. of Notre Dame Press, 1994), 4

³⁸ *Ibid.*

specific purposes.”³⁹ This move allows Walzer to characterize Habermas’s project⁴⁰ along the lines of thickness and thinness. Habermas’s detranscendentalized intersubjective ideal speech situation is meant to be thin, minimal, almost merely procedural. Politically, the reason is, among others, neutrality. The independence of the political realm with respect to the substance of the ethical realm is argued by Habermas precisely from the principle of state neutrality: “The requirement of a justification of law interwoven with politics that is neutral with respect to world-views explains why the principle of democracy acquires an independent status vis-a-vis the principle of morality.”⁴¹ Habermas’s “ideal conditions” are meant to be that set of thin universal principles that we later work out thickly into particular metaphysical substantial world-views. This is what Walzer sees as a transformation of thin minimalism into “procedural terms”: “a thin morality of discourse or decision that governs every particular creation of a substantive and thick morality.”

Minimalism, on this view, supplies the generative rules of the different moral maximums. A small number of ideas that we share or should share with everyone in the world guides us in producing complex cultures that we don’t and needn’t share –and so they explain and justify the production. Commonly, as in Jürgen Habermas’s critical theory, these shared ideas require a democratic procedure of articulate agents, men and women who argue endlessly about, say, substantive questions of justice.⁴²

For Walzer, the problem with this thin-to-thick picture of democratic deliberation is that the thin procedural design “turns out to be rather more than minimal.” Habermas’s “rules of engagement” are “designed to ensure that the speakers are free and equal” in a specific thickly

³⁹ *Ibid.*

⁴⁰ According to Rorty, “Michael Walzer is at the other extreme from Habermas.” Rorty, *Op. Cit.*, 45

⁴¹ Jürgen Habermas, “On the Architectonics of Discursive Differentiation: A Brief Response to a Major Controversy”, in *Between Naturalism and Religion*, trans. Ciaran Cronin (Cambridge, UK: Polity Press, 2008), 78

⁴² Walzer, *Op. Cit.*, 12-13

way.⁴³ Indeed, in a controversy with Karl-Otto Apel⁴⁴, Habermas concedes that the “normative” content of the pragmatic presuppositions that participants in argumentation “must implicitly accept” are “transcendental” in the weak Davidsonian sense of being “unavoidable.” This is why he calls them “the starting point.”⁴⁵ But, in his view, this weak transcendentalism does not overstep the boundaries of metaphysics. Let us see these presuppositions one by one.

The four most “important” of these unavoidable pragmatic presuppositions receive different formulations in Habermas’s different works, but they can be summarized as a) “inclusivity,” b) “equal opportunity,” c) “truthfulness,” and d) absence of “coercion,”⁴⁶ – where a), b), and d) subject “one’s behavior in argumentation to the rules of egalitarian universalism.”⁴⁷ Walzer rejects the idea that these presuppositions are minimal. Their normative content, translated into procedural rules, is meant “to liberate” participants in argumentation “from domination, subordination, servility, fear, and deference. Otherwise, it is said, we could not respect their arguments and decisions.” This is not minimal because:

Once rules of this sort have been laid out, the speakers are left with few substantive issues to argue and decide about. Social structure, political arrangements, distributive standards are pretty much given; there is room only for local adjustments. The thin morality is already very thick -with an entirely decent liberal or social democratic thickness.⁴⁸

⁴³ Walzer, *Op. Cit.*, 12-13

⁴⁴ Habermas, *Op. Cit.*, 82 and ss

⁴⁵ *Ibid.*, 82

⁴⁶ *Ibid.*

⁴⁷ Jürgen Habermas, “Communicative Action and the Detranscendentalized ‘Use of Reason’”, in *Between Naturalism and Religion*, trans. Ciaran Cronin (Cambridge, UK: Polity Press, 2008), 50

⁴⁸ Walzer, *Op. Cit.*, 12-13

However, once we go over these presuppositions in detail, the problem is not only that they are ideologically biased, say, in favor of social-democracy. I propose that the key qualifiers that make these presuppositions “norms” in any meaningfully applicable sense suggest that its normativity is a matter of belief. When Habermas speaks of “inclusivity” as the first “unavoidable” presupposition he specifies it as the rule that “no one who could make a *relevant* contribution may be prevented from participating,” (my italics).⁴⁹ What is a “relevant contribution”? Relevant for what or for whom? Doesn’t disagreement, at least in part, concern what counts as relevant? Doesn’t political oppression, at least in part, concern what counts as relevant? Doesn’t the disagreement over what counts as relevant partly make up the “thick” difference between world-views? This relevance does not belong to the intersubjective space, but to the subjective perspective of world-views.

The second “unavoidable” presupposition, “equal distribution of communicative freedoms,” translates into the rule: “everyone has an *equal opportunity* to make contributions,” (my italics).⁵⁰ What is an equal opportunity? Material equality? Enjoying the same time to contribute? Or should it take into account, say, historical discrimination when it comes to enlarge one’s contribution? Should there be affirmative action for marginalized views or individuals in minority groups? Should there be a proviso for status-quo bias? Is it equally fair that one argues better than another, or has received a better, and more expensive, rhetorical education? What does an “opportunity” entail in the context of heterodoxical life-experiences? Isn’t the very question of

⁴⁹ Jürgen Habermas, “On the Architectonics of Discursive Differentiation: A Brief Response to a Major Controversy,” *Between Naturalism and Religion*, trans. Ciaran Cronin (Cambridge, UK: Polity Press, 2008), 82 (emphasis added in “relevant”). The word “relevant” is repeated in the slightly different formulation of *Communicative Action and the Detranscendentalized Use of Reason*, in *Ibid.*, 50 “no one could make a relevant contribution concerning a controversial validity claim must be excluded”

⁵⁰ *Ibid.*, 82 (emphasis added on “equal opportunity”). The slightly different formulation of *Communicative Action and the Detranscendentalized Use of Reason*, *Ibid.*, 50: “everyone must have the same opportunity to speak to the matter at hand”

equality – its meaning, scope, and necessity – what is at stake in any political discussion? Isn't equality more of a desideratum than a presupposition?

The fourth presupposition completes “the rules of egalitarian universalism,” and is affected by the same problems as the previous two: “communication must be free from *restrictions* that prevent the better argument from being raised and determining the outcome of the discussion,” (my italics).⁵¹ Again, we can repeat the same questions suggested by “relevance” and “equality”: what counts as a restriction? This presupposition, however, adds another layer to the problem: the notion of the better argument. Or what Habermas’s calls “the unforced force” of the better argument. Let us leave aside for a moment that what counts as a restriction for the prevalence of a ‘better argument’ is again part of the political discussion that this content is meant to arbitrate, and the pointed question that Walzer asks —“How can it ever be certain that the better argument in any particular conversation is the best possible argument?”⁵² More problematic here is that the notion of a “better” argument —the exact purpose of the conversation— becomes in this “rule” external to the conversation. Contrary to what the pragmatic theory of meaning would suggest if taken to its most contextualist conclusions, for Habermas what makes an argument better and thus prevailing is not agreement. On the contrary: it is the fact that the argument is better, under the appropriate conditions, —i.e., ideal conditions procedurally regulated— which produces agreement.

⁵¹ *Ibid.*, 54 (emphasis added) The alternative formulation of the fourth presupposition in *On the Architectonics of Discursive Differentiation: A Brief Response to a Major Controversy*, *Ibid.*, 82 is rather different but equally problematic, and for the same reasons: “absence of contingent external constraints or constraints inherent to the structure of communication: the yes/no positions of participants on criticizable validity claims should be motivated only by the power of cogent reasons to convince.”

⁵² Michael Walzer, “A Critique of Philosophical Conversation” in *Hermeneutics and Critical Theory in Ethics and Politics*, (Cambridge: MIT Press, 1990) 189.

I'll go over the problematic of the better argument further below, but before that, to complete Walzer's critique of Habermas, it must be noted that these not-so-thin procedural "rules of engagement" not only presuppose a certain political horizon, as I said above, but also "constitute a way of life."

How could they not? Men and women who acknowledge each other's equality, claim the rights of free speech, and practice the virtues of tolerance and mutual respect, don't leap from the philosopher's mind like Athena from the head of Zeus. They are creatures of history; they have been worked on, so to speak, for many generations; and they inhabit a society that "fits" their qualities and so supports, reinforces, and reproduces people very much like themselves. They are maximalists even before they begin their rule-governed discussions.⁵³

Walzer is thus claiming that these allegedly universal minimal norms express in truth a historically grounded, thick, ideological, political worldview. On the one hand, these rules govern political dialogue because the institution that must govern the plurality of voices in dispute must be understood as neutral. But this posture of neutrality is already smuggling in political content before the conversation even begins: a historically situated substantial content.⁵⁴ The presence of neutrality presupposes the need for intersubjective consensus which, in its turn, presupposes the kind of procedural rules that are pregnant with historical and anthropological presuppositions. But in Habermas's ideal conversation what is really idealized is the speakers themselves, "designed or programmed in such a way that certain words, and not others, will come naturally to their lips."⁵⁵ This stance by Walzer belongs to his general critique of "Philosophical Conversation" in liberal thinkers such as Rawls, Ackerman or Habermas, which follows an

⁵³ Michael Walzer, *Thick and Thin*, (Notre Dame: University of Notre Dam Press, 1994) 12-13

⁵⁴ This content can be related to Rawls's critique of *Sittlichkeit*.

⁵⁵ Michael Walzer, "A Critique of Philosophical Conversation," in Michael Kelly (Ed.) *Hermeneutics and Critical Theory in Ethics and Politics*, (Cambridge: MIT Press, 1990), 185

intuition by Seyla Benhabib, according to which, in Habermas's case, "one extracts from the ideal speech situation what one has...put into it,"⁵⁶ like the rabbit pulled out of the prestidigitator's hat. Walzer's critique of this kind of "designed" conversation focuses on the idea that the preconditions are excluded from the conversation itself, even though it is clear that even philosophers themselves, to name a particular exemplar of civil discourse, seem unable to reach minimal agreement on the meaning, scope, and truth of these preconditions.⁵⁷

Design cannot help us, since all its elements, formal and substantive, necessarily precede hypothetical speech: they have to be worked out (and are worked out) independently of any ideal procedure.⁵⁸

The reason Walzer can attack with alacrity Habermas's rules of engagement in contextualist terms has to do with Habermas's attempt to preserve a certain principle of universality. In Habermas's design, speakers are constrained to assert only what can be universalized, at the expense of situated interests – a strategy that Rawls, albeit with some differences, also adopts with his veil of ignorance. For Walzer, universalization is a powerful constraint, with specific theoretical aims: it is intended to rule out bargaining and compromise (the negotiation of particular interests), and to press the speakers toward a preordained harmony. It arms social theory "against the indeterminacy of natural conversation."⁵⁹ The contrast between what is repressed – the natural bargaining of interests – and the usual indeterminacy of "real talk" suggests to Walzer "the dominance of design over discourse."⁶⁰ The means by which Habermas

⁵⁶ Seyla Benhabib, *Critique, norm, and Utopia: a Study of the Foundations of Critical Theory* (New York: Columbia, 1986), pp292-293

⁵⁷ The conversational equality that the presuppositions imply, for example, reflects "a hypothetical social equality (but is not the conversation supposed to produce, among other things, an argument for or against social equality?)" Michael Walzer, *Op. Cit.*, 185

⁵⁸ *Ibid.*, 194

⁵⁹ *Ibid.*, 186-187

⁶⁰ *Ibid.*, 186

accomplishes subordinates contingent discourse to rational design is by embedding in conversation the need of universality as a constraint.

With this constraint, Habermas stitches up procedure with (weak) transcendental – “unavoidable” – presuppositions, which are preconditions for any agreement, and even for communication itself, between subjects spilled over this mysterious transsubjective space. Indeed, where Walzer sees thickness in moral-historical terms, I propose that this thickness is also metaphysically committed in the sense that it is not only that they are historically rooted, as some non-metaphysical views may be, but also that this particular moral-historical view carried over Habermas presuppositions is metaphysically committed, as I will try and show.

Out of the four aforementioned presuppositions, only the third remains intact. It happens to be the most important when it comes to highlighting the metaphysical commitment that I claim is presupposed by Habermas: “participants must *mean* what they say,” (my italics).⁶¹ I assume that Habermas is saying here that interlocutors who try to reach an agreement should not deceive one another. Good faith communication, in which interlocutors do not engage in deception, is but one meaning of *meaning what you say*. Indeed, the idea of *meaning what one says* stresses the connection between the possibilities of language – including the interlocutor’s specific capacities for linguistic expression – with the subjectively available content of one’s mind. This presupposition assumes that such a connection is actually available to us and that it’s a feature of language: to truly mean what one says. This third presupposition is striking because, even though it mainly asks of the interlocutor to assume that the other is being truthful, and as such it

⁶¹ Jürgen Habermas, *Op. Cit.*, 82 (emphasis added in “mean”). Exactly the same formulation in *Communicative Action and the Detranscendentalized ‘Use of Reason’*, in *Ibid.*, 50

describes a moral attitude, it actually implies a broader epistemological picture. It might be helpful to reverse this proposition and see that it is not tautologously identical in reverse form: to say what one means does not precisely equal meaning what one says. The latter is more a question of deception/honesty, while the former (to say what one means) is more a question of adequately or inadequately, as the case might be, expressing one's intended meaning. Habermas conflates both meanings when he assumes that truthfulness bridges the gap between subjects in the intersubjectives space of communication. Such an epistemological picture is indeed related to Habermas's description of the way thoughts work in communicative action. Recall that for Habermas, meaning is pragmatic; it is obtained not through accuracy of representation, but through agreement. There is no way one can mean what one says before the social use of language unless there is *a priori* a structure of rationality that makes the connection between meaning and saying one of correspondence. The description of the possibility of correspondence is pre-social, pre-pragmatic, from a high altitude perspective, from God's vantage point. In other words, such a description would be metaphysical.

Meaning what one says is as a presupposition is different from the other three not only because the other three presuppositions subject one's behavior to the constraint of universal egalitarianism, as Habermas concedes, but also because it presupposes a subject with transcendental qualities. In order for "saying" to be coextensive with "meaning" and for meaning to be able to breed agreement, Habermas needs "linguistic expressions" to have the "ability to preserve the same meaning for different people in different situations." This ability is predicated upon two "preconditions:" "The fact that thought overshoots the boundaries of a spatio-temporally individuated consciousness, and the independence of the ideal thought content from

the stream of experience of the thinking subject.”⁶² The possibility of meaning what one says and saying one what means overcomes the situated nature of the subject and its perspective, and solves the problem of the potential incomunicability of experience. And this way, his theory fails in overcoming this modern dualist picture and makes meaning extra-pragmatic. The endurance of this theoretical picture is precisely what articulates praxis, and not the other way around, as it was initially suggested.

4. Gordon’s Critique of Habermas’s *Einverständnis*

With this in mind, if we now go back to the four remaining claims that Habermas made in the quote from *Between Facts and Norms* that I cited above, the need for this more-than-weak transcendental necessity slowly outcrops. I want to relate – and modify ever so slightly so as to show its metaphysical import – Gordon’s critique of Habermas with the second claim in that fragment: “In this speech situation, persons for and against a problematic validity claim thematize the claim and, relieved of the pressures of action and experience, adopt a hypothetical attitude in order to test with reasons, and reasons alone, whether the proponent's claim stands up.”⁶³ Reasons appear here to be these units of thought, pure in their freestanding character once they are thematized, independent of action and experience, that is, independent of the material world and its existential development –otherworldly, if I may. A speech act contains implicit reasons in the form of validity claims that are accepted or rejected by interlocutors; when there is no agreement, discourse appears. The goal of discursive meta-arguments is to make

⁶² *Communicative Action and the Detranscendentalized ‘Use of Reason’*, in BNR, p54

⁶³ Jürgen Habermas, *Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy*, (Cambridge, MA: The MIT Press, 1996), 228

explicit and thematize —making them central— these implicit reasons: these units of thought that preserve meaning from subject to subject. Reasons become the standard by which to “test” whether the former implicit validity claims actually “stand up.” Testing reasons under the ideal conditions is what propels understanding, i.e., the *telos* of language. The ideal conditions actually liberate reasons from experience, action, and all the non-ideal conditions that occasion misunderstanding and disagreement in the first place. James Gordon believes that in the relationship between agreement and understanding we see a non-sequitur around a certain “ambiguity” in Habermas’s use of the terms *Verständigung* and *Einverständnis*:

The claim that social order rests on shared understandings and meanings is significantly different from the claim that social order rests on intersubjective agreement. Shared understandings and shared meanings might fall well short of agreements. (...) Habermas has often been accused of the non-sequitur that the members of a society, simply by virtue of understanding what one another mean, will adhere to the same social and moral rules.⁶⁴

It is this connection between understanding and agreement that makes it necessary for the presuppositions of the ideal conditions to include “meaning what you say.” It is the only way a correspondence between political order and understanding can be said to rest entirely on intersubjectivity. When agreement and understanding are completely separated, as Gordon suggests, intersubjective agreements are still in place, but they might depend on strategic reason, *modus vivendi*, alliance with “lesser evils”, or simply relations of force. All these are very contingent and rather too fragile to constitute a formal rationality with the potential for universal validity. And it also threatens the basis for political legitimacy in the Habermasian sense; after all, his project aims at unveiling a profound, morally relevant, legitimacy that would go beyond contractualism. This is why, in the absence of traditional metaphysical-theological justifications,

⁶⁴ Gordon, *Op. Cit.*, 39

intersubjective understanding/agreement must be center stage. Indeed, and most important, agreement based on intersubjective understanding is what allows Habermas to establish a normative standard by which to salvage the status of rationality. But where Gordon's position amounts to the following question: "what if we understand one another and share meanings, but don't agree with one another on the rules that should govern us, nor do we accept to live by them?" – I propose the converse question: "what if we agree but in truth do not understand one another?" The fact that interlocutors claim to endorse the same rules does not entail that they understand the same thing by them, nor that they share the "meanings" that underline them. Just as *tout comprendre* does not entail *tout pardonner*, *être d'accord* does not entail mutual understanding.

Not only does agreement not require understanding, but the two are often at odds with each other. Agreement can be, and often is, an "externalization" of understanding, a kind of pseudo-understanding, in which different parties adhere to, or uphold, the same authority (be it to a person or to an idea) without the subjective elements of understanding. In Habermas' ideal speech conditions, we are all approaching understanding and overcoming mere belief (be it true or false). The (usually) unfulfilled desideratum is understanding not agreement. Agreement is a kind of natural or primitive state, while understanding is not. The path to understanding must of course first pass through disagreement (particularly those we agree with in our "primitive" or sub-rational condition, which perhaps amounts to agreement via *identification*). Habermas is probably right to say (or merely imply) that the possibility of *universal* agreement requires understanding and not mere belief.

Habermas's theory requires a certain condition, namely, that an interlocutor mean what she says and say what she means, and at the same time it requires that the auditor understand what the speaker says to be what the speaker means. Furthermore, the auditor would understand it in such a way that several already shared meanings are at work, so that meaning can become a social affair. Simultaneously, with all the universally egalitarian ideal preconditions in place, the hearer becomes a speaker and the roles are reversed. Reasons are given and their meanings remain from one agent-speaker to the other in a purely rational-pragmatic exchange. And then, understanding gives way to agreement due to the clear unforced force of the better argument. Thought becomes the standard, the object captured in language, the object captured in meaning. You can mean what you say – you must mean what you say – in order for meaning to “overshoot the boundaries of” contingency, i.e., of “a spatio-temporally individuated consciousness.” The ‘ideal thought content,’ the meaning of what you say, is independent from “the stream of experience of the thinking subject.”⁶⁵

Habermas is aware of the problem this picture of intersubjective exchange of reasons poses. He discusses this “independence” in the context of showing up the shortcomings of Frege's philosophy of language. Frege distinguished between representations and thoughts, so that representations belong to a particular subject “who is individuated in space and time”, whereas thoughts overstep the boundaries of ‘individual consciousness *unaltered*.’⁶⁶ The later Frege, however, was led to favor a certain reification of language with the aim of overcoming the psychologism implied in the “mentalistic two-world architectonic, which juxtaposes a subjective

⁶⁵ Jürgen Habermas, *Communicative Action and the Detranscendentalized 'Use of Reason', Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy*, (Cambridge, MA: The MIT Press, 1996), 54

⁶⁶ *Ibid.*, 53-55

world of representations to an objective world of things.” In order to overcome this dualism, he thought this picture had to be supplemented by “a third, ideal world of propositions,” thereby creating a further problem, a double gap between propositions and the world, and propositions and representing subjects. In other words, it does not solve the problem of how to situate reason without lapsing into a kind of Platonic transcendentalism. While Habermas is fully aware of this problem, and his theory represents a complex attempt at solving it, he also concedes that “the members of a language community must assume in practice that the grammatical expressions they utter have a *general* meaning that is *identical* for all participants across their various contexts of use.”⁶⁷ Something must remain. Even if the “ideal generality of the meaning of grammatical expressions” often “proves to be wrong,” it is nevertheless “unavoidable when language is used with the goal of reaching understanding.”⁶⁸ To solve this problem, inherited from the mentalistic picture of reason, two avenues are open: “Either the concept of reason itself is liquidated along with the mentalistic paradigm, or it is freed from its mentalistic framework and is transposed into the concept of communicative reason.”⁶⁹ Habermas of course chooses the second avenue. The problem is solved, apparently, by the adopting the stance of communicative action and discursive argumentation as a social practice, the rules of which – and the assumptions behind which – are mastered in practice because “they could not regulate their own application without trapping the agent in an infinite regress.”⁷⁰

Participants in argumentation who persist in trying to reach understanding remain, on the one hand, bound into a shared practice; on the other hand, they must adopt a justified

⁶⁷*Ibid.*, 54

⁶⁸ *Ibid.*, 53-55

⁶⁹ *Ibid.*, 56

⁷⁰ *Ibid.*, 67

stand on the thematized validity claims, that is, they must be led by the gentle force of reasons to their own autonomous judgement.⁷¹

Behind the social practice and the possibility of transmitting meanings intersubjectively, then, there is not an explicit reification of language, à la Frege, resting on a sharp distinction between thoughts and representations, nor an ontological picture that reinstates the even older dichotomy of world and mind, but rather a fluid picture of social practice that depends on the assumption of rationality. Once again, this rationality is present in the normative presuppositions of communication and it is unavoidably included in the norm “mean what you say” if the “gentle force of reasons” must guide a subject’s judgement. That the only *being* behind all social *becoming* is the assumption of rationality – expressed in the affirmation of the *reality* of communication – is seen in the fact that for Habermas *unconditional* validity claims “point beyond every factual agreement”: “What is accepted as rational here and now can turn out to be false under more favorable epistemic conditions, before a different public, or when confronted with future objections.”⁷² This seems like a strong version of contextualism and it is only mitigated by the assumption of rationality, a rationality that links the subjects in the intersubjective encounter, which makes their utterances meaningful, and which “gently” makes reasons convincing in validity. The assumption of truthfulness that we see in the presupposition of a correspondence between what one means and what one says, which expresses the conditions of possibility of intersubjectivity, is the last bulwark of “unavoidable” transcendence, and the core of a metaphysical anthropology that Habermas tries to “detranscendentalize” with his social conception of situated reason. This anthropological presumption is both affirmed and fostered

⁷¹ *Ibid.*, 75

⁷² *Ibid.*, 75

when these ideal *preconditions* are turned into political norms, and is the crack through which the light of metaphysical commitments flood in.

5. Rorty's Critique of Habermas's 'better argument'

Rorty sees this bulwark precisely in the appeal to the “unforced force of the better argument,” which is the supposed *telos* all these preconditions ultimately point to, even if it is more like a vanishing horizon than a concrete destination. Going back to the five claims from the quote that opened this section, I want to revisit the already analyzed second claim, together with the third one, which will be the focus of this epigraph. The aim of my analysis is to highlight the force of the metaphysical assumption behind Habermas's argument when Rorty criticizes it from a radically contextualist position.

Habermas says:

(2) In this speech situation, persons for and against a problematic validity claim thematize the claim and, relieved of the pressures of action and experience, adopt a hypothetical attitude in order to test with reasons, and reasons alone, whether the proponent's claim stands up. (3) The essential intuition we connect with this practice of argumentation is characterized by the intention of winning the assent of a universal audience to a problematic proposition in a non-coercive but regulated contest for the better arguments based on the best information and reasons.

To bring out the metaphysical commitment behind the last claim, I shall invoke two objections that Rorty poses against Habermas. The first has to do with Habermas's attempt at salvaging the notion of universal validity. Both Rorty and Habermas “mistrust metaphysics,” but, as Rorty puts it, Habermas thinks that it is philosophically sound and politically necessary to find a “metaphysics-free interpretation of the notion of universal validity in order to avoid the

seductions of Romanticism,” whereas Rorty thinks that “that notion and metaphysics stand or fall together.”⁷³ This criticism is an instance of a more general point by Rorty concerning the understanding of reason as a set of social practices. Rorty thinks, like Habermas, that it is important to do away with the philosophy of consciousness. One of the most important aspects of the long tradition of philosophy of consciousness, according to them both, is the idea that our minds and the nature of things are somehow connected, so that we can have access to the “really real,” —given a set of rules of rationality. Over and against this notion, the idea of a communicative reason that Habermas champions means an attempt at giving up all non-human standards of knowledge in favor of human consensus, socially achieved. For Rorty, however, changes in culture and science, changes in the way we see religion or the meaning of life, have very little to do with actual convincing and very much to do with the slow changing of vocabularies. Rorty believes that once we give up the attempt at finding correspondence between some sentences and the nature of the world, once we stop looking for a language that can speak the jargon of nature, and once we stop understanding what’s real as a set facts able to be chunked in true sentences, we can see that there is no intrinsic superiority of one vocabulary over another. Some aspects of the world may warrant us to hold certain beliefs, and we can express those beliefs in propositions in a given vocabulary, but the vocabulary itself is just a tool, ready to use when it is useful, slowly becoming irrelevant when desuetude makes it obsolete, or simply when its inconveniences multiply, and then a new vocabulary can appear as a substitute.⁷⁴ In this paradigm, Habermas’s attempt at finding a non-metaphysical standard of universal validity can be seen as the attempt to impose his vocabulary as the one and true vocabulary capable of

⁷³ Richard Rorty, “Grandeur, profundity, and finitude”, in in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 78

⁷⁴ Richard Rorty, “Response to Habermas” in Robert Brandom (Ed.), *Rorty and his critics*. (Malden: Blackwell, 2008)

delivering the truth. This is why Rorty regards Habermas's insistence upon retaining the ideal of universal validity "as an unfortunate concession to Platonism."⁷⁵

By hanging on to it, it seems to me, Habermas remains in thrall to the philosophical tradition that burdened us with the idea of reason as a human faculty that is somehow attuned to the really real.⁷⁶

Going back to Habermas's quote, the adoption of a "hypothetical attitude" is aimed at purifying the moment of discursive exchange of reasons —i.e., the moment of thematizing a disagreement — from all contingencies. What Habermas calls "reasons and reasons alone" is what is left once we are relieved of the pressures of "action and experience". Habermas states that we should connect the fact that "the validity claims raised *hic et nunc* and aimed at intersubjective recognition or acceptance can at the same time overshoot local standards for taking yes/no positions," with the overall idea that "this transcendent moment alone distinguishes the practices of justification oriented to truth claims from other practices that are regulated merely by social convention."⁷⁷ Habermas is trying to salvage universality from the wreckage of the philosophy of consciousness by substituting the practice of justification of truth claims – with reasons, and reasons alone – as a mark of a shared capacity for rationality, for the *anciene* notion that deep inside we somehow possess a privileged mode of access to universal truth, in a Platonic sense.⁷⁸ This relief of contingency towards a moment of transcendence beyond parochial social conventions is built into the belief that in the very act of communication, when its aim is understanding, there lies a

⁷⁵ Rorty makes a similar point against Rawls in his Rorty, Richard. "Chapter 11. "The Priority of Democracy to Philosophy." in *Prospects for a Common Morality*.

⁷⁶ Richard Rorty, "Grandeur, profundity, and finitude", in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 78

⁷⁷ Jürgen Habermas, *Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy*, (Cambridge, MA: The MIT Press, 1996), 15

⁷⁸ That is, that being is ultimately ideal in the sense of rational and the sensible is merely a contingent "copy", a mere play of shadows, poor images of true being.

structural set of universal *nomoi*, which are the only constitutive aspects of human nature that Habermas allows in the justification of his democratic, liberal politics. The “intention of winning assent of a universal audience” makes sense only if this shared nature informs the transcendent moment, relieves us of contingency and social conventions, and gives us access to reasons and reasons alone.

Ultimately this argument is circular: reasons and contingency mutually define each other, in a differential manner. A reason is a pure reason only in the absence of contingency, and contingency remains only in the absence of pure reasons, but neither obtains a clear and distinct demarcation in the context of communication without a certain tension in its conceptual relation with its contrary. However, for our purposes, the key aspect of this argument is that the presence of reasons alone is what makes it possible for the better argument to win assent. All the characteristics at play in this moment define what Habermas understands by “a reason” and a “better argument”: universality, relief from contingency, accurate information, non-coercive attitude, a uniform audience. All of the above are meant to cancel the cacophony that usually makes it hard for the better arguments to win assent...universally. The relevance of this assertion cannot be overstated. If reason is a social practice, as opposed to the ultimate, metaphysical foundation of the self, to claim that there is a hypothetical attitude that allows for a universally better argument to garner assent amounts to saying that the convincing aspect of such an argument will, in theory, remain unmatched in front of every single audience, by virtue not of the sociality of reason, but of the intrinsic quality of the argument itself. Politically, this is problematic and contradictory for the same reasons – specifically tolerance – that made liberalism desirable in the first place, since, as Rorty puts it, “the idea that there is one right thing

to do or to believe, no matter who you are, and the idea that arguments have intrinsic goodness or badness, no matter who is asked to evaluate them, go hand in hand.”⁷⁹ For Rorty, the appeal to a transcultural ‘better argument’ in Habermas does not go much further than a secularized theological argument:

More generally, the idea of ‘the better argument’ makes sense only if one can identify a natural, transcultural relation of relevance, which connects propositions with one another so as to form something like Descartes’ ‘natural order of reasons.’ Without such a natural order, one can only evaluate arguments by their efficacy in producing agreement among particular persons or groups. But the required notion of natural, intrinsic relevance – relevance dictated not by the needs of any given community but by human reason as such – seems no more plausible or useful than that of a God whose Will can be appealed to in order to resolve conflicts between communities. It is, I think, merely a secularized version of that earlier notion.⁸⁰

Habermas disagrees with Rorty on this crucial aspect because he thinks that renouncing the idea of a better argument leaves us defenseless against the bewitching rhetoric of fascists and romantics; that, somehow, we need to overcome the idea that there is no purely rational difference between the persuasion propaganda of a Nazi and the logical discourse of a social democrat. More generally, in his ideal speech situation, the supposedly “better” arguments are irreducible to the expression of a given tradition, romantically dug out of history, shining through our ethnic identities. Habermas fears that Rorty’s contextualist position leads ineluctably to an version of moral relativism that eschews the normative force of rationality.⁸¹ We might be able to agree with both in this if we were to conclude here that that’s why we actually need metaphysics.

⁷⁹ Richard Rorty, “Justice as Larger Loyalty” in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 54

⁸⁰*Ibid.*, 54

⁸¹ BRANDOM, Robert (Ed). *Rorty and his critics*. (Malden: Blackwell, 2008), 11

At the core of this dispute there lies the tension between justification and truth, as I mentioned above. In Habermas's system, a justification that wins assent under ideal conditions of rational discourse implies that there exists an intrinsic relationship with truth, even if this relationship is merely a regulative ideal. It is a "fact" that "a justification successful in our justificatory context points in favor of the context-independent truth of the justified proposition."⁸² In Rorty's ears, this sounds like the smuggling in of metaphysical content, in a least two senses. First, it assumes a "convergentist" theory of truth, dating back to Peirce, according to which the greater the audience that accepts a given justification the greater the chance that the assertion is true in a context-independent way.

Such a belief sees the space of reasons as finite and structured, so that as more and more audiences are satisfied more and more members of a finite set of possible objections are eliminated. One will be encouraged to see the space of reasons in this way if one is a representationalist, because one will see reality (or at least the spatio-temporal hunk of it relevant to most human concerns) as finite and as constantly shoving us out of error and toward truth, discouraging inaccurate representations of itself and thereby producing increasingly accurate ones. But if one does not take knowledge to be accurate representation of reality, nor truth as correspondence to reality, then it is harder to be a convergentist, and harder to think of the space of reasons as finite and structured.⁸³

Second, it relies on the idea of an ideal audience:

The notion of universal validity can only be explicated by invoking the notion of the ideal audience - the audience that has witnessed all possible experiments, tested all possible hypotheses, and so on. To explain what any of these idealizations amount to you have to resort to the idea of an audience whose standards of justification cannot be improved on. But such an audience seems to me as hard to envisage as the largest number, the largest set, or the last dialectical synthesis - the one which cannot become the thesis of a new dialectical triad.⁸⁴

The difference between Rorty's and Habermas's versions of pragmatism is crucial; both versions point to the problem I am trying to elucidate. To Rortian critiques, Habermas contends,

⁸² Jürgen Habermas and Maeve Cooke, *On the pragmatics of communication* (Oxford: Polity, 2003), 360

⁸³ Rorty *Op. Cit.*, 12

⁸⁴ *Ibid.*, 60

pragmatically, that the debate is not epistemological, as Rorty would have it. Habermas thinks that the connection between justification and truth “explains why we may, in light of the evidence available to us, raise an unconditional truth claim that aims beyond what is justified” because he is not drawing a distinction “between being and appearance”, and this “is not an epistemological debate.” Instead, the tension between justification and truth is a pragmatic one, because “in everyday practices, we cannot use language without acting, and thus, the connection between justification and truth cannot be reinstated simply as the old inner/outer problem of Enlightenment epistemology.” Instead:

What is at stake is not the correct representation of reality but everyday practices that must not fall apart. The contextualist unease betrays a worry about the smooth functioning of language games and practices. Reaching understanding cannot function unless the participants refer to a single objective world, thereby stabilizing the intersubjectively shared public space with which everything that is merely subjective can be contrasted.⁸⁵

Rorty contends that if this is all that’s at stake, it is not necessary that there be an internal connection between truth and justification, and furthermore we do not need the notion of a transcultural “better argument” nor that of universal validity. We do not need to presuppose context-independent truth as an aim. Following Rorty, we are always forced to ask: what context and how is it delimited? Truth for what purpose? For him, to “describe someone as having come to a certain conclusion for bad reasons is simply to say that the reasons which convinced her would not convince us,” because “our criteria for betterness of argument are relative to the range of arguments at our disposal, just as our criterion for betterness of tool are relative to the technology at our disposal.”⁸⁶ In turn, for Habermas, the practical relationship between

⁸⁵*Ibid*, 41

⁸⁶ *Ibid.*, 59

justification and truth lends credence to the idealization of justification as a way to transcend the localism and particularity of a given argument.

As social practices, idealized justification procedures are transcendent because they can overcome the audience, time, and place where justification is provided and be seen as valid in ever-larger intersubjective audiences, times, and places, to the point of being a basis for laying claim to the ideal of universal validity in a *constitutive* way, as embedded and accepted in the presupposition of discourse once we enter into an understanding-driven conversation. In the context of communication, these idealized procedures transcend the world as such as it is given in all possible experience.

This is not the place to resolve the complex debate between Rorty and Habermas, but I want to note that their disagreement clearly shows that Habermas' account of intersubjectivity is materially different from other possible accounts, precisely in the way it deals with metaphysical commitments. Rorty shows us that by attempting to salvage universal validity and the connection between justification and truth, Habermas is trying also to save the possibility of saying that the argument in favor of liberal democracy is more rational, intrinsically better, than the argument in favor of Nazism or a caste system or religious fundamentalism. In other words, Habermas needs to argue that one can side with liberal democracy while remaining neutral, if we grant to the neutral stance a presupposition of certain universal rules of rationality, proper to the very fabric of human beings. And this is exactly what Rorty disputes:

We should admit that we have no neutral ground to stand on when we defend such politics against its opponents. If we do not admit this, I think we can rightly be accused of attempting to smuggle our own social practices into the definition of something universal and ineluctable, because presupposed by the practices of any and every language-user. It

would be franker, and therefore better, to say that democratic politics can no more appeal to such presuppositions than can antidemocratic politics, but is none the worse for that⁸⁷

To the question “do our democratic and liberal principles define just *one* possible political language game among others?” Rorty answers an “unqualified yes,” whereas Habermas wants to defend that even if its expression is phrased in a given language-game, it also expresses the very presuppositions of rational discourse, supposedly shared by all human beings in every culture. This need for a neutral perspective from which to defend the superiority and universality of the liberal ethos is precisely what requires certain metaphysical commitments.

6. Metaphysical empathy (Michael Kelly and Seyla Benhabib)

The fourth claim from the quote I have been unpacking throughout this chapter shows the final step in this process of committing Habermas’s position to a metaphysical perspective.

(4) It is easy to see why the discourse principle requires this kind of practice for the justification of norms and value decisions: whether norms and values could find the rationally motivated assent of all those affected can be judged only from the intersubjectively enlarged perspective of the first-person plural.⁸⁸

This claim relates the practice of discourse we have just seen – the processes of justification, which is ultimately the question of legitimacy and that of moral choice – to the principle of universality. But instead of doing it by means of universal a priori (i.e., metaphysical) maxims – Habermas attempts to “naturalize” it as rational unanimity, that is, as the “rationally motivated

⁸⁷ *Ibid.*, 2

⁸⁸Jürgen Habermas, *Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy*, (Cambridge, MA: The MIT Press, 1996), 228

assent of all those affected.” Yet, since literal assent of all those affected is not what Habermas is proposing here, this claim goes beyond that, and asserts that a conceptually enlarged perspective, able to go beyond the subjective, individual view, is the “only” way to judge its potential universality. This is what he calls the “we perspective” or the first person plural. How the genesis of such a perspective can be possible is what matters, and I propose to show that this is where his metaphysical commitments ultimately lie.

Consider this “we-perspective”: it is neither a collective consciousness, nor a *Zeitgeist*. It is a perspective enlarged by means of intersubjectivity. At a first glance, this formulations seems to mean that agreement and consensus enlarge our subjective view. But it is important to note that Habermas is not speaking here of a conclusion reached intersubjectively. It is not a consensus built from shared interests and compromises, as Walzer showed (above). Rather, Habermas claims that intersubjectivity enlarges our perspective – it actually creates a perspective – and that “only” from such perspective can we *judge* “whether norms and values could find the rationally motivated assent of all those affected.”⁸⁹

Recall that this long quote from *Between Facts and Norms* that we have been unpacking through this chapter preceded the one that Rawls used in his critique of Habermas’s position as a “metaphysical” project. There Habermas speaks of this we-perspective as a “universalized role taking,” by means of which we see the other’s perspective. Yet, “the *Other*” is an abstraction: one is supposed to take the role of *any* other. This is what intersubjective universality means. The only way this role-taking makes sense is if we abstract from this other any concrete features – interests,

⁸⁹ A sidenote may be warranted here: intersubjectivity is not, properly speaking, a thing, but rather a substantivized adjective and hence a concept. On the other hand, “communication,” which is not just a substantivized adjective but a kind of thing, can be the agent of the action. I think part of the problem in Habermas account can be seen as the substantivation of a quality, the reification of intersubjectivity, so as to make it carry the metaphysical weight that he rejects in principle.

situations, preferences, desires – and we connect with her through our common nature. Indeed, a shared human nature, conceptually available to our powers of reflection, is what allows us to try to see from the perspective of the other in a way that transcends the limits of our particular horizon. The metaphysics emerges in the notion of nature here, and what nature has to be in order to justify those claims of universality.

For Habermas, this practice of rational argumentation, with all its discursive presuppositions – profoundly moral, as we have seen – “distinguishes itself socio-ontologically...by a complete reversibility of participant perspectives.” It can be called a social ontology because it assumes the presence of a social being *qua* being, purified of all particularity. More importantly, the entire view redounds almost point by point to the Plato’s fundamental distinction between being and becoming and, not unlike Plato, it separates being from becoming as a privileged mode of access to the universal perspective. Or in Habermas’s words: “it unleashes the higher level intersubjectivity of the deliberating collective.”⁹⁰ Such *metaphysical* empathy is only possible if we abstract from all the particular qualities of a situated being, which politically means detaching oneself from all interests; and we leave only the very thin residue of a common human nature that we understand, a priori, as transcultural, decontextualized, and universal.

Michael Kelly and Seyla Benhabib have shown how problematic this view is. In her critique of Habermas, Benhabib repurposes the Hegelian criticism of Kantian universalism, which she thinks Habermas fails to take into account sufficiently. Hegel criticizes Kant’s universalism for being inconsistent at best and empty at worst.⁹¹ “Out of the pure form of the moral law alone,

⁹⁰ *Ibid.*, 378.

⁹¹ Seyla Benhabib, “In the Shadow of Aristotle and Hegel: Communicative Ethics and Current Controversies in Practical Philosophy” in KELLY, Michael Kelly (Ed.) *Hermeneutics and Critical Theory in Ethics and Politics*, (Cambridge: MIT Press, 1990), 4

no concrete maxims of action can follow and, if they do, it is because other unidentified premises have been smuggled into the argument.”⁹² This critique is the basis for the ongoing debate between historicism and universalism. Benhabib, however, sees this opposition as no longer compelling; she thinks that there are ways to reconcile both positions, that is, discourse ethics with neo-Aristotelian and neo-Hegelian communitarianism, by means of a “historically self-conscious universalism.”⁹³ In order to accomplish this, Benhabib argues that Habermas should drop his principle of universality for the same reasons that Hegel thought that Kant’s was dialectically unstable: it is either trivial or inconsistent. Habermas’ demand for universalizability, however, is different from Kant’s. Kant’s principle aspires to transcend the subject from the subject itself, whereas Habermas’ principle is *intersubjective*: “Instead of thinking of universalizability as a test of *noncontradiction*, we think of universalizability as a test of *communicative agreement*.”⁹⁴ Or in Habermas’s own words, from the block quote above, “the rationally motivated assent of all those affected.” But since, as Walzer shows – and Benhabib accepts – theorists can design conversations to guarantee certain outcomes, but we still face the “dismal choice”: “either models of practical discourse...are defined so minimally as to be trivial in their implications, or there are more controversial substantive premises guiding their design which do not belong among the minimal conditions defining the argumentation situation, in which case they are inconsistent.”⁹⁵ The way out of this dilemma, according to her, is to see that there is already universal content in Habermas’s pragmatic presuppositions. What she calls *the principle of universal respect*: “the right of all beings capable of speech and action to be participants in the moral

⁹² *Ibid.*

⁹³ *Ibid.*, 8

⁹⁴ *Ibid.*, 6

⁹⁵ *Ibid.*

conversation” and what she calls *the principle of egalitarian reciprocity*: “each has the same symmetrical rights to various speech acts.”⁹⁶ By preserving this universalism, while dropping the need for universal agreement, “the emphasis is less on *rational agreement*, and more on sustaining those normative practices and moral relationships within which reasoned agreement as *a way of life* can flourish and continue.”⁹⁷ The reason why this shift resolves the dismal dilemma is that by dropping the need for universal agreement we avoid the emptiness of the universability test; in Agnes Heller’s words (quoted by Benhabib): “we cannot obtain any positive [moral] guidance from the Habermasian version of the categorical imperative.”⁹⁸ But, at the same time, by reconstructing Habermas’s pragmatic presuppositions as universalist moral demands, we obtain a “universalist moral theory that is self-conscious about the historical horizon of modernity within which it is situated.”⁹⁹ This is “historically self-conscious universalism,” in other words, because it is what we post-Kantian and post-Hegelian moderns mean when we say the word “universal” in the context of moral conversation: “The principles of universal respect and egalitarian reciprocity are our philosophical clarification of the constituents of the moral point of view from within the normative hermeneutic horizon of modernity.”¹⁰⁰ Benhabib’s critique of Habermas’s universalism shows that behind the constraint of universal acceptance there may be some transcendental projection of what is morally acceptable: the attempt at introducing situated material content into a formal-procedural principle. But more importantly, it shows that the

⁹⁶*Ibid.*, 7

⁹⁷ *Ibid.*, 13

⁹⁸ *Ibid.*, 11

⁹⁹ *Ibid.*, 12

¹⁰⁰ *Ibid.*, 8

procedure itself “involves the utopian projection of a way of life as well.”¹⁰¹ This way of life is predicated upon a particular understanding of human nature. How this human nature is described has to do with the presence of the formal principle of universalizability.

Following McIntyre, Kelly shows that Habermas’s “universalizability,” that is, the power that his procedural obtains in turning its moral content into something universally defensible, is just the outcome of a particular historical tradition, namely, Western modernity. We can only understand universality from this tradition, and our political and moral justifications along the lines of universality are only acceptable in the context of that tradition. Kelly argues convincingly that Habermas would concede that the principle of universalization is “unintelligible unless its conceptual links to the Kantian moral tradition are clear; but he would object that it does not follow from this concession that the principle itself cannot be justified independently of this same tradition.”¹⁰² Or, in other words, intelligibility and justification are two separate things, one may be tradition bound, the other, by means of rational procedure, may be independent. This “independent” here is crucial because it means ‘neutral.’

Practical rationality seeks to establish “the conditions under which moral principle – for example, about justice – can be chosen impartially and then used to resolve conflicts – again, about justice – impartially,” and for this purpose, traditions must not be the final “arbiter.”¹⁰³ Kelly contends that even if Habermas were capable of justifying his procedural, provisional universalism as independent from the Kantian tradition, he would still remain dependent upon the tradition of

¹⁰¹ *Ibid.*, 13

¹⁰² Kelly, “MacIntyre, Habermas, and Philosophical Ethics,” in Michael Kelly (Ed.) *Hermeneutics and Critical Theory in Ethics and Politics*, (Cambridge: MIT Press, 1990), 79

¹⁰³ *Ibid.*, 79

liberalism itself, with its principles of equality, freedom and justice. And this, in a way similar to what Rawls criticizes in Habermas, implies, all in all, a sort of *Sittlichkeit*, à la Hegel. Habermas tries to “incorporate” this point in discourse ethics:

by conceiving of the logic of historical development underlying the project of the Enlightenment as a learning process which culminates in the conditions of reciprocity and symmetry and on which we cannot willingly turn our backs.¹⁰⁴

This way, Habermas might introduce both the idea of provisionality, necessary if we are to eschew all presuppositions about universal, transcendental, and eternal truths, as well as the political values belonging to his tradition, without renouncing the possibility of a neutral justification that might achieve universal consensus through pragmatic rationality. Historically, we have learned to appreciate these political values to the extent that they have become unavoidable, but once we have reached this moment, we can elevate it conceptually to the heights of universality – so he claims.

The critique of Habermas’s universalism may be illuminated if we think of universality not as a grand accomplishment, but rather as a natural feature of language. Simply put, there is no universality in nature. Even if we have something like natural kinds à la Aristotle, every object in nature is a particular “this,” while language is thoroughly constituted of universals, leaving aside the special case of proper names, because words are essentially conceptual not material. Even if we take *this* particular word, it is particular as it we take it materially as merely an assemblage of graphemes or phonemes. Insofar as it signifies something beyond its graphic or phonic elements though, when it is properly speaking ‘a word’, the word is thoroughly conceptual, i.e., universal. Academics and law-makers are tasked with the problem of discovering (stable)

¹⁰⁴ *Ibid.*, 80

universals, while poets and “regular” people face the challenge of giving expression to the particular. In a certain sense then, language is intrinsically exoteric, while emotions and other pre-conceptual “particulars” in human experience are esoteric.

We may be able to track the development of civilization (in society and in the individual alike) as progress in universality, but strictly speaking it would be wrong to say that this progress goes from particularity to universality; rather it goes from lower degrees of universality to higher degrees, encompassing more and more particular content (including of course the relations between particulars). The transition from particularity to universality occurs, let's suppose, with the mysterious advent of language in human history. But back to growth in universality as concepts encompassing more content: this brings us back to the definition of metaphysics in terms of a “comprehensive view”. The problem with metaphysics as a project is that it seeks to grasp the whole, a noble effort, but it routinely mistakes a certain part (however large) for the whole. Habermasian liberalism, with its strong emphasis on universality, distorts the whole (the set of all particulars) by privileging abstract universals over concrete particulars. The interesting thing is to note how Habermas, without using these Hegelian terms, tries to discover a concrete universal, that is not in the Platonic heavens, but here on Hobbesian earth, so to speak.

Kelly, however, thinks that his appeal to a learning process does not capture enough of the historical depth presupposed by the principle of universalization, “because Habermas reduces the complex learning process to the competence to utilize the universalization principle” or, in Benhabib’s words, to the ability to engage in the ideal role taking that results in the ‘generalized

other’.”¹⁰⁵ The learning process, in other words, is reduced to the principle of universalizability, which in turn, presupposes the learning process. Moreover, the principle itself is dependent upon the possibility of a metaphysical empathy, the idealized role taking.

Recall that “whether norms and values could find the rationally motivated assent of all those affected can be judged only from the intersubjectively enlarged perspective of the first-person plural.” This enlarged-perspective, with Hegelian and Arendtian undertones,¹⁰⁶ refers to the reversibility of perspectives from oneself to an other. But this other, in the context of this *naturalized* categorical imperative, is presented as a formal other, as an empty other, where one can project the content of what is expected of her or him to accept in a given ideal discursive argumentation. The erasing of the concrete other is what allows this we-perspective to be thought, and then re-concretize the other, now with the material content that Habermas has assumed in the first place. In his debate with Rawls, Habermas puts it like this:

Under the pragmatic presuppositions of an inclusive and non-coercive rational discourse among free and equal participants, everyone is required to take the perspective of everyone else, and project herself into the understandings of self and world of all others; from this interlocking of perspectives there emerges an ideally extended we-perspective from which all can test in common whether they wish to make a controversial norm the basis of their shared practice; and this should include mutual criticism of the appropriateness of the languages in terms of which situations and needs are interpreted.

¹⁰⁵ *Ibid.*, 80

¹⁰⁶ *Ibid.*, 83

In the course of *successively* undertaken abstractions,¹⁰⁷ the core of generalizable interests can then emerge step by step¹⁰⁸

Both Kelly and Benhabib argue that one of the main problems that arises in this role-taking is that the ‘concrete other’ is excluded and abstracted out of the process, and in truth, if we are to rely on the acknowledgement of the other’s perspective, this should not be so. To include the concrete other implies situating this other historically, and being aware of their identity¹⁰⁹ in a specific way, specially if we are to expect that they are part of that ‘learning process’ that leads us to expect them to be able to apply the universalization principle as part of a transcendently weak, or ‘unavoidable’, set of pragmatic presuppositions of speech action. Benhabib believes, and Kelly concurs, that by excluding the concrete other, Habermas is also excluding the generalized other; you can’t have one without the other. “It is with a concrete other that we engage in the role taking that is supposed to culminate in the criteria of reversibility and universalizability said to be constituents of the moral point of view.”¹¹⁰ The exclusion of the concrete other in favor of the generalized other, I propose, clarifies why the very possibility of

¹⁰⁷ This development, that of increasing abstraction, is what I was talking about before. I’ll just point out again that language is essentially abstract. A consequence of this is that the “I”, the ego, the locus of autonomy, etc., is no more particular than it is universal. This comment doesn’t really address the block quote from Habermas, but it’s worth noting that “discourse” is already the space of universal abstractions, whereas empathetic relations or emotion-laden face-to-face encounters involve particular stuff that is not necessarily expressed in or captured by language. For political consequences of this, imagine an example: We could take the case of a mother of a black African-American kid killed by police and she’s interviewed by reporters and expresses herself emotionally. There is certainly something of rational discourse in her words, but it’s the concrete, non-abstract, emotive force of her words that stirs empathy within us. In short, what I’m trying to say here is that there is a certain tension between Habermas’ ideal abstraction and the role of intersubjectivity in his theory, given that he wants to think it along lines of empathy *Einführung* (I am indebted to Matt Hackett for this point).

¹⁰⁸ Habermas, Jürgen, *Reconciliation Through the Public Use of Reason: Remarks on John Rawls’s Political Liberalism*, *The Journal of Philosophy*, Vol 92, No. 3 (Mar., 1995), pp 109-131, p 117-118

¹⁰⁹ Identity, as we usually use the term, involves various forms of universality (predicate terms), such as race, nationality, gender, religion, etc. Note that none of these terms express singularity, the concreteness of the other person. The identity, or concreteness of the other, consists not in his physical (spatio-temporal) existence, which is true is singularity, but in the idiosyncratic assemblage of predicate terms that apply to the individual.

¹¹⁰ Kelly, *Op. Cit.*, 80.

taking the perspective of "everyone else" – this is the core of this universalism – is the possibility of taking the metaphysical, all-encompassing, perspective.

The "ideally extended we-perspective" is the old God's view which is now secularized in its aims, but not in its presuppositions, or to paraphrase Habermas's own words: it is a "social ontology", pregnant with moral and metaphysical content, decidedly committed to not only a particular conception of human nature, but to the view that our common fundamental nature, expressed through our rational powers, is what gives us the ability to go beyond our particularity and "project" ourselves "into the understandings of self and world of all others." This interlocking of perspectives, creating an increasingly enlarged perspective by which to decide upon moral controversies, is what ultimately will allow Habermas to claim that in his ideally constructed conversations the "better argument" wins over without coercion. Or in other words: that in the case of conflicting perspectives, this conversation can play the role of decider due to its assumed neutrality.

7. An integrative perspective

The fifth and last claim of the long quote I have been commenting on summarizes what is the function of the neutral perspective, why it is needed in the overarching schema, and how it conceals its metaphysical content:

- (5) This perspective *integrates* the perspectives of each participant's worldview and self-understanding in a manner that is neither coercive nor distorting. (My italics)

Neutralists are after a standpoint that can integrate all world-views. If they could get it, they would be able to observe the quarrels between different world-views from the outside, from a

position that has no opposition. Of course, for this to be possible, one must be able to see the world as a whole, from the outside, which means that one must endorse the idea that there exist at least two fundamental perspectives: one from the inside and one from the outside. The inside perspective is possessed of a world-view (in the sense of an ideology), it is partisan, and it projects one's self-understanding onto everything external. The outside perspective lacks a world-view, is non-partisan, and is detached from all selfhood. The outside perspective is not inter-subjective, and it can explain intersubjectivity as the space between subjects, or selves, because it can see this space not from the loaded point of view of oneself, but from the self-less view from nowhere. Because it sees the intersubjective space from without, it can see the content that makes this space possible, and convert it into an abstract set of *nomoi*. These *nomoi* are universal because they express the sublimated content of our rationality – our human nature. And we see this human nature as something universal that channels political content because we can see it from the outermost perspective that integrates without coercion or distortion. Once this is set up, the best argument can win, but in truth it is unnecessary to have the debate anymore, because the bestness of the argument has already, implicitly, been deduced a priori, before the intersubjective process of giving and receiving reasons.

neutrality is not a way to avoid the smuggling in of metaphysical content in political and social procedures. On the contrary, it is the precondition for metaphysics. We leave the contingent, situated context of our political quarrels, we are able to settle them universally, we obtain from our dialogical action some *nomoi*, we offer a comprehensive view of human nature *qua* being,

because there is a neutral perspective. Without it, everything Habermas and other neutralists assert would be just the social expression of a temporary context.

Chapter IV

Reasonableness's Metaphysical Content in Rawls

*“Que se consiga el efecto, sin que se note el cuidado”
attributed to Don José Rodrigo Villalpando ¹*

What distinguishes the Kantian form of constructivism is essentially this: it specifies a particular conception of the person as an element in a reasonable procedure of construction, the outcome of which determines the content of the first principles of justice.²

This is the first sentence of John Rawls's lectures on *Kantian Constructivism in Moral Theory*. The main aim of this chapter is to disentangle to what extent the conception of the person that lies behind Rawls's conception of justice rests on metaphysical assumptions. The notions of procedure and construction that appear in this passage will be central in my analysis, but the key aspect will be the qualifier that precedes them: reasonable. In the role played by reasonability it will be seen how the “content of the first principles of justice” is metaphysically loaded and how it fosters a particular anthropology—a conception of the person—consistent with that content.

1. From comprehensive to political

Rawls's lectures on Kantian constructivism belong to the second half of his oeuvre, *A Theory of Justice* being the epicenter of the first half. To claim that Rawls's original theory of justice is a

¹ *Fiscal del Consejo de Castilla*, in his *Instrucción Secreta* sent to the *Corregidores del Principado de Cataluña* in January 29th 1716: “That the effects be achieved without the effort being noticed”

² John Rawls, “Kantian Constructivism in Moral Theory” *The Journal of Philosophy*, Vol. 77, No. 9, pp. 515-572 (Sep. 9, 1980), 516

comprehensive doctrine of the good is an uncontroversial view: he himself says so in the introduction to his subsequent works. Indeed, the second half of his *œuvre* is an explicit attempt at solving the problems that *A Theory of Justice* created in presenting a comprehensive moral theory as a tool to organize political affairs. According to Rawls, the practical problem with that particular conception of justice —namely, ‘justice as fairness’— is that it cannot ensure stability. The practical instability of the first formulation of ‘justice as fairness’ is a function of an “unrealistic idea of a well-ordered society as it appears in *Theory [of Justice]*” because citizens are supposed to endorse it “on the basis of what I now call a comprehensive philosophical doctrine.” One of the main aims of *A Theory of Justice* was to oppose utilitarianism with a bold conception of justice, and “it is clear...that the text regards justice as fairness and utilitarianism as comprehensive, or partially comprehensive, doctrines.”³

A more realistic idea of justice, according to Rawls, demands that we secure stability in view of the “fact of a plurality of reasonable but incompatible comprehensive doctrines —the fact of reasonable pluralism.” The problem is solved by presenting ‘justice as fairness’ “from the outset as a political conception of justice,” instead of as a comprehensive moral view.⁴ However, this change “forces many other changes and requires a family of further ideas”, namely: “overlapping consensus of comprehensive, religious, philosophical, and moral doctrines,” “public basis of justification and public reason,” and “certain facts of commonsense political sociology” like “the burdens of judgement.”⁵

³ John Rawls, *introduction to Political Liberalism* (New York: Columbia University Press, 1993), *xvi*. “Partially comprehensive” must be a metaphorical paradox.

⁴ *Ibid.*, *xvii*

⁵ John Rawls, *introduction to Justice as Fairness. A restatement*, (Cambridge: The Belknap Press of Harvard University Press, 2001), *xvii*

For over 25 years, Rawls developed a revision of *A Theory of Justice* to accomplish this shift from comprehensive moral view to a political conception, and from this effort has sprung an abundance of texts, some of them published and republished in journals and books with constant revisions. From “Basic Structure as a Subject”⁶ (1977) which addresses the range of the application of his idea of impartiality as justification, as we have seen in chapter I, to a more complete statement of his position in the lectures on “Kantian Constructivism in Moral Theory,”⁷ (1980), which address the distinction between so-called moral and political constructivism, followed by a series of papers on the new ideas and positions that were necessitated by the attempt at distancing his theory from a comprehensive view: “Justice as Fairness: Political not Metaphysical”⁸ (1985), which, as we have seen in chapter II, concerns the “method of avoidance” that he uses for theoretical justification, while “The Idea of an Overlapping Consensus”⁹ (1987), “The Priority of Right and Ideas of the Good”¹⁰ (1988), and “The Domain of the Political and Overlapping Consensus”¹¹ (1989), all address central components of the general argument, as we will see below. All of these papers, reviewed and refined, together with his “Reply to Habermas,” which, as we have seen in chapter III, continue to deal with the aspect of his doctrine that seems to harbor metaphysical assumptions, and his “Public Reason Revisited”—one of the key text to discuss neutrality—are all published in the volume *Political Liberalism*¹² (1993, 1996, and 2005 —the last one posthumous—, with subsequent

⁶ *American Philosophical Quarterly* Vol. 14, No. 2 (Apr., 1977), pp. 159-165

⁷ *The Journal of Philosophy*, Vol. 77, No. 9, (Sep. 9, 1980), pp. 515-572

⁸ *Philosophy and Public Affairs*, Vol. 14, No. 3. pp. 223-251 (Summer, 1985)

⁹ *Oxford Journal of Legal Studies* Vol. 7, No. 1 (Spring, 1987), pp. 1-25

¹⁰ *Philosophy & Public Affairs* Vol 17, No. 4 (Autumn, 1988), 251-276

¹¹ *New York University Law Review* 64 (1989): 233–55

¹² (New York: Columbia University Press, 2005)

revisions and additions). “Political Liberalism” and the also posthumous “Justice as Fairness: A restatement,”¹³ form the final product of the purification of the original edition of *A Theory of Justice* (1971).¹⁴ Also, there is a canonical revised edition of *A Theory of Justice* which incorporates the first changes made by Rawls in 1975 when he was preparing the text for its German translation, addressing the first criticisms that eventually led to the other revisions. Incidentally, it became the standard for all other translations, but it was not published in this form in English until 1999¹⁵.

This effort is a testament to Rawls’s dedication to making his thought consistent throughout and to his openness to criticism and objections, however it also raises a question. Which version of his argument are we to take as the most relevant? In 1992, Rawls wrote: “I reached a clear understanding of political liberalism —or so I think— only in the past few years.¹⁶” On the other hand, the constant tweakings and the never-finished character of Rawls theory perhaps provides some support for my argument: one of the main reasons for this constant revision is the problem of nailing down the relationship between metaphysics and his theory of justice. The never-ending changes can be traced back to the fundamental yet difficult requirement that political liberalism “has to be impartial between points of view of reasonable comprehensive doctrines”¹⁷. The problem of instability follows from the problem of imposing a comprehensive doctrine like his original formulation of ‘justice as fairness’ because reaching an agreement on a comprehensive doctrine is as difficult as it is to find consensus in religious matters. The solution to

¹³ (Cambridge: The Belknap Press of Harvard University Press, 2001)

¹⁴ “Original Edition” (Cambridge: The Belknap Press of Harvard University Press, 1971)

¹⁵ “Revised Edition” (Cambridge: The Belknap Press of Harvard University Press, 1999)

¹⁶ *Political Liberalism, Introduction*, p xxx

¹⁷ *Political Liberalism*, introduction, p xix

this problem is to design an idea of justice that may be understood in “a metaphysically, epistemologically, and morally neutral way.”¹⁸ In other words: the key difference between the alpha and omega of his thinking, (namely, his two majors works, *A Theory of Justice* and *Political Liberalism*), is neutrality.

2. Exclusion and Inclusion

In the introduction to the last edition of *Political Liberalism*, Rawls explains that his first attempts at introducing his new political conception of justice and the idea of an overlapping consensus were “misleading” and “led to objections” that made him realize that in order to remove the problem he would have to develop some “missing pieces”. Among these “essential” pieces, the three “central ones” are:

1. the idea of justice as fairness as a freestanding view (...)
2. the distinction between simple pluralism and reasonable pluralism, together with the idea of a reasonable comprehensive doctrine; and
3. a fuller account of the reasonable and the rational worked into the conception of the political (as opposed to the moral) constructivism, so as to bring out the bases of the principles of right and justice in practical reason.¹⁹

Let us see them one by one.

2.1. Freestandingness as the Exclusion of Metaphysics

The first of these pieces is the most explicitly related to neutrality. Freestandingness stems from the hunch that the problem of instability that arises specifically out of *A Theory of Justice* might in truth be a more general problem. It may be the case that a comprehensive moral or political view

¹⁸ I am following here Samuel Freeman, Introduction to *The Cambridge Companion to Rawls*, Samuel Freeman Ed., (Cambridge: Cambridge University Press, 2003), especially p 33-34

¹⁹ John Rawls, Introduction to *Political Liberalism*, (New York: Columbia University Press, 1993), xxx

that can gain the assent of all (or the majority of) citizens of a democratic regime simply may not exist, even if they behave in a rational and reasonable way, and thus enjoy freedom and equality, in an ideally liberal way. If that were the case, then the only way to ensure stability would be to apply coercion against reasonable and rational people, which would then involve a breach of both the ideals of freedom and equality proper to liberalism. As we have seen in chapter II, Rawls's strategy in this respect consists of 'bracketing' or avoiding "philosophical questions of the original source of moral principle, how we ultimately come to know them, and even the question of their ultimate truth."²⁰ The method of avoidance does not entirely preclude the possibility that the political conception of justice that ultimately outcrops is grounded in some deep comprehensive view; it merely renounces the path of justification that would lead to posing and answering of the question concerning grounding.²¹ This is done in order to avoid the problems that it would create, but also in order to allow for different comprehensive doctrines, (whatever their nature: religious or otherwise), to approach such a political conception from their own grounding and accept it as warranted.

However, Rawls's aim is not to build a conception of justice that can work as a synthesis of existing comprehensive views, nor as a sort of negotiated compromise between these views. The process of justifying 'justice as fairness' as an exclusively political conception involves two steps. In the first, it relies on ideas implicit in a democratic culture, it applies the method of avoidance, and it tries to construct its principles from an idea of the person as free and equal to all others (in the "political" sense). This means that a person, in this construct, is a citizen, and as such enjoys the freedom and equality presupposed in the idea of moral agency, but it does not require, or this

²⁰Freeman, *Op. Cit.*, 33-34

²¹ It becomes difficult to understand in which sense this is a method, since it entails not doing something explicitly.

is what Rawls claims, any citizen to actually endorse an idea of moral autonomy that extends beyond the political sphere.²² The citizen and the person are distinguished in the same way that right and the Good are distinguished, that is, in the same way that a political conception of justice is distinguished from a comprehensive view, and in the same way that political and metaphysical spheres are separated. The citizen is understood as free and equal in the political sphere only, so that the person can enjoy the freedom to believe and live according to the idea that people might not be autonomous or ontologically equal.

In this first step, even though Rawls has renounced the idea of congruence, that is, an important idea in *A Theory of Justice* according to which the content of ‘justice as fairness’ is congruent with one’s idea of the Good (because that would amount to saying that there is a comprehensive consistency underlying the system), there still is the belief that it is rationally sound for people to endorse this idea of citizenship, because accepting free and equal agency in the political sphere allows them to pursue their own idea of the good in their private or communal, non-political sphere. In this first step of justification we see how the political sphere is the sphere in which neutrality operates in the sense that it excludes metaphysical views from it, and it offers an idea of the political separated from metaphysical or epistemological debates that shape the deep sense in which we say that a society is plural. It is already apparent in this first step that neutrality plays the role of an outermost view that distinguishes between warranted and unwarranted uses of controversial ideas, such as equality and liberty, around which all that it is at stake in the political domain is organized. The only meaning of person (or of equality and liberty) that is warranted in the idea of justice proper to the political domain is the one that neutrality allows; all the rest are excluded. The content of these warranted ideas will be discussed further below.

²² The topic of autonomy will be dealt with separately in the following chapter.

2.b Overlapping Consensus as Moral Inclusion

However, this is not enough to secure stability, according to Rawls. So a second step in the process of justification is required, and it involves what he calls ‘overlapping consensus.’ An overlapping consensus is predicated upon the possibility that reasonable citizens endorse the political conception of justice on the basis of “nonpublic reasons stemming from their comprehensive views.”²³ The political conception of justice that is at the basis of the system should be compatible with each reasonable person’s self-understanding of moral agency and, with that, of the other characteristics of their comprehensive view. When all reasonable comprehensive views satisfy this requirement, there is an overlapping consensus.

In this second step of justification, we see how the political sphere (with the conception of justice that articulates it) is neutral in the sense that it can include a plurality of comprehensive views. If in the first step it excluded metaphysical positions by being only political, in the second step it reintegrates them by offering an account of justice that can be compatible with a diverse set of metaphysical groundings, and thereby it reintroduces the moral aspect of the political conception of justice in the overlapping content that is shared by different reasonable perspectives. An overlapping consensus assures stability “for the right reasons” because it “is not a mere *modus vivendi* but is moral in both its object and its content.”²⁴ In what sense does Rawls say that this can be moral but not comprehensive?

First, the object of the consensus, the political conception of justice, is itself a moral conception. And second, it is affirmed on moral grounds, that is, it includes conceptions of society and of citizens as persons, as well as principles of justice, and an account of the political virtues through which those principles are

²³ Freeman, *Op. Cit.*, p 37

²⁴ Rawls, *Political Liberalism*, 126

embodied in human character and expressed in human life. An overlapping consensus, therefore, is not merely a consensus on accepting certain authorities, or on complying with certain institutional arrangements, founded on a convergence of self- or group interests. All those who affirm the political conception start from within their own comprehensive view and draw on the religious, philosophical, and moral grounds it provides.”²⁵

It is worth noting that this argument is developed by Rawls as a response to communitarian criticisms of liberalism that see it as a mere *modus vivendi*, and not as the basis for a political *community*.²⁶ This is why Rawls stresses here its most moral aspect. Even if it seems contradictory to claim that ‘justice as fairness’ is a moral conception, as if there were no separation of the right from the Good, in Rawls’s view it is not contradictory because this moral conception is construed politically. Political constructivism is a social endeavor that manages to develop a political conception of justice, which is a moral conception that is not construed following Kantian moral constructivism – *a priori*, individually, in a purely rational way, etc. – but politically, following the rules of a specific social procedure, as represented in the ‘original position’.

Equally, the function of this moral conception —according to Rawls— is not to determine what one’s morality or comprehensive view should be. Rather, its function is to arbitrate different moral views anchored in different comprehensive views in a way that is more stable and deep than a mere *modus vivendi*. It not only secures a set of basic institutions, and it does so without some form of compromise, it also secures the legitimacy of a given set of institutions and practices because it finds a political conception of justice that responds to a moral understanding of the social relationships between reasonable persons. Its morality is, precisely, the fact that it treats reasonably diverging comprehensive moral and philosophical views equally. In other words,

²⁵ Rawls, *Political Liberalism*, 146

²⁶ This seems *prima facie* less metaphysically compromised than Habermas. Rawls discovers a common assumption in positions and he builds it into his theory. This allows him to make an immanent critique of opposing positions. Habermas is similar but the intersubjective process in Habermas is less intuitive and more questionable than Rawls’ principle here.

the reciprocity and equality —the ‘fairness’— that form the constraints of the original position (as we will see below) constitute not only a practical solution to instability but also a moral content that is carried over into the political conception of justice that crops out of the original position. The underlying neutrality of it all is already unveiled as a moral position, even if politically construed and thus, ‘freestanding.’

Moreover, it also provides a set of incentives for other moral positions to evolve accordingly. As Samuel Freeman claims, “the key assumption underlying Rawls’s conjecture here is that,

because of the reciprocity of the principles of justice, over time the public political culture and the benefits of a well-ordered society will lead all reasonable persons holding reasonable comprehensive views to accommodate their own comprehensive accounts of justice and morality to the public view. (...) Reasonable comprehensive doctrines in response to the benefits of justice they all enjoy in a well-ordered society, should evolve there so as to incorporate a liberal political conception into their values and principles.”²⁷

The inclusion-exclusion dynamic that freestandingness and overlapping consensus create is meant to offer a system of incentives for comprehensive views to evolve toward becoming more congruent with the political conception of justice of liberalism. This weak form of congruence is what in the argument is referred to over and again as “reasonable”. Recall the three central missing pieces that Rawls had to develop in *Political Liberalism* to make ‘justice as fairness’ more stable, the first being freestandingness, and the other two (as quoted above):

2. the distinction between simple pluralism and reasonable pluralism, together with the idea of a reasonable comprehensive doctrine; and
3. a fuller account of the reasonable and the rational worked into the conception of the political (as opposed to the moral) constructivism, so as to bring out the bases of the principles of right and justice in practical reason.

²⁷ Freeman, *Op. Cit.*, 37

The distinction between simple pluralism and reasonable pluralism shows that reasonability is a central piece because it signals the limit of what is acceptable within the range of plurality that neutrality arbitrates. A political conception of justice that has to organize a well-ordered society cannot simply be neutral among all ideas of the good, because that would imply being neutral between, say, rapists and non-rapists, or between Nazis and non-Nazis. However, the very point that neutrality tries to advance forbids liberals to argue against the inclusion of such-and-such doctrine by virtue of its intrinsic good, since there must be another, more neutral, way of doing so. Otherwise, they would be advancing a comprehensive doctrine that would oppress other views, views unlike pro-rape or Nazism. In this case, a comprehensive doctrine would be as oppressive as any other conception that rejects plurality. Therefore, 'justice as fairness' has to be impartial among acceptable ideas of the good, and it must be able to receive endorsement from different and conflicting perspectives, from their own groundings, as we have seen, according to an overlapping consensus. Thus, these perspectives, in order to be part of the plurality of acceptable views, must be *reasonable*.

This is why it is so important to decide which doctrines can be accepted in the game of consensus, because their grounding will be consistent with the conception of justice in the eyes of *reasonable* citizens that live by them. By making explicit the conditions on entering the space of reasonability, we make implicit a shared, acceptable grounding. Is this an implicitly metaphysical content, even if shared and overarching?

3. Reasonableness

3.1 Indetermination and self-containment

In order to answer this question in the affirmative, and to make that content explicit, we must go deeper into the notion of reasonableness. Rawls himself does not offer a clear definition of the concept. As Samuel Freeman notes, this “can be frustrating for the reader, since this concept is so crucial to Rawls’s argument.” So much so that its pervasiveness together with its “obscurity” may seem “to some” that it “explains little” and “only masks the lack of a proper argument,” to the extent that the most common objection to Rawls’s use of “reasonable” is “that the concept lacks a precise definition.”²⁸ Vague concepts allow theorists to exploit them for their purposes, whereas precise definitions leave the relevant concepts rigid and vulnerable to criticism. Definitions that are somewhat elastic and yet not overly vague make for potent/efficient and relatively invulnerable concepts.

Reasonableness lacks a sufficient definition not because Rawls simply neglects to provide one, or because he favors a Heraclitean obscurity, but rather because he is compelled by the structure of his argument to leave it vague and give it an all-purpose utility. Rawls cannot offer a detailed and comprehensive definition of reasonableness without thereby undermining his purpose and endangering his method of avoidance. This could end up dismantling the entire edifice he has built separating the political from the comprehensively moral and epistemological outlook, the right from the Good, the reasonable from the rational and the true. Along with all the other dichotomies employed in *Political Liberalism*, they serve the claim that ‘justice as fairness’ is a political conception and not a metaphysical doctrine, thereby solving the problem of stability of

²⁸ Freeman, *Op. Cit.*, 32-33

A Theory of Justice. If Rawls were to establish a clear set of criteria for reasonableness, he would have to provide an explicit account of the relationship between our rational capacities and their connection with what is morally acceptable, turning his system into a comprehensive system. The incompleteness of his philosophy in this respect is the outcome of the method of avoidance. The purpose of using reasonability as a guardian of the public sphere is precisely to avoid judging each comprehensive view according to its metaphysical merits. What reasonability asks of comprehensive views is that they accept the idea that in the political sphere one cannot appeal to comprehensive views. A reasonable person is someone who accepts a sharp distinction between the demands of the political sphere and their own moral sphere, or with their own religious views, etc. This is why “the idea of the reasonable makes an overlapping consensus of reasonable doctrines possible in ways the concept of truth may not.”²⁹

However, this does not solve the problem of how to decide which “conceptions and principles, judgements and grounds, persons and institutions,”³⁰ to which the idea of reasonableness is applied to, are reasonable. Rawls says that “the content of the reasonable is specified by the content of a reasonable political conception,” which is consistent with the entire architecture of his system because the content of the reasonable cannot be external to the independent sphere of the political as he has constructed it, and thus the content of the reasonable can only be the fact that a conception of justice adapts itself to the criteria of that sphere. And that means that the content of the reasonable, as a function of the content of each political conception as a non-metaphysical conception, is related to the features that affirm such a separation:

²⁹Rawls, *Political Liberalism*, 94

³⁰*Ibid.*, 94

The idea of the reasonable itself is given in part, again for our purposes, by the two aspects of persons' being reasonable: their willingness to propose and abide by fair terms of social cooperation among equals and their recognition of and willingness to accept the consequences of the burdens of judgement. Add to this the principles of practical reason and the conceptions of a society and person on which the political conception is based. We come to understand this idea by understanding the two aspects of the reasonableness of persons and how these enter into the procedure of construction and why.³¹

This passage shows paradigmatically that Rawls's theory is, as Habermas puts it, "impressively self-contained,"³² because it has a systematic vocation: all of its parts are interrelated. It harbors the belief that if the system can cohere in itself while avoiding the question of metaphysics, then that question need not be addressed at all. However, as this very passage expresses, the one component that is determined by the others and not by means of its own definition is reasonability. "Reasonable" is always a quality predicated of something else in the system, and everything else is the predicate of reasonability, when reasonability is the subject.

3.b The Anthropology of a Social *Ethos* and an Imperfect *Nous*

According to this passage, the idea of the reasonable has six characteristics. First, it is given (in part) in the form of a person being reasonable, which already suggests that we are speaking of a moral quality, an *ethos*. This is confirmed by the fact that the two characteristics essential to this *ethos* have to do with the will: (2) the willingness to propose and abide by fair terms of social cooperation among equals and (3) the recognition of and willingness to accept the consequences of the burdens of judgement.

A reduced form of #2 could be: being reasonable is to want to cooperate, which, in the banal sense of the word, sounds reasonable enough. But we should note that Rawl's formulation is

³¹ *Ibid.*

³² Habermas, Jürgen. *Reconciliation through the use of reason: remarks on John Rawls Political Liberalism* *The Journal of Philosophy*, Vol. 92, No. 3. (Mar., 1995), pp. 109-131p 131

actually more nuanced: “Proposing” and “abiding by” express active and passive aspects of political life. To be reasonable then is to cooperate actively in the process of creating the social contract, and then resign yourself to the outcome of this process. It is to willingly partake in the process, to cooperate in it, as Rawls sees it. Adding the qualifier “fair” to the “terms of social cooperation” softens the stance somewhat. Of course, if the terms are fair, who would not want to cooperate?³³ However, since the question of fairness/justice is what is under discussion, when Rawls says that a reasonable person is someone who willingly proposes and abides by a *fair* social contract is saying nothing unless we know what he means by fair. If he means the willingness to propose and abide terms of cooperation that are fair according to the subject that proposes and abides by them, this would turn the phrase into a sort of truism that does not solve the problem it is supposed to solve. This is an empty tautology: a reasonable person is one who is fair; where fair is synonymous with reasonable. This becomes: a reasonable person is reasonable. X is X.³⁴

Some clues regarding this fairness are seen in the fact that social cooperation among equals is requisite for these fair terms. The requirement excludes any social contract that would legitimize anti-social behavior, and the freedom to not cooperate socially, as well as any terms that do not support the assertion of an *a priori* equality among members of the social body. These might seem perfectly valid requirements for a good society on our terms, but as a point of legitimate entry into the public sphere it seems to assert a social dimension of the person that hardly satisfies the conditions of neutrality or impartiality.

³³ We could phrase this rhetorical question in a negative form: if the terms were not fair, who would want to cooperate? Of course the answer is: cooperation under unfair conditions is only appealing if non-cooperation is worse. For example, workers cooperate with bosses under unfair conditions all the time, only because it's better than not having a job.

³⁴ Note that this is different than question begging. The problem is not precisely that he assumes what he sets out to prove but that he leaves the very concept under consideration empty. I am thankful to Matt Hackett for his help in clarifying this distinction.

Of course, we know that Rawls's concept fairness, as in the locution 'justice as fairness', has to do with the position that agents in the original position occupy, positions that are fair because they are ignorant of their intrinsic inequalities (social, biological, or otherwise), as well as of the higher-order goods their particular comprehensive views ask them to pursue. Thus, to be *willing* to propose and abide by these fair terms of social cooperation among equals is hardly an empty statement when it comes to moral requirements to enter the public sphere, but (#3) to be willing to recognize and accept what Rawls calls "the burdens of judgement" is yet a step further.

First of all, because a willingness to recognize entails a mental state that not only covers the will, *voluntas*, but also the mind, *nous*; it is a willingness to know and accept as true the fact that reason is imperfect and one's perspective is always partial and incomplete. Or, what is the same: a willingness to know that modern notions of rationality have been proven too optimistic, and that in truth our judgement is limited by the "hazards involved in the correct (and conscientious) exercise of our powers of reason and judgment in the ordinary course of political life."³⁵ This, in turn, explains the turn to reasonable pluralism. Rawls's list³⁶ of the burdens of judgement includes the following items:

(a) empirical and scientific evidence is often complex and conflicting; (b) we may reasonably disagree about the relative weight of different considerations; (c) concepts are vague and subject to hard cases; (d) the way we assess evidence and weigh values can be shaped by our total life experience; (e) different normative considerations on different sides can make overall assessment difficult; and (f) the number of values any social institution can incorporate is limited³⁷

³⁵Rawls, *Op. Cit.*, 55–56

³⁶ For this compilation I am following Jonathan Quon, "Public Reason," *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2013/entries/public-reason/>>.

³⁷ *Ibid.*

This account of the burdens of judgement can and has been criticized,³⁸ but our purpose here is not to dispute Rawls's view on the limits of reason; it is rather to note that this is a historical view, an understanding of both the limits of reason and, by implication, of one's own views of the world that responds to the failures of the modern project of reason. Reasonableness demands of persons that, when entering the public sphere, they not take their views too seriously. They must be able to act as if they did not believe in the truth of their views and accept that theirs might be as wrong as the views they deem wrong in principle.³⁹ They should assume that the reason for accepting to live as if their beliefs could be false is the imperfection of the tools of rationality and science, namely, evidence, considerations, concepts, normative assessments, and finite institutional possibilities.

Not surprisingly, these are anthropocentric considerations, perfectly valid and defensible, but in direct contradiction with other comprehensive views that might see the flaws that Rawls calls 'burdens of judgement' precisely as the outcome of trying to support one's views on anthropocentric facts and values, or on the attempt at grounding one's views in reason. To summarize these first three characteristics of the reasonable, it appears to be an *ethos* of the person who is willing to live by the belief that there must exist a fundamental equality and sociality of people in the public square, and to partake in it as if his or her private beliefs were not to be taken too seriously.

³⁸ e.g. Galston, W., *Liberal Pluralism: The Implications of Value Pluralism for Political Theory and Practice*, (Cambridge: Cambridge University Press, 2002), 46–47; Gaus, G. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, Oxford: Oxford University Press, 1996), 131–136; Wenar, L., 1995, "Political Liberalism: An Internal Critique," *Ethics*, 106(1), 1995: 41–48.

³⁹ This describes a certain type of skeptical position, particularly one of Agrippa's five tropes of skepticism, namely, dissent — the uncertainty demonstrated by the differences of opinions among philosophers and people in general.

3.c Reasonability as practical reason

The quote from Rawls continues: “Add to this the principles of practical reason and the conceptions of a society and person on which the political conception is based.”⁴⁰

Under the rubric “principles of practical reason” it is not entirely clear how much Rawls is willing to borrow from Kant’s moral theory, but he sees his own political constructivism as both distinct and compatible with “rational intuitionism” and Kantian moral constructivism (similar to the way he sees ‘justice as fairness’ as distinct yet compatible with reasonable comprehensive views).⁴¹ With respect to principles of practical reason, Rawls buys into Kant’s distinction between theoretical and practical reason in the sense that “practical reason is concerned with the production of objects according to a conception of those objects...while theoretical reason is concerned with the knowledge of given objects.”⁴² This distinction in human faculties is then used by Rawls to further distinguish two “different conceptions of objectivity appropriate from theoretical and practical reason.”⁴³ In this way, the notion of objectivity needed for political purposes can avoid the “special difficulties” that theoretical reason faces when it must satisfy a “causal requirement” for the objective view proper of “logic and mathematics”⁴⁴. In contrast, that requirement is not essential for a conception “suitable for moral and political reasoning” because it is sufficient that “the reasons offered be sufficiently strong.”⁴⁵

⁴⁰ Rawls, *Political Liberalism*, 94

⁴¹ for example, in *Political Liberalism*, p-89-106, and elsewhere.

⁴² *Ibid.*, 93

⁴³ *Ibid.*, 117

⁴⁴ *Ibid.*, 118

⁴⁵ *Ibid.*, 118

In the context of evaluating a possible objective point of view for moral and political theory and distinguishing it from “the point of view of any particular agent, individual or corporate, or of any particular group of agents, at any particular time,”⁴⁶ constructivism identifies that point of view with that “of certain reasonable and rational persons suitably specified,” which, in “justice as fairness is the point of view of free and equal citizens properly represented.” This freedom and equality form a “somewhere,” a vantage point for the objective view of moral and political constructivism on the world, “in contrast to what Nagel calls ‘the impersonal point of view.’”⁴⁷

The distinction between theoretical and practical reason allows Rawls to say that when questioning someone’s political or moral objectivity, there is nothing to question but “our sincerity and reasonableness.”⁴⁸ Again, the notion of moral and political objectivity, via Kant’s practical reason, ends up referring to the reasonable: reasonableness incorporates principles of practical reason in order to say that one is reasonable if she accepts the separation of practical and theoretical reasons, so that the notion of objectivity can be separated from the epistemological problems of theoretical reason. The price of that softening of conditions, which parallels the ‘method of avoidance’ and other separations in the system, is again a certain anthropology that describes the rational in a particular way, and the works of our moral behavior as somehow following Kant’s idea of rationality. That separation excludes every view, except for the notion of equality and freedom that ‘justice and fairness’ advocates. It is designed to protect it as neutral and morally objective.

⁴⁶ *Ibid.*, 111

⁴⁷ *Ibid.*, 115-116

⁴⁸ *Ibid.*, 118 This position runs the risk of assuming radical anti-consequentialist stance on ethics. This is indeed a kind of “as if” posture, as opposed to an engaged, being-in-the-world, position.

This becomes clearer in another aspect of practical reason that Rawls borrows from Kant. Even though Rawls rejects “Kant’s constitutive autonomy”⁴⁹ because it bears the mark of a comprehensive view, he writes:

Certainly, political constructivism accepts his view that the principles of practical reason originate, if we insist on saying they originate anywhere, in our moral consciousness as informed by practical reason. They derive from nowhere else. Kant is the historical source of the idea that reason, both theoretical and practical, is self-originating and self-authenticating.⁵⁰

This is particularly relevant since, as we already know, the concept of the reasonable functions as a gatekeeper for the public square. One can only enter into the debate if one adopts the principle of reasonable pluralism. And to enter into it implies accepting the fact that for moral purposes we operate with principles of practical reason, applied in a rational way, through chains of reasons, as it were, beginning with a reason that is self-originating and self-authenticating, an origin labeled as ‘moral consciousness.’ Around the notion of the reasonable, the liberal anthropology and its commitments becomes ever more clear.⁵¹

3.d The essential aspects of a moral person

To continue with the characteristics of the reasonable, after the principles of practical reason (#4), Rawls adds (#5): “the conceptions of a society and person on which the political conception is based.” These are what Rawls calls ‘model conceptions of a well-ordered society and of a moral person.’⁵² Their general purpose is:

⁴⁹ *Ibid.*, 100

⁵⁰ *Ibid.*, 100

⁵¹ This notion of autonomy will be criticized more deeply in the following chapter.

⁵² For example, in “Kantian Constructivism in Moral Theory” *The Journal of Philosophy*, Vol. 77, No.9, pp. 515-572 (Sep. 9, 1980), p520, and elsewhere.

to single out the essential aspects of our conception of ourselves as moral persons and of our relation to society as free and equal citizens. They depict certain general features of what a society would look like if its members publicly viewed themselves and their social ties with one another in a certain way.⁵³

A well-ordered society is the ‘idealized concept’⁵⁴ of the kind of society that an effective political proposal should produce according to liberalism. Its three main characteristics are the common acceptance of the very same principles of justice for the political domain, the effective application of those principles to the basic political structure of society, and the willful compliance with both the principles and their institutionalized form.⁵⁵ Going deeper into the model-conception of a well-ordered society we find the usual notions in Rawls’s “impressively self-contained” system. The reason for this model-conception is the existence in “the political culture of a democratic society” of three unavoidable characteristics: “the fact of reasonable pluralism”, “the fact of coercion” (no single comprehensive view can be the organizing force of a democratic society without coercion), and the need for a conception of justice “limited to what I shall call ‘the domain of the political’ and its values.” So a reasonable person is one who accepts the idea that a well-ordered society is one in which people consent to a freestanding political conception limited to the domain of the political because of the fact of reasonable pluralism. The circularity of the notion of reasonableness appears once again.

Perhaps more crucial in this context is Rawls’s conception of a moral person:

we take moral persons to be characterized by two moral powers and by two corresponding highest-order interests in realizing and exercising these powers. The first power is the capacity for an effective sense of justice, that is, the capacity to understand, to apply and to act from (and not merely in accordance with) the principles of justice. The second moral power is the capacity to form, to revise, and rationally to pursue a conception of the good. Corresponding to the moral powers, moral persons are said to be

⁵³ *Ibid.*, 520

⁵⁴ Rawls, *Political Liberalism (Op. Cit.)*, 35

⁵⁵ *Ibid.*, 35

moved by two highest-order interests to realize and exercise these powers. By calling these interests "highest-order" interests, I mean that, as the model-conception of a moral person is specified, these interests are supremely regulative as well as effective. This implies that, whenever circumstances are relevant to their fulfillment, these interests govern deliberation and conduct.⁵⁶

These two moral powers express the relationship between the reasonable and the rational in Rawls' understanding of the person. The first moral power, the capacity for acting from the principle of justice expresses the reasonable side of the moral sphere in the sense that it implies that a person can behave according to the rules of justice that the 'fair system of cooperation' has established. The second moral power expresses the rational side inasmuch it shows the connection between the self and the good, as an end for that self. It is an end that is self-centered, even in the case where it happens to express a fundamentally altruistic comprehensive view, because in its lesser version, it is a capacity to calculate the best means to achieve a given benefit, and in its highest version, it is the capacity to order one's life according to the comprehensive view or conception of the good that one holds. In this tension between the reasonable side and rational side of the moral sphere, the rational side informs the reasonable, and the reasonable limits the rational. One's means to an end, or one's comprehensive view, are limited by the public principles of justice, or, in other words, right limits the Good.⁵⁷ A reasonable person with an effective sense of justice is willing "to act in relation to others on terms that they can also publicly endorse."⁵⁸

⁵⁶ John Rawls, "Kantian Constructivism in Moral Theory" *The Journal of Philosophy*, Vol. 77, No.9, pp. 515-572 (Sep. 9, 1980) p 525

⁵⁷ *Ibid.*, 528

⁵⁸ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 19.

Rawls admits that this conception of the person with moral powers is an idealized and simplified form.⁵⁹ We will see what this idealization means further below, but so far it is clear that saying this does not reduce an obvious problem with this conception of the person and its moral powers: it clearly describes an idea of human nature that, together with the Kantian notion of self-originating and self-authenticating claims, suggests a formal, a priori understanding of the self as a moral agent that is difficult to articulate without a commitment to metaphysical ideas of freedom, or from a shared, fundamental human nature that justifies a fundamental equality, or a particular instance of a greater comprehensive view. In order to solve this, Rawls claims — in a footnote — that this understanding of a moral person is just ‘normative,’ and, again, that it applies only to the domain of the political.

In the present case the conception of the person is a moral conception, one that begins from our everyday conceptions of the person as the basic units of thought, deliberation, and responsibility, and adapted to a political conception of justice and not to a comprehensive doctrine. (...) As a normative conception, it is to be distinguished from an account of human nature given by natural science and social theory and it has a different role in justice as fairness.⁶⁰

In this context, this is a distinction without a difference. Recall that, as shown above, these two moral powers are related to two higher order interests that are “supremely regulative as well as effective,” in the sense that these interests both “govern deliberation and conduct”.⁶¹ Any understanding of the person that hinges on its moral capacity, as a force that determines deliberation and conduct from within a hierarchy of what is reasonable (i.e. social and political), and what is rational (i.e. self-centered and comprehensive), —privileging the former over the latter as the basis for the success of a conception of justice that must govern a society—, is an

⁵⁹ *Ibid.*, 20

⁶⁰ *Ibid.*, 18, footnote 20.

⁶¹ John Rawls, “Kantian Constructivism in Moral Theory” *The Journal of Philosophy*, Vol. 77, No. 9, pp. 515-572 (Sep. 9, 1980) 525

idea of human nature in the strong sense, relying on a metaphysical understanding of moral personhood that is assumed to exist universally. This is not a metaphysical commitment in the weak sense in which every theory might involve some position on the ultimate causes of the universe, even if only tacitly. This is a metaphysical commitment in the strong sense of an a priori understanding of the place and powers of human nature to regulate our conduct and place in the social order, most notably in the context of the understanding of our rational nature as being a failed nature via the ‘burdens of judgement.’ The fact that it is applied to the domain of the political is not only useless as a disclaimer; if anything, it adds to our claim that it is an instance of metaphysical commitment. The effective sense of justice that is connected to stability is an account of obedience that rules out coercion: it explains a sort of natural form of allegiance based on natural moral powers found a priori, that is, independent of all circumstances of existence.

4. The Original Position

All this is better seen in the ‘original position,’ which yields the principles of justice. Recall that the original quote about the meaning of the reasonable ends with a final (#6) characteristic: that we understand the aspects that define the reasonableness of persons in terms of “how these enter into the procedure of construction and why.” The procedure of construction is the process by which we are able to obtain principles of justice for the sphere of the political independent of any metaphysical commitments. This is the original position, a model-conception that mediates between the ideal of a well-ordered society (in which individuals obey the principles of justice effectively and for the right reasons) and the moral understanding of persons:

The original position is a third and mediating model-conception: its role is to establish the connection between the model-conception of a moral person and the principles of justice that characterize the relations of citizens in the model-conception of a well-ordered society. It serves this role by modeling the way in which the citizens in a well ordered society, viewed as moral persons, would ideally select first principles of justice for their society. The constraints imposed on the parties in the original position, and the manner in which the parties are described, are to represent the freedom and equality of moral persons as understood in such a society. If certain principles of justice would indeed be agreed to (or if they would belong to a certain restricted family of principles), then the aim of Kantian constructivism to connect definite principles with particular conception of the person is achieved.⁶²

The criticisms that Dworkin and Walzer direct against the procedure of the ‘original position’ capture the two weaknesses that best explain why at the core of Rawls’s system there is a material idea of human nature with metaphysical undertones. They are also helpful in showing how the reasonable is a carrier that smuggles that material determination in the system. Let us see them in detail.

4.a The Original Position Determines the Outcome

The original position is a way for Rawls to answer the question ‘how are we to specify the fair terms of cooperation?’ in a way that can fulfill the condition that they must be “settled by an agreement reached by free and equal cooperation, and made in view of what they regard as their reciprocal advantage or good.”⁶³ It is obvious that we face a circularity problem here. To stipulate fair terms of cooperation for free and equal persons we must reach agreement by free and equal cooperation. The original position works as the center of that circle. “The difficulty is this”—says Rawls—:

⁶² *Ibid.*, 520

⁶³ John Rawls, *Justice as Fairness. A restatement*, (Cambridge: The Belknap Press of Harvard University Press, 2001) 14-15

we must specify a point of view from which a fair agreement between free and equal persons can be reached; but this point of view must be removed and not distorted by the particular features and circumstances of the existing basic structure. The original position, with the feature I have called the ‘veil of ignorance’, specifies this point of view.⁶⁴

As we have already seen in Chapter III, Walzer contends that most, if not all, of the procedural devices designed to obtain principles of justice are ultimately flawed, in the sense that they pose as conditions for the design precisely the kinds of things that will be found as conclusions at the end of the process. These designs translate normative content into formal procedures. The procedure is meant “to liberate” participants in argumentation “from domination, subordination, servility, fear, and deference. Otherwise, it is said, we could not respect their arguments and decisions.”⁶⁵

But once rules of this sort have been laid out, the speakers are left with few substantive issues to argue and decide about. Social structure, political arrangements, distributive standards are pretty much given; there is room only for local adjustments. The thin morality is already very thick—with an entirely decent liberal or social democratic thickness.⁶⁶

Here, Walzer has both Habermas and Rawls in mind. This idea that one finds at the end of the process what one placed at the start is intended to rule out bargaining and compromise (the negotiation of particular interests), and to press the speakers toward a preordained harmony. It’s pulling a rabbit from a hat: it’s not real magic because you hid it there from the start. It arms social theory “against the indeterminacy of natural conversation.”⁶⁷ The contrast between what is repressed—the natural bargaining of interests—and the usual indeterminacy of “real talk” suggests to Walzer “the dominance

⁶⁴ *Ibid.*, 15

⁶⁵ Michael Walzer, *Thick and Thin*, (Notre Dame: University of Notre Dam Press, 1994), 12-13

⁶⁶ *Ibid.*, 12-13

⁶⁷ Michael Walzer, “A Critique of Philosophical Conversation,” in Michael Kelly (Ed.) *Hermeneutics and Critical Theory in Ethics and Politics*, (Cambridge: MIT Press, 1990), 186-187

of design over discourse.”⁶⁸ It is a theoretical construction, not an organic social development; a product of theoretical reason and therefore transcendent, as opposed to the alleged immanence of the intersubjective process.

Rawls’s conditions, expressed in the veil of ignorance, imply that the parties to the negotiation of the social contract are not allowed to know “their social positions or the particular comprehensive doctrines of the persons they represent. They also do not know a persons’ race and ethnic group, sex, or various native endowments such as strength and intelligence.”⁶⁹ Behind the hypothesis of the original position reasonability and neutrality coincide. These actors are, in a very specific way, impartial. The various sources of their partiality are eliminated in the process of decision. They are purely neutral. We see how in this “device of representation”⁷⁰ it is affirmed explicitly that one can actually make decisions in a purely formal way; that there exists the possibility of reasoning about one’s general advantage while subtracting one’s own identity and obtaining, so to speak, a view from nowhere.

This might very well be so, but it may be equally legitimate to see it as a historically rooted idea, it is very much conditioned by and tied to the historical time and place of its emergence. These roots nourish and tether what crops out of them. The ideal of the person as free and equal that undergirds this very formal possibility is the outcome of a particular anthropological theory – particularly Kantian, but generally speaking liberal

⁶⁸ *Ibid.*, 186

⁶⁹ John Rawls, *Justice as Fairness. A restatement*, (Cambridge: The Belknap Press of Harvard University Press, 2001), 14-15

⁷⁰ *Ibid.*, 17

through and through – that defends the possibility of a mind, or a soul, or a consciousness, free from the contingencies of body, life, and history. This can be traced back to some form of Platonism, but it is more precisely linked to the political development of the *modern* tradition beginning with Luther’s Reformation. As we see in the introductions to *The Freedom of the Christian* and *Concerning Governmental Authority* Luther can argue in favor of separating the domain of the political from religion on the basis of “the freedom of belief”. I will develop this idea later on, but it is already worth noting here that this freedom, opposed to the contingencies of the physical and historical presence of the body, is what allows for the quest of justice, and justifies, as is well known, the comprehensive doctrine that Rawls himself calls liberalism and articulates in *A Theory of Justice*. The veil of ignorance is a postmodern or hypermodern form of that same separation. Justice does not appear as a bargaining of interests, but as the outcome of a rational process that is free because it is equal, stripped of all contingent circumstances and physical embodiments, and equal because it is free from contingencies. This is the perspective that allows for fairness to emerge.

4.b The Original Position Is Not a Contract

Dworkin agrees with Walzer’s point that what Rawls draws from the system as a conclusion is already put in it earlier. But he also notices that this stripping away of all contingencies makes for a very uniform outcome:

Since no one knows anything about himself that would distinguish him from anyone else, he cannot rationally pursue any interest that is different. In these circumstances nothing

turns on each man having a veto, or, indeed, on there being more than one potential party to the contract in the first place.⁷¹

This interpretation, along with the fact that Rawls does not claim that the contract stemming from the original position is anything more than a “non-historical” hypothesis, nor that it has ever been upheld in history “or indeed ever could actually be entered into,”⁷² leads Dworkin to claim that the purpose of the original position cannot “supply an independent argument for the fairness of enforcing” the two principles that ‘justice as fairness’ proposes. “A hypothetical contract is not simply a pale form of an actual contract; it is no contract at all.”⁷³

The reason for the unbinding nature of the ‘contract’ and its unsuitability to serve as a source of independent argument is a precise consequence of the veil of ignorance, since it violates the reason for which a contract is useful as a social device for reaching agreement. The fact that it expresses, as an outcome, the conditions imposed as ‘fair’ for the decision itself, is what cancels its contractual nature.

A judgement of antecedent interest depends upon the circumstances under which the judgement is made, and, in particular, upon the knowledge available to the man making the judgement. (...) The fact, therefore, that a particular choice is in my interest at a particular time, under conditions of great uncertainty, is not a good argument for the fairness of enforcing that choice against me later under conditions of much greater knowledge. But that is what, on this interpretation, the original position argument suggests, because it seeks to justify the contemporary use of the two principles on the supposition that, under conditions very different from present conditions, it would be in the antecedent interest of everyone to agree with them.⁷⁴

⁷¹ Ronald Dworkin, “The Original Position,” *University of Chicago Law Review*. (500 1972-1973), 532

⁷² Rawls, J., *Op. Cit.*, 17

⁷³ Dworkin, R., *Op. Cit.*, 501

⁷⁴ *Ibid.*, 503

If the circumstance of the original position is the one in which differences are cancelled through the veil of ignorance, it is normal that the outcome expresses these circumstances. But that only means that the outcome is posited as a premise. To say that the outcome is the premise means that the premise comes from somewhere else, and carries a force different from the analytical force of the ‘original position’, even as a “thought-experiment for the purpose of public- and self-clarification,” as Rawls would say in the watered-down version of his last restatement of ‘justice as fairness.’⁷⁵

4.4 The Method of Reflective Equilibrium

This is why Richard Dworkin argues that Rawls’s original position is a device to test the competing arguments that can be derived from a basic, natural right, which is posited by Rawls as he seeks a fundamental grounding for his political theory, and in particular, it is the core of his political conception of ‘justice as fairness.’ In order to do so, Dworkin analyses three features of Rawls’s theory: “the technique of equilibrium, the social contract, and the original position itself.”⁷⁶

His argument begins with an analysis of the method of reflective equilibrium. This method implies that one must be able to adapt one’s general principles of moral theory and one’s moral intuitions reciprocally —back and forth— to make them coherent. The requirement of coherence is a rational requirement, which reflects the general rationalistic paradigm of the theory. The idea that a conception of justice must be coherent and bridge the distance between moral intuitions and theoretical principles casts light on the formal aspect of the theory. As we

⁷⁵ Rawls, J., *Op. Cit.*, 17

⁷⁶ Dworkin, R. *Op. Cit.*, 508

have seen above, this coherence, camouflaged as logical consistency, depends upon a conception of our moral and rational powers endowed with the kind of freedom-from-contingency proper to the tradition of Christian liberalism, with its modern dualisms, and its *a priori* conditions. Precisely to avoid the problems that arise when we endorse certain aspects of rationalism, Rawls opposes the idea of the reasonable to the idea of the rational, but he does so, in part, by means of the method of equilibrium.

Dworkin begins by saying that there are two kinds of models that allow us to achieve this coherence: naturalist models and constructivist models. Naturalist models propose that moral “intuitions are clues to the nature and existence of more abstract and fundamental moral principles, as physical observations are clues to the existence and nature of fundamental physical laws.”⁷⁷ In this view, moral theories are discovered, not designed, in that they describe an objective moral reality. Moral intuitions can never be adapted to fit a theory, because that would imply that the observed data is flawed. Such intuitions are, in this sense, uncompromising. The constructivist model treats “intuitions of justice not as clues to the existence of independent principles, but rather as stipulated features of a general theory to be constructed.”⁷⁸ As we know, Rawls’s original position, as a case of social contract theory, is a constructivist theory. But it is a highly specific kind of contract, with highly specific constraints.

Before we go into why these constraints lead Dworkin to claim that at its core ‘justice as fairness’ is a natural right theory, let us further examine the consequences of endorsing the method of reflective equilibrium as if it were a contingent and flexible form of moral reasoning that avoids any metaphysical commitments. In order to enrich Dworkin’s take on reflective equilibrium, I

⁷⁷ *Ibid.*, 510

⁷⁸ *Ibid.*

want to claim that the method is a device that makes it possible for a certain set of beliefs (namely, those that constitute the comprehensive liberal doctrine pertaining to *A Theory of Justice*), to be presupposed and camouflaged as neutral. One way to make cast light on the significance of this camouflage is to articulate a version of the criticism that Michael Della Rocca directs against the “method of intuition” in general, and, in particular, against the most widely accepted form of Rawls’s method of reflective equilibrium.⁷⁹

The principal aim of Della Rocca’s work is to explain a certain domestication of philosophy in the last 100 years through the prevalence of the ‘method of intuition’ (or MI, as he abbreviates it)⁸⁰ in analytic philosophy. The meaning of MI is straightforward: it is the method whereby intuitive responses are central to one’s philosophical theorizing. The aim of any overarching theory should be to reconcile these intuitions —as the everyday, recognizable psychological experiences that they are— with the fundamental principles a theory, to the extent that theoretical principles should, at the very least, be modified to comply with the content of those intuitions.⁸¹ Preserving those intuitions is the main goal, even if the adherents of MI endorse a moderate version of it, and thus accept that a well-grounded principle might force him or her to discard or modify some aspects of a given intuition. Hardcore adherents of MI will retain the intuitions come what may, while moderate adherents of MI will adopt the method of reflective equilibrium (“MRE”⁸²). More specifically, Della Rocca identifies three types of MI adherents. Particularists —who only accept concrete, particular intuitions—, hardcore intuitionists —they

⁷⁹ Michael Della Rocca “The Taming of Philosophy,” in *Philosophy and its History*, Ed. Mogens Laerke et al., (Oxford: Oxford University Press, 2013), 178

⁸⁰ *Ibid.*

⁸¹ *Ibid.*, 179

⁸² *Ibid.*, 180

conserve intuitions come what may and adapt principles to them—, and flexible intuitionists, MRE adherents (the most common ones) —who believe that intuitions are subject to revision and reconstrual when they conflict with well-reasoned principles, and principles can also be modified in a sort of reflexive flexibility, back and forth in search of congruency.⁸³

By intuitions, Della Rocca means opinions, judgements and beliefs regarding particular cases, especially when these are deemed “common sense” or “ordinary,” and are accepted as points of departure for philosophical work. However, Della Rocca acknowledges that when Rawls defines his method, he never speaks of following mere intuitions or common sense beliefs, but rather of “considered judgements,”⁸⁴ which, while they stem from common sense and from intuitions, entail something more, as we will see shortly. This way, the basic criticism that Della Rocca directs against the method of intuition, that is, that it is conservative and thus arbitrary (because one must return to common sense as soon as possible, when theorizing, for no significant reason other than an unwarranted love for ordinary beliefs), is more difficult to direct against Rawls. Considered judgements are not simply moral judgements about a particular case that must carry some weight in deciding which principles of justice we may reasonably endorse. They are *considered* in the sense that they are the best possible judgements we, as moral and rational agents, are capable of, and under conditions in which our capacities are most likely to have been fully exercised.⁸⁵ These conditions and these capabilities are the conditions and capabilities of neutrality:

⁸³ In a certain sense, this discussion parallels the difference between *Sittlichkeit* and *Moralität* discussed in the previous chapter

⁸⁴ Rawls, J., *Op. Cit.*, 29

⁸⁵ *Ibid.*

the positions of judges, umpires, and referees are designed to include conditions that encourage the exercise of the judicial virtues, among them *impartiality* and judiciousness, so that their verdicts can be seen as approximating considered judgements, so far as the case allows.⁸⁶

The method of reflective equilibrium expresses the virtue of neutrality in judgement.

Della Rocca is as unimpressed with the degree of subtlety that the word ‘considered’ might play in Rawls’s judgements, as he is with the idea advanced by Rawls (in response to criticism of the original version of *A Theory of Justice*), according to which his reflective equilibrium is “wide” and not “narrow.”⁸⁷ It is ‘wide’ in the sense that it includes the contrasting of different moral and political theories and the presence of “background facts concerning persons and morality” which may provide “more room for criticism and revision of ordinary beliefs and of considered judgements.”⁸⁸ He is unimpressed because his first charge against the method still holds:

Rawls may still face a charge of undue privileging of certain states or grasping over others. For we can ask: why should considered judgements in Rawls’s sense be our starting points or focal points? It may be that reality requires a sharper, more radical departure from even our refined intuitions or considered judgements which are, as presumably Rawls admit[s], fallible. There is no reason to think antecedently that we are more likely to arrive at the truth by attuning our theories to fallible intuitions —of whatever kind— than by adopting some other, extra-intuitive starting point.

That the MRE suffers from a status quo bias is the first great criticism of the three that he directs at reflective equilibrium. The remaining two are:

2) the focus on intuitions and/or judgements is a focus on psychological bits rather than on the world.

⁸⁶ *Ibid.*, 29-30 [my *italics*]

⁸⁷ *Ibid.*, 30-31

⁸⁸ Della Rocca, M., *Op Cit.*, 189

3) the method arbitrarily decides when a principle is prevalent over an intuition or judgement, and vice-versa.

These criticisms are relevant not only epistemically, but also because the method of reflective equilibrium is the main counterbalance to the original position. A reflective equilibrium is achieved by going back and forth between different levels of generality, from case to theory, when our moral and political judgements are in conflict with other judgements that we (or others) may hold. This is in fact the actual way to reach the principles of justice that Rawls advocates. It is the alternative to the ‘original position,’ which is described as a hypothetical “thought-experiment,” and a “device of representation.”⁸⁹

While Rawls’ original position appears to privilege a rationalistic approach to the problem of justice, by supposedly eliminating from the equation all possible personal biases —through the veil of ignorance—, later he also argues that the principles arrived at in the original position must be checked against, and changed according to, our considered judgements about justice, that is, our most sophisticated beliefs and intuitions (as Della Rocca sees them). Thus, the presence of those considered judgements, a form of common sense, that must be in reflective equilibrium with rational principles, introduces a historically grounded form of bias: the beliefs of a time are introduced as forms of justice together with the rational principles, preventing revolutionary or hitherto unconsidered ideas from becoming the currency of political philosophy. The taming of philosophy, indeed.

⁸⁹ Rawls, J., *Op. Cit.*, 17-19, 29-31

The endorsement or common sense, however sophisticated, implies a form of conservatism, of latitudinarian adherence to the relative prevalence of a morality belonging to a particular time in history. This is further stressed by the requirements that Rawls poses to evolve from a narrow to a wide reflective equilibrium. A person evolves from a narrow equilibrium to a wide one when he or she “has considered the leading conceptions of political justice found in our philosophical tradition.”⁹⁰ As an evolved form of common sense, it hardly goes beyond its own theoretical echo. In the same way that the method of intuition compares and contrasts common sense beliefs and principles, the wide reflective equilibrium obtains its legitimacy through checking one’s beliefs and principles with the common, historical theories of the time, which function as limits and warranted arguments, pushing the reflection into the historical boundaries of what has been already accepted in the present.

I believe this to be crucial because it partially defeats the purpose of the ‘original position.’ The ‘original position’ is meant to clean out the injustices of history from the procedure that designs the ideal way to decide which conception of justice is best. But when the parties in the original position make their decision, they also select among the same leading conceptions of political justice that have survived in our philosophical tradition.⁹¹ In the original position, the conditions to which the parties are a subject constitute the “reasonable” limits to which the rational, in turn, is completely subordinate. Those conditions express the need for a non-biased, symmetrical structure of the decision. This is problematic on its own terms, as I will show below, but let us assume for a moment that it holds. That being so, these leading theories would re-introduce

⁹⁰ *Ibid.*, 31

⁹¹ *Ibid.*, 80-85. There seems to be a disclaimer in *Ibid.*, 83, when Rawls admits that this is not the way to obtain the best possible conception of justice but only to obtain a moral basis to work with and “hold its own against the known alternatives”, which reinforces the role of the status quo, of holding ground in a battle of existing ideologies.

history as a condition. Why would they be leading if not due to the same unfair conditions that made the original position necessary in the first place?

Now, the need for a wide reflective equilibrium is a further step in this direction. After the criticism that Dworkin leveled at the hypothetical contract in the original position in *A Theory of Justice*, seen above, according to which a hypothetical contract is in fact no contract at all, Rawls responded by saying that the original position is merely a device of representation or a thought-experiment. Its aim was to model certain qualities and limits of 'justice as fairness' after the fashion of rational choice theory, so as to show the predictability of the choices that parties thus positioned would prefer.⁹² But, most importantly, he complemented the thought with the need for wide reflective equilibrium; so while the original position represents the rational moment of the theory and the person that is supposed to achieve a wide reflective equilibrium is assumed to be rational, this idea of the rational is twice constrained by the standard of the reasonable: first, in the conditions established in the design of the original position, and second, with the need for and subsequent hegemony of the method of reflective equilibrium over the original position. Both instances of reasonableness reach their limit in the historically grounded 'leading' theories of the time.

The reasonable is here a vehicle that carries certain values with overriding power, expressed in the constraints and in the limited range of common sense views from which history allows us to choose a conception of justice. Taking this argument to its logical extreme shows that the ideas of justice that neutralists like Rawls propose rely on and foster particular understandings of what

⁹² *Ibid.*, 17, footnote 17

the good is, and that these particular understandings are ultimately made of ordinary beliefs, and common sense theories that reflect the biases of the time, —and by no means impartial views of justice. Those who hold or live according to ideas that exist on the margins of the mainstream are sometimes thrust into the mainstream or sometimes further marginalized, according to the interests of the regime in place, which is what every regime in history has done. In this case, however, they are pushed into it with the pretext of impartiality and fairness, when in truth what neutrality conceals is a set of views stemming from the historical status quo, or what Blumenberg calls, “the latitude.”⁹³ These ignored, non-mainstream ideas are characterized as unreasonable because they cannot measure up to the standards of Rawlsian reasonability. They are left out of the basic structure of society, but not permanently: over time, a well-ordered society trims out the unreasonable ideas so that the people who hold them slowly conform to the limits of the reasonable.⁹⁴ After all, the conception of the person summoned by what Rawls describes as “citizens’ rights, liberties, and opportunities,” the conception of person that stems from “social institutions and traditions,” the conception of person that embodies a reasonable moral conception, is not only a political actor. It also “assumes a wide role as part of public culture.” What are the consequences of this wide role? “Citizens are made aware of and educated to this conception.”⁹⁵ This is crucial because not only the separation between the political sphere and the non-political sphere seems to collapse, but also it is implied that the values smuggled into the political sphere as if they were neutral form a set of incentives and coercions that force people to

⁹³ Hans Blumenberg, *The Genesis of the Copernican World* (Cambridge: MIT Press, 1987)

⁹⁴ Rawls, J., *Op. Cit.*, 43

⁹⁵ John Rawls, “Kantian Constructivism in Moral Theory” *The Journal of Philosophy*, Vol. 77, No. 9, pp. 515-572 (Sep. 9, 1980), 553

become, in their non-political sphere, more like the idea of the person proper of the political sphere.

Thus, the fact that reflective equilibrium precludes the possibility of radical, untamed, unthought-of ideas from taking part in the public sphere on an equal footing with other traditional ideas or in the construction of the conception of justice that Rawls promotes is not the only problem; a further problem is that it excludes certain beliefs from the public square in the name of a neutrality that is as metaphysically loaded as those beliefs it excludes, precisely insofar as they are metaphysical and unreasonable. With Della Rocca's idea of common sense as a form of taming, it can also be seen that in the process of reflective equilibrium, more mainstream, historically grounded values, are introduced into the idea of justice, in the name of flexible rationality.

The second criticism that Della Rocca levels at the method of intuition in general shows this problem to run much deeper. If the first criticism emphasized the fact that the focus on intuitions puts forth an "arbitrary conservatism in our psychological economy,"⁹⁶ Della Rocca's second criticism is that the emphasis on intuitions is a way of focusing on psychological bits instead of on the world, like older philosophers, such as Spinoza or Hume, used to do.⁹⁷ This renders the method of intuition "inherently unsuited for providing an account of reality."⁹⁸ By doing this, Rawls's method places too much weight on opinions and beliefs, which makes his project of impartial fairness fraught for the same reasons it tries to avoid affirming a version of reality. His

⁹⁶ Della Rocca, M., *Op. Cit.*, 191

⁹⁷ *Ibid.*, 191 and ss.

⁹⁸ *Ibid.*, 91

method of avoidance, as we have seen in Chapter II, rejects the debate about truth because that would turn his conception of justice into a comprehensive view. The alternative, however, is that the focus on intuitions or beliefs or leading theories gives rise to an arbitrary debate on doxography:⁹⁹ “It’s as if Rawls and other proponents of the MI are guilty of placing not so much a veil of ignorance but rather *a veil of intuitions* between us and reality.”¹⁰⁰

For Della Rocca, this is a case of the taming of philosophy, but what makes this charge of psychologism relevant to us is that it shows how both the modeling of the original position and the method of reflective intuition are means to discard political options or conceptions of justice based on certain understandings of the human psyche that are not warranted by any theory or hypothesis available for testing. The choice of one epistemology over another is arbitrary, like the prevalence of intuition-based methods over, say, a direct engagement with reality that Della Rocca proposes to his readers.¹⁰¹ The epistemological theories or hypotheses are not available for testing because Rawls denies that his conception of justice is a doctrine about any particular truth about the world or about human nature, but only a set of common starting-points that can wind up in an overlapping consensus. Della Rocca’s points on the conservatism of common sense and the irrelevance of the adjective *considered* next to the word *judgments*, together with my criticisms of the supposedly wide version of reflective equilibrium, show that the problem is not so much that Rawls prefers an intuition-based method, which is debatable, or that he favors a particular form of liberal democracy, which I myself find appealing, the problem is not even that he strives for the possibility of a neutral perspective that might offer a wide space for a pluralistic

⁹⁹ *Ibid.*, 192

¹⁰⁰ *Ibid.*, 192

¹⁰¹ *Ibid.*, 193

public sphere, which is a noble and obviously moral endeavor even if it is argued from the perspective of stability (another moral value disguised as pragmatism, by the way). The problem is, precisely, that the presence of neutrality as a goal makes it necessary for these choices to be obscured, camouflaged, or hidden, even if unwittingly so, in order to methodically avoid the discussion on groundings. But to avoid having discussion does not mean that one refrains from presuppose its terms, as we know. Moreover, from the fact that it can gather consensus from the majority of people, even if they are plural in their beliefs, it does not follow that it is not grounded in, and fostering, an idea of human nature strongly committed to metaphysical principles or beliefs. This is especially true when the system designed for these purposes appears to be biased in favor of the forces of the status-quo, which are usually blind to their commitments because they are normalized historically, as feminist critics, animal-rights activists, and critical race theorists have shown time and again.

If the method of reflective equilibrium tames philosophy, as Della Rocca says, for the same reasons it tames the public sphere. It is crucial to understand that this is not a side effect but its explicit function, as it is also the function of the ‘original position’: to select “acceptable restrictions on the reasons” to choose this or that political principle.¹⁰² The instrument to tame the public sphere is the idea of the reasonable, which, as we have seen, works as the *Cerberus*¹⁰³ of the public sphere. It is the need for impartiality that leads Rawls toward the need for a mechanism, a method, that can explain individuals’ choices for one principle over another as something detached from the historical realities of personhood. It thus requires an understanding

¹⁰² John Rawls, *Justice as Fairness. A restatement*, (Cambridge: The Belknap Press of Harvard University Press, 2001), 17.

¹⁰³ In Greek mythology, Cerberus (/ˈsɜːrbərəs/;[2] Greek: Κέρβερος Kerberos [ˈkerberos]), often called the "hound of Hades", is the monstrous multi-headed dog that guards the gates of the Underworld to prevent the dead from leaving. Wikipedia. Accessed, April 19th 2017

of the person that can deal with his own intuitions in a way that avoids the question of their accuracy vis-à-vis reality or truth, (as Della Rocca points out), while checking their validity at more and more abstract levels of generality, against “leading” political philosophies, until the right principles are chosen.

The separateness that impartiality requires, then, makes the final choice, the point of reflective equilibrium, appear arbitrary. This is Della Rocca’s final criticism on the method of intuition in general, and the method of reflective equilibrium in particular.

Principles are not by themselves in a position to override intuitions, nor are intuitions by themselves able to overrun a principle that conflicts with those intuitions. Rather, there is a familiar give and take...And we must adjust principles and intuitions so as to come up with the most coherent overall system.¹⁰⁴

Its apparent flexibility vis-à-vis intuitions and principles does not yield any proper rule to decide when the principle or the belief or the judgement must recede in favor of the other: “For any line one draws between intuitions and principles there are other such lines that one could equally well draw within a coherent system of principles and intuitions.”¹⁰⁵ The method simply demands generally that one take into consideration “all levels of generality” so that they can be accommodated fairly.¹⁰⁶ Della Rocca half-jokingly mentions that this may be the reason why Rawls’s conception of justice is called ‘justice as fairness,’ which highlights the fact that this very general requirement begs the question, since this is supposedly a process to achieve a fair conception of justice. For Della Rocca, it is just another instance of the taming of Philosophy because it again introduces an arbitrary factor.

¹⁰⁴ Della Rocca, M., *Op. Cit.*, 194

¹⁰⁵ *Ibid.*, 194

¹⁰⁶ Rawls, J., *Op. Cit.*, 29-30

The MRE requires arbitrary distinctions and inexplicable relations that are not to be tolerated because these inexplicable relations constitute a failure of the MRE to engage non-arbitrarily with reality and thus they constitute an illegitimate narrowing of philosophy.¹⁰⁷

Della Rocca stresses that this failure implies an impoverishment of the range of philosophical inquiry and finds it particularly problematic precisely because “the very flexibility of the MRE — the very point that earns this method its august status—” is the reason for the failure.¹⁰⁸

What Della Rocca does not expressly recognize is that there is a precise reason for this apparent flexibility. In order to provide a clear set of instructions on how to decide in case of a conflict between principles and intuitions, one must be ready to make a claim about the nature of reality or the exact reach of our understanding, which would violate the method of avoidance that lets Rawls claim that his conception of justice is freestanding: it would turn his system into a comprehensive view on his own terms. As we know, however, the method is not that flexible, because a decision is ultimately made in favor of the principles that lie at the core of ‘justice as fairness,’ and in this process of decision the role of the ‘original position’ is central. Even if it is understood as a device of representation, it constitutes the fixed pole around which this flexibility operates.

4.d Dworkin and the Natural Right to Equal Respect

After this excursus, let us return to Dworkin’s criticism of the original position. Dworkin understands the original position as a device that expresses certain moral values, and not as a

¹⁰⁷ Della Rocca, M., *Op. Cit.*, 197-198

¹⁰⁸ *Ibid.*, 198

thought-experiment that attempts to show why the principles of justice are the best available. His view stems from the fact that the method of reflective equilibrium, once the harmony between intuitions and principles has been achieved, does not need the imprimatur of the ‘original position’ to be legitimate. Since the original position is not really a contract —as shown above—, nor anything that one can enter into except hypothetically, and is offered as a device for modeling certain reasons, it all suggests that the original position is just a “schematic representation of a particular mental process” that is predicated of most, if not all, human beings. It is, in other words, an expression of the underlying human nature that is assumed in Rawls’s system. Rawls himself, in the first version of *A Theory of Justice* says that the conditions embodied in the original position are the fundamental “principles governing our moral powers, or, more specifically, our sense of justice.”¹⁰⁹ This underlying anthropology will shape the way *other* comprehensive views are treated in the system, as we will shortly see.

For Dworkin, the original position “must be seen as a kind of halfway point in a larger argument, as itself the product of a deeper political theory that argues for the two principles *through* rather than *from* the contract.”¹¹⁰ Dworkin’s point is that this deeper theory is a natural rights kind of theory and that this is seen in the set-up of the original position. A fundamental natural right is posed at the beginning and then, the theory of justice, the conception of justice that outcrops from the original position, is the expression of that natural right posed as a precondition.

¹⁰⁹ John Rawls, *A Theory of Justice* “Original Edition” (Cambridge: The Belknap Press of Harvard University Press, 1971) 51, quoted in Dworkin, R., *Op. Cit.*, 509

¹¹⁰ Dworkin, R., *Op. Cit.*, 519

In order to offer a better understanding of this point, it is worth mentioning briefly that Dworkin develops the idea that almost all theories of justice might be classified as a goal-based, duty-based, or rights-based theory depending on which concept gives it “ultimate pride”: “some overriding goal, or some set of fundamental rights, or some set of transcendent duties, as fundamental, and other goals, rights, and duties subordinate and derivative.”¹¹¹ Examples of overriding goals would be “improv[ing] the general welfare,” corresponding to a theory of justice of the utilitarian family; of an overriding right, it could be “the right of all men to the greatest overall liberty,” which would correspond to a classical form of liberalism à la Mill or Locke; and an example of a fundamental duty could be “the duty to obey God's will as set forth in the Ten Commandments” or a theory of justice based on a Kantian categorical imperative.¹¹² Dworkin believes that all such types of theory are associated with their own “metaphysical or political temperaments and that one or another would be dominant,” depending on which political economy they design.¹¹³

Rawls's theory is a rights-based theory because in the original position a fundamental right is posited in the beginning: the right to equal respect. That is why the constraints eliminate all differences that could be used as arguments for self-interest or veto power—which is the point of a contract, that is, to have the right to veto the agreement. Of all the possible forms of equal respect that are available, the original position enforces the one that Rawls deems to be most fundamental, through the constraints. This is seen in the structure of the original position:

¹¹¹ *Ibid.*, 522

¹¹² *Ibid.*

¹¹³ *Ibid.*, 523

Since no one knows anything about himself that would distinguish him from anyone else, he cannot rationally pursue any interest that is different. In these circumstances nothing turns on each man having a veto, or, indeed, on there being more than one potential party to the contract in the first place.¹¹⁴

It is worth noting here that historically, social contract theory stems from an idea present most explicitly in Hobbes, which states an original, natural right to everything.¹¹⁵ The contract appears as a giving up of that right in exchange for some positive law that can secure peace and prosperity. This positive law mirrors a deep natural law, namely, the prohibition to act against one's preservation, and which expresses the rational desire for peace. From the first natural law, a second stems: reciprocity. "That a man be willing, when others are so too,...to lay down his right to all things."¹¹⁶ The reciprocity that the Hobbesian political covenant demands is based on the reciprocity that exists in the state of nature: we can kill one another with similar ease. Therefore, this 'I should accept if everyone else accepts' is the very articulation of the power of veto: the core natural right that the theory defends and Hobbes calls, not surprisingly, "liberty."¹¹⁷ The equality of treatment under the law (with the exception of the sovereign) is the outcome of the contract and the product of the natural power of veto that equalizes all of us.

In contrast, Rawls's theory assumes the existence of inequality as a point of departure. That is the reason why in his original position, all the information that could trigger a bargaining advantage by one of the parties over the others is concealed behind the veil of ignorance. The same goes for comprehensive views or social classes: parties know that the people they represent

¹¹⁴ *Ibid.*, 524

¹¹⁵ Thomas Hobbes, *Leviathan*, Ed. Edwin Curley (Cambridge: Hackett Publishing Company, Inc., 1994) 78 and throughout.

¹¹⁶ *Ibid.*, 80, (italics in the original).

¹¹⁷ *Ibid.*, 80 and throughout

hold comprehensive views or may belong to a social class, but they do not know which ones and “cannot design institutions” or “act to favor” their own class or those who hold one ideal over those who hold another.

The original position is well designed to enforce the abstract right to equal concern and respect, which must be understood to be the fundamental concept of Rawls's deep theory.¹¹⁸

They can only defend a system that would allow the interests of comprehensive views in general to be advanced, that is, formally —regardless of content— and, thus, equally. Why equally? Since the veil of ignorance conceals the content of the comprehensive view —seen by Rawls as “non-negotiable convictions”¹¹⁹—, all comprehensive views are understood formally, as being the same for the purposes of the contract, and they must be treated equally. However, as we know, this is only true of *reasonable* comprehensive doctrines, and this reasonableness corresponds with the set-up of the original position, which is based on the fundamental natural right to equal respect. This right is “natural”, according to Dworkin, in the sense that it is not “the product of any legislation, or convention, or hypothetical contract”¹²⁰ but the precondition for it. And he adds: “I have avoided that phrase because it has, for many people, disqualifying metaphysical associations.”¹²¹ Rawls ranks among these people.

The force of the natural right that original position defends connects Rawls’s anthropology with the conditions of the public sphere:

Justice as fairness rests on the assumption of a natural right of all men and women to equality of concern and respect, a right they possess not by virtue of birth or

¹¹⁸ Dworkin, R., *Op. Cit.*, 531-532

¹¹⁹ Rawls, J., *Op. Cit.*, 123

¹²⁰ Dworkin, R., *Op. Cit.*, 527

¹²¹ *Ibid.*

characteristic or merit or excellence but simply as human beings with the capacity to make plans and give justice. This right, he [Rawls] says, is "owed to human beings as moral persons," and follows from the moral personality that distinguishes humans from animals. It is possessed by all men who can give justice, and only such men can contract. This is one right, therefore, that does not emerge from the contract, but is assumed, as the fundamental right must be, in its design.¹²²

However, Dworkin does not believe that the assumption of natural rights is “a metaphysically ambitious one.”¹²³ In part, because he sees all political doctrines as somehow metaphysically loaded, which makes his understanding of ‘ambitious’ not a function of the intensity with which one is metaphysically committed but more as one of statistical variance: everyone is metaphysical, so being metaphysical along the normal line of the bell is not very ambitious. For example, he says that Rawls’s metaphysics

requires no more than the hypothesis that the best political program, within the sense of that model, is one that takes the protection of certain individual choices as fundamental, and not properly subordinated to any goal or duty or combination of these. This requires no ontology more dubious or controversial than any contrary choice of fundamental concepts would be and, in particular, no more than the hypothesis of a fundamental goal that underlies the various popular utilitarian theories would require.¹²⁴

For Dworkin, any right-based theory requires that the fundamental right or rights that it protects work as “independent grounds for judging legislation and custom,”¹²⁵ which is the same work that the goals or the duties do in their respective theories. Using Dworkin’s tripartite classification, it can be shown that Rawls’s original position mutes out goals and duties in favor of the right his theory protects. In the original position, parties represent what he calls “developed moral persons,” which means that they are aware of their final ends and the duties that their conception of the good might entail, but “the parties do not know the content of their conception of the

¹²² *Ibid.*, 532

¹²³ *Ibid.*, 527

¹²⁴ *Ibid.*, 527-28

¹²⁵ *Ibid.*, 528

good: its final ends.”¹²⁶ So the original position yields an “interest that moves the parties: a higher-order interest in protecting and advancing their conception of the good as best they can, whatever it may be.”¹²⁷

4.e Formal Comprehensive Views

It may be, as Dworkin says, that the metaphysical position that Rawls’s set up assumes is not as ambitious as other theories. Be that as it may, there is a stark contrast between the right that the set-up protects and the other rights, goals, and duties that the theory includes in the concept of “comprehensive view.” These can be defended as an abstract set of values of which we only know that they comply with the material content of the right to equal consideration and respect, which is why they are deemed reasonable and can be included in the political domain.

The fact that one can defend comprehensive views formally, without knowing their content, implies that Rawls understands comprehensive views primarily as mental adherence to a realm of mental objects that fall under a particular sphere of beliefs. Or, in other words: that the condition to satisfy in order to be considered a legitimate comprehensive view, so that it can be defended by the parties in the ‘original position,’ is one that constrains comprehensive views to be the kind of mental object that can be defended formally with the same arguments and limits that any other set of final ends or duties would be.

The presence of the natural right that the system protects at all costs, not only works as a limit on public sphere, as represented in the original position, or as the standard to evaluate everything

¹²⁶ John Rawls, “Kantian Constructivism in Moral Theory” *The Journal of Philosophy*, Vol. 77, No. 9, pp. 515-572 (Sep. 9, 1980), 524

¹²⁷ *Ibid.*

else. The fact that it also assumes a certain idea of what comprehensive views are and how they work within the human psyche shapes the limits of how they can be in the public domain, because the norms that obtain from the contract create the actual limits of the reasonable in the public sphere.

4.f Rawls's theological tradition

This shaping of comprehensive views as a mental object that can be dealt with formally, independently of content, also shows that Rawls's account of these views is historically grounded, fashioned after the Christian tradition and its quarrel with the domain of the political (he himself mentions the wars of religion in seventeenth-century Europe as a source for his idea of tolerance and respect).¹²⁸

The tradition it belongs to is the same that understands the underlying conflict variously as one of the realm of the secular against the spiritual, or *foro externo* against *foro interno*,¹²⁹ or public manifestations of allegiance against private exercise of religion,¹³⁰ or the spiritual aspect of freedom (as coextensive with consciousness and the self) in contrast with the material aspect of determination of body. The underlying tension can be observed from Descartes's *cogito* to Levinas' understanding of liberalism and Marxism as opposed to "Hitlerism."¹³¹

That these divisions or dichotomies are part of the European and Christian traditions (or at least, in opposition to Christian denominations) is the basis for the critique that Abdullahi A. An-Na'im

¹²⁸ In the introduction of both *Political Liberalism* and *Justice as Fairness*.

¹²⁹ Hobbes, T., *Op Cit.*, Part One, 93-95, and throughout

¹³⁰ See, for example, Baruch Spinoza, *Theologico-Political Treatise*, trans. R. H. M. Elwes, (New York: Barnes & Noble Inc., 2009), chapters XIX and XX.

¹³¹ Emmanuel Levinas, "Reflections on the Philosophy of Hitlerism," trans. Seán Hand, *Critical Inquiry* 17:1 (1990), pp. 62-71. This argument will be further developed in the following chapter.

directs against Rawls. An Na'im believes that the difference "between the theological history and political context of Rawls's thinking and that of Islamic politics in postcolonial Africa and Asia"¹³² is essential to understanding why Rawls was partially wrong when he asserted that An-Na'ims's *Toward an Islamic Reformation* is a "perfect example of overlapping consensus."¹³³ An-Naim offers a relativistic approach to the issue, that grounds a "profound suspicion of universalizing assertions emanating from North Atlantic societies," such as the appeal of Rawls's idea of public reason, which he finds "too reminiscent of the 'civilizing mission' of European colonialism."¹³⁴ This concrete criticism is turned into a general understanding of the problem of religion and state:

It is extremely difficult to conceive of "religion" in terms that are sufficiently inclusive to be applied to all human societies in their varied contexts, and it is only to be expected that, consciously or not, any theorist will develop his or her theory of the relationships between religion, state, and politics with reference to a specific religion in a particular sociopolitical context.¹³⁵

It is not our aim to judge whether such a culturally grounded perspective is warranted or not, but rather to show how from this perspective, Rawls's conceptual divisions, in connection with his formal view of comprehensive views, are seen from a different tradition. An-Na'im mentions various examples, most prominently the distinction between 'the so-called religious and secular domain,' which he sees linked to the Christian tradition but not applicable to Islam: the "inherent consistency" of both domains "induce Muslims to think of both as entwined," and any

¹³² Abdullahi A. An-Na'im, "Islamic Politics and the Neutral state, A Friendly Amendment to Rawls?" in, *Rawls and Religion*, Tom Bailey and Valentina Gentile (Ed) (New York: Columbia University Press, 2014), 248

¹³³ John Rawls, "The Idea of Public Reason Revisited," *The University of Chicago Law Review*, Vol. 64, No. 3 (Summer, 1997), pp. 779, n46, quoted in Abdullahi A. An-Na'im, *Op. Cit.*, 259

¹³⁴ Abdullahi A. An-Na'im, *Op. Cit.*, 245

¹³⁵ *Ibid.*, 260

separation “is not a workable solution for Muslims.”¹³⁶ The same is also applicable to the notions of the “spiritual and the material” and the meaning of “human agency,”¹³⁷ that he considers central in Rawls’s theory. An-Na’im’s criticism tries to make clear that some of the distinctions that are necessary to make the veil of ignorance workable are the remnants of a Christian tradition. Especially, the possibility of formalizing and psychologizing the content of a comprehensive view, and the kind of adherence and commitment that is supposed by the world ‘belief’ is predicated upon the acceptance of certain divisions proper of our metaphysical tradition.

4.g Rawls’s Answer

Going back to Dworkin’s criticism, after having confined comprehensive views to this formal, and possibly Christian, understanding, after having muted out all content except that of equal respect and having shaped accordingly all other views, rights, duties and goals, the women and men who have to live under the system that comes out of the original position “cannot choose other than to protect” that right, because in the original position everyone’s interest lies in the same solution.¹³⁸

The right of each man to be treated equally without regard to his person or character or tastes is enforced by the fact that no one else can secure a better position by virtue of being different in any such respect.¹³⁹

¹³⁶ *Ibid.*, 260

¹³⁷ *Ibid.*, 260

¹³⁸ Dworkin, R., *Op. Cit.*, 530

¹³⁹ *Ibid.*

Since the original position limits the discussion to what the right of equal concern and respect allows, and since it is more abstract “than the standard conceptions of equality that distinguish different political theories,” it “may now be seen as a device for testing these competing arguments.”¹⁴⁰ This is the range permitted by the fact of reasonable pluralism.

It may be a wide range, and the right that lies at its core might easily be agreed upon, after all it is not the liberty of Mill, as Dworkin himself says,¹⁴¹ but an abstract form of equality of respect, a highly spiritualized form of equality. However, as we have seen, the fact that at the core lies a fundamental right regarded as natural imposes unwarranted limits on other comprehensive views, and the fact that it is designed and presented as if it were freestanding, and neutral, involves the muting out of the content of those views, a muting out that is predicated upon the assumption that one can separate the content and the form of the comprehensive views so that they can conform to the limits of the reasonable.

Rawls answers Dworkin in his paper *Political, Not Metaphysical*, and up until the last of his works, the posthumously published *Justice as Fairness: A Restatement*, he mentions Dworkin’s paper and his response.¹⁴² Rawls’s answer is worth quoting in length:

I prefer not to think of justice as fairness as a right-based view. (...) I think of justice as fairness as working up into idealized conceptions certain fundamental intuitive ideas such as those of the person as free and equal, of a well-ordered society and of the public role of a conception of political justice, and as connecting these fundamental intuitive ideas with the even more fundamental and comprehensive intuitive idea of society as a fair system of cooperation over time from one generation to the next. Rights, duties, and goals are but elements of such idealized conceptions. Thus, justice as fairness is a conception based, or as Elizabeth Anderson has suggested to me, an ideal-based view, since these fundamental intuitive ideas reflect ideals implicit or latent in the public culture of a democratic society.

¹⁴⁰ *Ibid.*, 530

¹⁴¹ *Ibid.*, 533

¹⁴² John Rawls, *Justice as Fairness. A restatement*, (Cambridge: The Belknap Press of Harvard University Press, 2001), 17, footnote 17

In this context the original position is a device of representation that models the force, not of the natural right of equal concern and respect, but of the essential elements of these fundamental intuitive ideas as identified by the reasons for principles of justice that we accept on due reflection. As such a device, it serves first to combine and then to focus the resultant force of all these reasons in selecting the most appropriate principles of justice for a democratic society. (In doing this the force of the natural right of equal concern and respect will be covered in other ways.) (...) Others may prefer his account.¹⁴³

I fail to see how this addresses the main point that Dworkin directs against Rawls's 'original position', especially because this is the only moment in which Rawls, in addition to rejecting it, attempts an answer. What Rawls offers instead of a counter-argument, however, is a variation on the topic of reflexive equilibrium. The "working up" of certain intuitive ideas, including freedom and equality, into "idealized conceptions" that political philosophy finds in the background culture of democratic societies, has all the characteristics that critics of the method of reflective equilibrium criticize, as seen above.

In particular, the strong bias towards the status quo, seen in the idea of the culture of democratic societies as a source for these intuitive ideas, lends credence to the culture-based criticism of An-Na'im, as well as Walzer's and Della Rocca's approaches. Also, it's hard to see how it's not merely a distinction without a difference to say that the original position "models the force" of these intuitive ideas, "combine[s]" them, and then "focus[es]" them to select the appropriate principles of justice, instead of saying, as Dworkin or Walzer do, that the original position is a device to protect a certain fundamental right or idea posed *a priori*, because it is seen as essential, universal, or natural. There is no difference either to saying, as Della Rocca does, that this "working up" of intuitive ideas creates a "veil of intuitions" instead of a "veil ignorance."

¹⁴³ John Rawls, "Justice as Fairness: Political not Metaphysical", *Philosophy and Public Affairs*, Vol. 14, No. 3. pp. 223-251 (Summer, 1985), 237, footnote 19

The method of avoidance, that leads Rawls to simply reject Dworkin's (and others') criticisms, and offer an "ideal-based view" as an alternative level, is not a sufficient response. Not using the vocabulary of metaphysics so as to gather consensus does not entail that one is avoiding tacit metaphysical assumptions. However, in this instance, Rawls admits that the final goal of the whole procedure is to connect "fundamental" intuitive ideas "with the even more fundamental and comprehensive intuitive idea of society as a fair system of cooperation over time from one generation to the next."¹⁴⁴ To what extent a "fundamental comprehensive intuitive idea" is different from a comprehensive view remains a mystery, —not that in its content one can see much of a distance. After all, when Rawls discusses the evolution from a narrow to a wide reflective equilibrium, discussed above, he concludes that his ideal is a "full reflective equilibrium": "the adjective 'full' we reserve for features as realized in a well-ordered society" in which the public point of view is "affirmed by all in a full reflective equilibrium".¹⁴⁵ Or in other words: a reasonable society. Rawls claims that this is 'non-foundationalist' because "considered judgements may have an intrinsic reasonableness to reasonable persons that persists after due reflection,"¹⁴⁶ which, after what we have seen, is circular reasoning.

To escape this circularity, Rawls needs this 'intrinsic' reasonableness to be social. Rawls introduces the social, intersubjective, aspect of reasonableness (that, in a way, we have already seen in Habermas, in the previous chapter), through his idea of public reason.

¹⁴⁴ *Ibid.*

¹⁴⁵ John Rawls, *Justice as Fairness. A restatement*, (Cambridge: The Belknap Press of Harvard University Press, 2001), 31

¹⁴⁶ *Ibid.*

5. The Idea of Public Reason

Public reason must be able to bridge the distance that we find between the rational assent of a theoretical problem and the urgency of practical exigencies that all political communities must deal with.¹⁴⁷ In other words: if the modern project of ‘objective reason’ expressed in ‘rational intuitionism’¹⁴⁸ had succeeded to the extent of definitive truths, even in moral issues, this gap would not exist. But it still does. This is the reason why plurality is regarded not only as a fact of human nature *qua* social, or a natural result of liberal democracy, but it is also a necessary outcome of the weakness of reason. Conversely, plurality and disagreement serve as proof of the limits of reason and, at the same time, of the relevance of perspectives in cognitive standpoints. We are not heading towards a radical form of skepticism here, nor towards a form of sophistic moral relativism; but the idea of neutrality, as seen in most neo-Kantian theories of the liberal state, such as Rawls’s, is meant to guarantee a fundamental openness not only due to moral standards —such as the ones that can underline the right to liberty of consciousness or to free speech— but also due to (the lack of) epistemic legitimacy. The idea of neutrality here tries to be a negative idea, its main focus resting on restraining reason from trusting its powers completely. The embodiment of this neutrality, politically, is public reason.¹⁴⁹

¹⁴⁷ This is also the reason why Rawls defends an anti-utopian position regarding political philosophy. Although he stresses the importance of ‘ideal’ constructions, he also introduces the idea of ‘possibility’ as a value for political conceptions. John Rawls, “The Idea of Public Reason Revisited,” *The University of Chicago Law Review*, Vol. 64, No. 3 (Summer, 1997), 768-71

¹⁴⁸ “The first feature of rational intuitionism says that moral first principles and judgements, when correct, are true statements about an independent order of moral values; moreover, this order does not depend on, nor is it to be explained by, the activity of any actual (human) minds, including the activity of reason.” John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993) 91, see footnote 3.

¹⁴⁹ For the discussion on public reason I am following both *Political Liberalism* and, especially, *The Idea of Public Reason Revisited*, both *Op. Cit.*

These two facts taken together —the presence of pluralism and of the limits of reason— open the door for a new kind of legitimacy, namely, intersubjective legitimacy. To compare, this can be applied both to science and politics, but in different manners. McCarthy develops the point by separating natural sciences from human sciences, thereby placing human sciences in the same group as political conceptions and comprehensive views:

The carefully marshaling and weighing of evidence and argument for and against different positions in the pursuit of reasoned agreement remains the hallmark of the best work in history, sociology, psychology, economics, political science, and other areas of organized knowledge of the human world. It is just that the *einheitsstiftende Idee* that informs and structures it, the idea of a single objective world that is the same for all of us —as Habermas, echoing Kant, characterizes it— functions as a regulative idea in the human sciences in a way different from that in the natural sciences. In the latter, following the maxim to seek one right account of natural event is regularly rewarded with universal agreement on just such an account. In the human sciences, however, pursuit of *the* truth concerning human affairs frequently leads to disagreement that is reasonable in the sense of in line with the canons of evidence and argument proper to the discipline in question, and sometimes to disagreement about those very canons. In these domains, we rarely encounter “proof” or “demonstration” in any strictly logical sense, rather we normally have to do more or less *cogent* arguments, and cogency is a very different sort of relation than logical proof.¹⁵⁰

And I have argued elsewhere¹⁵¹ that from the pragmatist, Peircean point of view on science some moral lessons can be learned for social praxis, especially when it develops the idea of a community of inquirers. Behind all forms of falsification there is an account of fallibility, and behind all forms of social cooperation —in science and elsewhere—, there is an account of beneficial opposition and an overlapping of perspectives. For half a century now, many social and political theorists have tried to build up different approaches to this topic: theories of communication, inter-subjective cognitive processes, or social action. All of them point to the

¹⁵⁰ Thomas McCarthy, “The Burdens of Modernized Faith and Postmetaphysical Reason in Habermas’ “Unfinished Project of Enlightenment.” in *Habermas and Religion*, Ed. Craig Calhoun et al., (Cambridge, UK: Polity Press, 2013),125

¹⁵¹ Graupera, Jordi, “Dependences between logic and community: Philosophical implications of Peirce’s categories for Praxis,” *Ramon Llull Journal of Applied Ethics*, Issue 2, 179-194 (Barcelona, 2011)

contradiction that informs the idea of democracy: that there is strength to be found in the sum of weaknesses.

Key concepts in Rawls' theory of the liberal state, such as reciprocity, overlapping consensus, and, especially, reasonability deal with this same contradiction. These concepts are meant to address the problem of the limits of perspectives, and the limit of plurality. This is where science and politics diverge. A limit in scientific discourse cannot be the same as a limit in political (or moral) discourse, in the same manner as a political law is not the same as a scientific law. The tension between positivism and naturalism, when translated into the contemporary notions of *ius* positivism and *ius* naturalism, opens a new set of problems, not in the way a genus relates to species, but, rather, in the way a genus relates to another genus. Questions on what, why, where, when and how things are differ fundamentally from the question how things *should* be.¹⁵² Not all claims on nature are scientific, whereas all claims on the social fabric, even if they are scientific, are political to some extent.

Although the border between science and pseudoscience is far from settled, within science, among other standards, results and means work as facts; whereas in political and moral discourses results work as *telos* and, together with means, they form an ideology. This is the reason why neutrality in the eyes of science (or philosophy, for that matter) might be seen as different from political neutrality. It is a rudimentary version of the fact/value problematic. A neutral point between facts and interpretations is eased because although natural laws must be posed to exist—in one way or another—, scientific laws are secondary to them. A neutral point between

¹⁵² This is the perennial problem of the *sein/sollen* distinction.

goals and ideologies is burdened because political laws —especially in a democracy, and even more specifically in a liberal democracy— must exist.¹⁵³

In setting the limits of plurality, even if they are merely procedural, there must always be an assessment concerning the value of inter-subjectivity. They are mutually dependent. However, the neutral position is different in each case. The first is political, the second, at least, epistemic, if not metaphysical. The first needs to be settled to some extent (*it is necessary*), the second is an open question fundamentally, (*on necessity*). The key to any liberal conception of the state is to be able to settle the former, without settling the latter. That is, a form of political neutrality must be present without, at the same time, settling the question on necessity. There is much doubt that this can be done, but many proposals for a theory of the liberal state, like Rawls', assume that this is possible.

Rawls' public reason is a limited form of reason in many ways, but most centrally there are limitations that concern which content it must be applied to and who is required to use it. Public reason must be applied to questions of 'fundamental political justice', which Rawls defines as "constitutional essentials" and "matters of basic justice."¹⁵⁴ The former are political rights and liberties, and the deliberation guided by public reason decides whether they "may reasonably be included in a written constitution, when assuming the constitution may be interpreted by a supreme court." Matters of basic justice "relate to the basic structure of society and so would concern questions of basic economic and social justice and other things not covered by a constitution."¹⁵⁵ Likewise, public reason must be used to deliberate on these matters only by

¹⁵³ Even their absence is a form of law, as Hobbes showed.

¹⁵⁴ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 37

¹⁵⁵ This can be found both in *The Idea of Public Reason Revisited*, p767, and *Political Liberalism*, pp 227-237.

government officials—including, especially, judges—and candidates for public office.¹⁵⁶ A further limitation implies that these individuals are required to make use of public reason only in discussions “of those questions in what I refer to as the public political forum. (...): the discourse of judges in their decisions, and especially of the judges of a supreme court; the discourse of government officials, especially chief executives and legislators; and finally, the discourse of candidates for public office and their campaign managers, especially in their public oratory, party platforms, and political statements.”¹⁵⁷

However, these two limits point to two different aspects of neutrality. The limitation affecting the ‘who’ and ‘where’ of public reasoning is directed towards differentiating the sphere of the public from all other spheres, be they purely private or somewhere in between public and private, like public opinion.¹⁵⁸ Besides what has already been discussed about this distinction, practically speaking, it only demands of certain people to behave in a certain way. Admittedly, this could be required from a comprehensive point of view. For example, it could be argued that a church might ask their clergy to discuss moral matters from the pulpit making use exclusively of theological reasoning, (or even an official theoretical view, as Kant does in his *What is Enlightenment?*¹⁵⁹) In this case, plurality would be relative to a common standard, namely, theological argumentation and its sources; but intersubjective neutrality would be trespassed,

¹⁵⁶ Although later on Rawls will ask of citizens to see themselves as if they were public officials as a way of fulfilling their duty of civility regarding political decisions, this duty still maintains the difference between the public official behavior and the civil behavior.

¹⁵⁷ John Rawls, “The Idea of Public Reason Revisited,” *The University of Chicago Law Review*, Vol. 64, No. 3 (Summer, 1997), 767, see footnote 1

¹⁵⁸ Jürgen Habermas; Sara Lennox; Frank Lennox “The Public Sphere: An Encyclopedia Article” *New German Critique*, No. 3. (Autumn, 1974), pp. 49-55

¹⁵⁹ Immanuel Kant, “Answering the Question: What is Enlightenment?” (New York: Columbia University online archive), <http://www.columbia.edu/acis/ets/CCREAD/etscc/kant.html>, accessed, March 2017

precisely because this standard works as if it were an objective standpoint —embodied, ultimately, by an explicit God’s-eye view—. Similarly, in public reason, the fact that there is a limit that separates public officials and the general population, in itself only assures that plurality is not formally restricted, but it does not guarantee that there will not be an objective standard that will impose itself when this general population tries to translate its plurality into public action. It bans certain reasons from being used as coercive reasons, but it has yet to establish which reasons may count as legitimately coercive. The key, then, is if it is limited enough.

The second limitation, that is, the topics in which this public reason is allowed to have a coercive impact, is meant to address this, and points to the other kind of neutrality, i.e., inter-subjective neutrality. The problem with this second limitation is that, in the end, it must affirm something. By delimiting not the content of public reason but the content to which this public reason can and must be applied, Rawls is highlighting which aspects of a communal existence ensure the possibility of living according to one’s own cognitive stance. Necessarily, this implies a restriction on an individual-to-individual or group-to-group basis, as much as unrestricted liberty for affirming one’s view and living according to it. This a very delicate moment because it demands a standard that must appear to be, at the same time, external to any comprehensive view and internal to intersubjective legitimacy. This tension is seen clearly in the use of the words ‘essential’ and ‘justice’ in Rawls’ account. As we have seen above, however, the delimitation of the domain of essential justice, when Rawls is pressed for clarification, falls either on the side of arbitrariness (Della Rocca), natural right theory (Dworkin), or unwarranted universalism (Walzer).

Besides the ‘who’ and the ‘whither’ the other three aspects of public reason that Rawls establishes are intended to solve this riddle.¹⁶⁰ The first one is content: “its content as given by a family of reasonable political conceptions.” The key word of this aspect is ‘reasonable’, otherwise the aspect would again lapse into a vicious circle. The second one is transparent objectivization: “the application of these conceptions in discussions of coercive norms (are) to be enacted in the form of legitimate law for a democratic people.” The key to this aspect is “legitimate”, otherwise it would be a truism. The third, critical self-consciousness: “citizens checking that the principles derived from their conception of justice satisfy the criterion of reciprocity.” Here the spotlight is on ‘reciprocity,’ otherwise the work of citizens would be solipsistic.

These three concepts —reasonability, legitimacy, and reciprocity— are intimately interdependent. There is no legitimacy without reasonability, there is no reasonability without reciprocity and vice-versa, and there is no real reciprocity without legitimate procedures. Likewise, the aspects they point to are consistent with the problems that a comprehensive view shows: content, objective standards and self-consciousness. A religion, or a moral view, has a content derived from a standard claimed to be objective that can lead to an impermeable self-consciousness. What Rawls wants to find by connecting reasonability, legitimacy, and reciprocity is a content derived from a standard that does not claim to be objective in the modern way; that is, not like rational intuitionism. Rather it should bear a sort of objectivity that is just neutral, so as to produce a permeable self-consciousness. The six essentials for objectivity that he states in *Political Liberalism* are all meant to deny the ‘objective eye’ of the moderns, by affirming common grounds: frame, aim, distinctive point of views, the burden of reason, and agreement among

¹⁶⁰ For this I am following John Rawls, “The Idea of Public Reason Revisited,” *The University of Chicago Law Review*, Vol. 64, No. 3 (Summer, 1997), 765-807

reasonable agents.¹⁶¹ So, again, the aspects of reasonableness, reciprocity and legitimacy appear to inform the validity of an ‘as if’ for objective reason for politics, which finds the strength in the claim for an unepistemic neutrality.

Acting from and following the idea of public reason, according to Rawls, means explaining “to other citizens their reason for supporting fundamental political positions in terms of the political conception of justice they regard as the most reasonable.”¹⁶² It is clear that public reason does not demand, to those who are to exercise it, a justification for truth.¹⁶³ A political conception is, strictly speaking, the place for the content of justice to affirm itself. But of all political conceptions, that is, of all possible contents, the ones that count for the constitutional essentials and matters of basic justice are only those that can claim to be reasonable. Reasonableness, again, works as a substitute for truth in the sense that its presence excludes the possibility of asserting views based on truthfulness.¹⁶⁴

In this substitution, the word ‘reasonable’ points to the idea of process both in its grammatical structure and in the faculty it highlights. The suffix ‘-able’ signals a potentiality, not an actuality. As such, it could mean ‘able to think or understand,’ but then it would only refer to its active voice: one who is reasonable, and thus judges according to some logical laws. Rawls, however, uses this term also in its passive form: something is reasonable if it is able to be thought, digested through reason alone¹⁶⁵. The coincidence of these two voices brings a meaning that implies idealization. Reasonability is a feature of a conception, responsible for an expectation regarding

¹⁶¹ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 110-116

¹⁶² John Rawls, “The Idea of Public Reason Revisited,” *The University of Chicago Law Review*, Vol. 64, No. 3 (Summer, 1997), 769

¹⁶³ *Ibid.*

¹⁶⁴ Of course, one may hold a view of things that concords with the requirements of public reason, but the limited scope of the latter impedes the whole of her truth be the justification for it. It must be left aside.

¹⁶⁵ “Political constructivism specifies an idea of the reasonable and applies this idea to various subjects: conceptions and principles, judgements and grounds, persons and institutions. In each case, it must, of course, also specify criteria to judge whether the subject in question is reasonable.” John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993) 94

the reasonability of other citizens (especially if they hold a different comprehensive view). This expectation is an idealization of reason, projected onto others.

In this passage of *The Idea of Public Reason Revisited*, we find these two uses of reasonability together:

A citizen engages in public reason, then, when he or she deliberates within a framework of what he or she sincerely regards as the most reasonable political conception of justice, a conception that expresses political values that others, as free and equal citizens must also reasonably be expected reasonably to endorse ¹⁶⁶

The first usage of the word is passive, a quality of the conception; the last one is active, a quality of a subject, the one in the middle corresponds to the act of idealizing, and links the conception with the subject.

Now, the requirement for neutrality regarding a plurality of views seems to be met because the account of reason incorporates in principle the existence of the Other. Grammatically, the potentiality applied to reason is the acknowledgment of the weakness of rationality in its actuality. And therefore, is meant to moderate the act of idealization. This is further stressed by the use of the word ‘sincerely’, —an essential for political objectivity to appear— which implies willingness as a value, in order to compensate for the lack of accuracy in the projection that reason imposes on other subjects’ reason.

But, beyond grammar, the faculty highlighted by the use of ‘reasonable,’ namely, reason, is what keeps the problem unsolved. Reasoning and rationalizing are meant to be different, as we know. The implication is that there does not exist a rational political conception that is, at the same time, fully reasonable, unless it accepts as a standard the impossibility of any standard. In this account, reason is a mediating faculty that impedes the full assertion of truth. “Our aim in doing this is to express in that procedure all the relevant criteria of reasonableness and rationality that

¹⁶⁶ John Rawls, “The Idea of Public Reason Revisited,” *The University of Chicago Law Review*, Vol. 64, No. 3 (Summer, 1997), 774 . Emphasys added.

apply to principles and standards of political justice.”¹⁶⁷ This is where constructivism comes into play to assert freedom and equality as preconditions for deliberation, as we have seen above.

Being reasonable, then, means using reason and a sincere will to project an idea of the other as free and equally reasonable in asserting that their reasons and beliefs might not have anything to do with the truth. An unreasonable political conception then would be any conception that is unable to project this notion of reason onto others. This would exclude also this view: there is no political conception that one can reasonably expect others to reasonably endorse if they are free and equal. The limit of the intersubjective neutrality is the idealization of citizens as free and equal, including, both in freedom and equality, freedom of consciousness and equality of imperfection in reaching truth.

The content of public reason as expressed in reasonableness, then, is a correction not only of truth, but also of rationality. As such, it proposes a notion of the person as occupying a cognitive standpoint, in which the plurality of views is respected through the emphasis on the potentiality of reason as opposed to the actualization of rationality, but also, in which subjectivity is conditioned by the frame of freedom and equality. Subjectivity is deprived of its particularity through idealization —and by means of the ‘original position’—, and then the outcome is again a use of reasonability that unwarrantedly trespasses the kind of intersubjective neutrality that the system sets for itself. The particular risk, here, is that this imposed idealization, this content, could lead to an impermeable self-consciousness. If it does, it means there is no difference between this idealization and a comprehensive view.

To sum up, we end up with an idealization of the person through the idea of citizenship, from which citizens idealize the other as reasonable, in order to settle constitutional essentials and matters of basic justice, according to the qualities of persons as citizens, that is, as free and equal individuals, both in form and in matter. This is necessary because human political actions are imperfect, due to both the burdens of reason and the *fact* that we are not fully free and equal. So,

¹⁶⁷John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 103

what is lacking in terms of epistemic or metaphysic content —an account of freedom and equality—, is posited politically as a precondition. Rawls' liberal constructivism posits the requirement of freedom and equality to citizens to address the fact that they are not given as free and equal. These facts are what the veil of ignorance is meant to cover, so as to let the frame emerge. Reasonableness is the concept that smuggles the metaphysical content, it lets us talk about freedom and equality as if they were a neutral perspective, provided that one accepts the ideal of moral human being that it entails, before entering the conversation.

Chapter V

The Co-dependence of neutrality and Autonomy

*Plovia, aquell dia, Perquè vull!
Perquè tinc ganes que ploqués!
Sortia ella de casa. Perquè vull!
Perquè tinc ganes que sortís!
Tenia jo un paraigua. Perquè vull!
Perquè tinc ganes de tenir!
(...)
Tot era de tothom. Perquè vull!
Perquè tot és de tots!
Acabe la cançó. Perquè vull!
Tot comença en un mateix.*

Ovidi Montllor¹

Introduction

The debate around the concept of the person as it is formulated in liberal theory has defined most of the controversies between liberal authors and their critics. Intersubjectivity and reasonableness, as we have seen in chapters III and IV, among other topics, are part of this concept of the person, which is at stake when trying to assess the cogency of liberal neutrality. Within these debates on the person, the notion of autonomy is probably the most controversial. As John Christman puts it, “it is assumed that the person so conceived be able to reflect rationally on her desires, character, values and commitments, and revise them in light of such reflection.”² Part of the controversy has to do with the fact that if the “fundamental value to be assumed and

¹ Song by Ovidi Montllor. “It rained that day, because I want to!/ Because I feel like it should! She left the house, because I want her to! / “I had an umbrella because I want one! Because I feel like I should! (...) Everything belonged to everyone. Because I want it to!/ Because everything belongs to everyone!/ I finish the song because I want to!/ Everything begins in oneself.”

² John Christman, “Liberalism, Autonomy, and Self-determination” in *Social Theory and Practice*, Vol. 27, No. 2 (April 2001), 185

protected in a just society is the autonomy of the person,”³ then it is not clear to what extent liberalism accomplished its mission of being “value neutral.”⁴

This explains why the relationship between autonomy and neutrality is different than that of neutrality and intersubjectivity or reasonableness. While intersubjectivity and reasonableness, in diverse forms and intensities, are necessary concepts to explain and justify neutrality and the exclusion of metaphysics, (even if they fail in doing so, as I claim), autonomy is an odd guest in the conceptual apparatus. The reason, as partially explained in chapter II, is that for many neutralist authors, autonomy, even circumscribed to the political realm, is a metaphysical idea (or, at least, presupposes some metaphysical claims). Some authors, such as Brian Barry,⁵ reject it altogether precisely in order to be consistent with neutrality. Others, such as Ackerman,⁶ believe that neutrality is unintelligible without autonomy, and a third kind, Rawls,⁷ Rostbøll,⁸ or, more critically, Larmore,⁹ try to develop an idea of autonomy that is only political and “practical”¹⁰ and not metaphysical, in the familiar move that we saw throughout chapter III. Nagel stands out in his effort to problematize the relationship between neutrality and autonomy, as I will show below. There is a fourth category of liberal thinkers, relevant to our purposes: perfectionists, such

³*Ibid.*

⁴ *Ibid.*, 187

⁵ Brian Barry, *Justice as Impartiality* (Oxford: Oxford University Press, 1995)

⁶ Bruce Ackerman, *Social Justice in the Liberal State* (New Haven, Conn: Yale University Press, 1980)

⁷ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), *Justice as Fairness. A restatement*, (Cambridge: The Belknap Press of Harvard University Press, 2001), and “Kantian Constructivism in Moral Theory” *The Journal of Philosophy*, Vol. 77, No. 9, pp. 515-572 (Sep. 9, 1980)

⁸ Christian F. Rostbøll, “Kantian Autonomy and Political Liberalism” *Social Theory and Practice*, Vol. 37, No. 3 (July 2011), pp. 341-364

⁹ LARMORE, Charles, *Patterns of Moral Complexity* (Cambridge: Cambridge University Press, 1987, and *Equality and Partiality*. (Oxford: Oxford University Press, 1991)

¹⁰ Rostbøll, *Op. Cit.*, *passim*

as Raz.¹¹ They reject neutrality because they defend autonomy. So at the two ends of this spectrum, there is agreement: Raz and Barry agree that neutrality and autonomy are incompatible, whereas, to a different extent, Rawls, Ackerman, Nagel, Rostbøll, and Larmore, believe in some degree of compatibility.

My aim in this chapter is to show how, through the explicit or implicit endorsement of autonomy, neutralist liberals allow themselves to protect a conception of the self that is metaphysically committed. To a certain extent, to say that autonomy is a metaphysical idea is uncontroversial, even though some authors, such as Christman and Rostbøll, to name two recent takes on it, would reject that identification —and I will try to debunk their views. But my main task is to show how even when neutralist authors reject autonomy or restrict its meaning to the political domain in order to avoid metaphysics, they are still holding on to an idea of the self that is metaphysically committed. The first epigraph of this chapter will be devoted to a general overview of the co-dependent relationship between neutrality and autonomy. In section 2, I will put forth a historical excursus to show the Kantian standard that operates in the background of the liberal notion of autonomy, and then, via Levinas’s essay on the “philosophy of Hitlerism,”¹² I will show the theological import of the idea of freedom in that background both before and after WWII. The following three sections will be devoted to three different takes by liberal authors on the relationship of neutrality and autonomy: Raz and Barry will be first, since they agree on their incompatibility. Secondly, Rawls, Larmore and, partially, Rostbøll. And in the third place, Nagel, as an example of the paradox that I point out when one attempts to defend both

¹¹ Joseph Raz, *The morality of freedom* (Oxford: Clarendon Press, 1988)

¹² Together with Simon Critchley’s interpretation of it in “Levinas and Hitlerism” *Graduate Faculty Philosophy Journal*, Num 1., (2014), 223-249

notions. Finally, the 6th, and last, section will be dealing with the metaphysical implications of the rest of the chapter, by means of explaining a disagreement I have with Markus Gabriel's take on "liberal naturalism."

1. neutrality and Autonomy in the Political Domain

neutrality and autonomy have a mutually dependent relationship. One presupposes the other, and vice-versa. In pre-philosophical reflection, it is already conceivable that neutrality is thought to be necessary so as to protect the freedom of each individual to pursue his own life-plan, according to his values. In other words, neutrality is a response to the fact of plurality of world-views and so-called conceptions of the good.

But the most interesting perspective is not from the fact of plurality to the necessity of neutrality, but the other way around: from the perspective of neutrality towards the fact of diversity. The neutral standpoint is a sort of "view from nowhere":¹³ it is meant to be devoid of any value, except for an equal regard for each of the reasonable conceptions of the good that constitute the fact of plurality. The point of neutrality is to remain silent with respect to the value of the content of each conception of the good, except for their reasonableness, as we have seen in Chapter IV. These diverging views are regarded formally. Their pertinent characteristic is precisely that they are held by particular individuals—an individual here functionally meaning the unit of consent and dissent that neutrality articulates, even in the case of individuals who can be grouped together according to their common views.

¹³ I take this expression from Thomas Nagel, *The View From Nowhere*. (Oxford: Oxford University Press, 1986)

neutrality also articulates the encounter between these different views. Each of them must accept the presence of the neutral perspective as the horizon of the public sphere, the point from which all reasonable views are seen as equally legitimate. So neutrality demands of these views that they adapt to the formal, third-order value of neutrality: as will be explained below, the good is the first-order, autonomy is the second-order, and neutrality the third order value. neutrality thus demands that they be willing to compromise to a certain extent, so as to find policies and principles that are somewhat neutral instead of promoting sectarian policies that embody their idea of the good, whatever it may be.

This demand implies that one can somehow take a step back, and get some distance from one's own conception of the good. One's judgment can operate with one's 'conception', revise it, and adapt it to the needs of neutrality, and still appreciate the fruits of the system that neutrality articulates. From the perspective of neutrality, individuals are seen as being capable of rationally dealing with the content of their beliefs and the commitment with their beliefs, and then adapt the content and the commitment when in the public domain. This capacity is what all individuals hold equally, regardless of their conception of the good or comprehensive view.¹⁴

In order for this capacity to make any sense, one must go beyond asserting the existence of the fact of plurality. One must affirm with equal force the fact of autonomy, that is, that whatever the cultural or historical origin of one's comprehensive view, there remains the possibility of acting upon it, in the inner workings of the self, and adapting it, negotiating, compromising, revising it, changing it, rejecting it and so forth in an independent way. This is the point of autonomy:

¹⁴ Or, or relatedly, individuals have the capacity to act from a principle of justice.

independent self-creation, the belief that one's self is stronger than all circumstances, at least partially, and the belief that self is the origin of moral, political, and existential choices.¹⁵ There is one exception to the dominion of autonomy over circumstances and here is where the mutual dependency outcrops: the only circumstance that is conceived as having to be as strong as that capacity to overpower circumstances is neutrality. neutrality forces each member to enter the public domain rejecting the absolutist quality of their belief system, and then, from that same autonomy, give consent to the system of justice that is neutral.

One way to argue for this co-dependency would be to say that autonomously, one can only accept this constraint; that every decision not to accept it isn't autonomous but obeying foreign irrational decisions. neutrality is the constraint that autonomy demands, autonomy is the good that neutrality makes room for. neutrality thus doesn't force anything but what autonomy needs. Either it is necessitated through autonomy; or it is not recognized by the autonomous subject at all.

This is further stressed by the only other constraint that neutrality puts in place: reasonableness. As we saw in chapter IV, in reasonableness we see the demand to justify to others our views, to take them as equally deserving such an explanation, and, sometimes, as in Rawls's public reason, to justify our views following certain rules of theoretical reason and accepting (the typically post-modern) limits to our capacity for judgment. Being reasonable, among other things already discussed, presupposes that one can have the kind of relationship with one's conceptions of the good proper of autonomy: an ability to rationally construe reasons for one's adherence to an idea

¹⁵ I will explore this formulation of the belief in autonomy further below in 5.5: "Nagel's paradox."

of the good, and the ability to express them in a coherent, fallibilistic way. And then, to act from it. The assumption is of a self that can originate the claims and give oneself and others rational motives to pursue whatever their comprehensive views happen to demand, and act from them.

In the absence of these qualities —proper to autonomy— neutrality does not make any sense. If we are not capable of making moral and political claims, and ordering them according to an overarching, comprehensive view, if we are not able to separate our reason from the content of our belief so as to negotiate it in the relevant domain, if we are not in the position to revise our commitments to our beliefs and eventually modify them; finally, if we are not the kind of being that can rationally explain our motivations, then in the face of plurality of views, neutrality cannot provide an effective solution. If our comprehensive views were always traced back to some act of indoctrination, or if our commitments to a comprehensive view were determined by, say, our genes, our pathological impulses, or an evil deceiver, then the entire edifice that is safeguarded by neutrality would not stand, and maybe we would be arguing for and against, say, a segregationist system. Each comprehensive view would be but a symptom of some determined, inalterable forces, and even if they were somehow incompatible, maybe living apart would be the only debatable option. Maybe we would appeal to a libertarian model of government: the state should not intervene at all, since it's either useless or an anti-natural form of oppression. Or we would set out to liberate ourselves from 'false consciousness:' we would strive to substitute a new, truer view of the world for our old, determined, and false views, because we would not be able to change them with our own internal forces, but the truth could be brought to us from the outside. Or we would long for autonomy.

neutrality presupposes autonomy, and only because there is autonomy, can we understand second- and third-order values. Autonomy is a second-order value because it is empty of content: it only tells us the origin and the form of our relationship with our beliefs, that is, we are the independent creators of them, and they respond to reasons, however imperfect.¹⁶ That very distinction between the content of our beliefs and the holder of those beliefs, namely, the rational agent, already speaks of a very specific anthropology with metaphysical presuppositions lurking below the surface (more below). Autonomy as a second-order value shapes the form of our conceptions of the good (the name itself ‘conceptions of the good’ speaks volumes: the person conceives the good, which is a concept). But neutrality is a third-order value: it shapes autonomy. It explains our relationship with our commitments by means of a further idealization of their form, it takes them to be autonomous in the strong sense. We construe an idea of justice out of our idea of autonomy and the capabilities that it assumes autonomous agents to possess.

If the first main argument in this chapter aims to show the necessary relationship between neutrality and autonomy, the second, subordinate, argument shows how the authors that claim to reject autonomy, or to circumscribe it in the political domain in order to deflate it, are in fact endorsing it under a different name or by means of a roundabout maneuver. The third main argument aims at showing that that dependency on autonomy implies metaphysical commitments. But before all that, since different authors have different takes on the metaphysical meaning of autonomy, and express different intensities of agreement with Kant, (which establishes their level of commitment to autonomy), let us first consider the modern

¹⁶ Barry, *Op. Cit.*, 159

understanding of autonomy, via Kant and Levinas, to lay out the metaphysical stakes of the conversation.

2. Modern autonomy: Kant, Levinas, and Post-metaphysics.

2.a Henrich and Korsgaard on Kant's Pure Reason

It is commonplace to say that the cornerstone of Kant's moral philosophy is the autonomy of reason.¹⁷ The Kantian meaning of autonomy is related to his overarching idea of the pure reason and the limits of metaphysics: reason is autonomous in the sense in which it can criticize 'pure' reason. As Dieter Henrich notes: "reason is only autonomous when it 'purely', by itself alone, and without regard to grounds or drives distinct from it, suffices to determine the will."¹⁸

If in the first *Critique* the pure theoretical function of reason was interrogated,¹⁹ in the second it is its practical function that is examined²⁰. In the former, its pure-theoretical use was declared illegitimate, because it was transcendent, meaning beyond its boundaries;²¹ but in the second *Critique*, the categorical imperative that was deduced from the concept of the will and of duty in the *Groundwork* is applied without any reference to the embodied subject. It manages to be a

¹⁷ Besides Kant himself, there are several authors, textbooks, encyclopedia articles that state in similar ways this insight, which will be in the background of this sections, but for this explanation I am going to follow, in particular, Dieter Henrich *The Unity of Reason, Essays on Kant's Philosophy, essay: Ethics of autonomy*, Trans Louis Huny, (Cambridge MA: Harvard U press, 1994) and my own notes on Agnes Heller's 2009 seminar on Kant's Ethics at the Philosophy department of the New School for Social Research.

¹⁸ *Ibid.*, 94-95

¹⁹ Hereafter all internal citations are to the English translation taken from Immanuel Kant, *Critique of Pure Reason*, translated and edited by Werner S. Pluhar (Cambridge: Hackett Publishing Company, Inc., 1996)

²⁰ Hereafter all internal citations are to the English translation taken from Immanuel Kant, *Critique of Practical Reason*, translated and edited by Mary Gregor (Cambridge, Cambridge University Press, 1997)

²¹ Kant says that the use of reason for objects that transcend the bounds of possible experience is illegitimate (in other words, a priori synthetic judgments), but the precise reason for declaring transcendent reason illegitimate is because there is an insoluble antinomy when reason makes claims on the "unconditioned."

causality beyond the domain of its a priori concepts. If there exists such a thing named pure practical reason, it does not require a critique. As long as it is a pure reason, it contains already the complete rules for its use. According to Henrich, there are two reasons “for reason to be determinative of the will purely by itself....It must ground action in two senses.”²² The first is:

Reason must contain principles of action which state *what* the will wills. It must designate specific actions as rational without appealing to grounds other than those which result from the structure of reason itself. Reason is autonomous only if it can supply a ground for knowing the right will before it has acquired knowledge of the structure of being or the fundamental tendencies of action of human nature. There must be a *principium diiudicationis bonitatis* inherent in the form of reason.

This first *principium* establishes also a form of self-consciousness. The process through which the *what* of what the will wills is taken as ground presupposes the interplay between the faculty of reason and the content of the will, in which different options can be weighed. As Christine Korsgaard states: “The source of Reason is a particular form of self-consciousness that characterizes the human mind. Human beings are conscious of the potential grounds of our beliefs and actions *as potential grounds*.”²³ When reason designates specific actions as rational without appealing to anything outside itself, a “space of awareness, a “reflective distance” is opened up...

...between the potential ground of a belief and the belief itself, or between the potential ground of an action and the action itself, we must step across that distance with some awareness that we are doing so, and so must be able to endorse the operation of that ground as the basis for what we believe or do.²⁴

²² Dieter Henrich, *Op. Cit.*, 94-95

²³ Christine M. Korsgaard, *The Constitution of Agency, Essays on Practical Reason and Moral Psychology*, (Oxford: Oxford University Press, 2008.), 5

²⁴ *Ibid.*, 4. This awareness, Korsgaard believes, “is the source of Reason.” (*ibid.*)

This already shows something that I will develop later (in section 3): the attempt at grounding autonomy on just the practical domain, as if it did not entail a certain understanding of the self, as Rostbøll, Larmore or Rawls attempt, is flawed. Rostbøll, who in his attempt at salvaging autonomy for liberalism, —and making it compatible with neutrality,— offers an understanding of autonomy as merely “instrumental” in treating others according to the substantive value of equal respect. He claims that this view does not entail metaphysical commitments, because “the important Kantian suggestion is that the autonomy we should respect does not belong to what there is, to what can be explained empirically or metaphysically, or to what can be known, but rather to the practical standpoint from which we ask what we should do and how to treat ourselves and others.”²⁵ In this account we treat others as *noumena*, but only as an instrument for practical reason. Rostbøll seems not to be aware that this treatment of others and of oneself as *noumena* presupposes the entire edifice of the *purity* with which reason designates the content of the will as rational, independently of external contingencies.

Korsgaard’s and Henrich’s assertions open for practical reason the possibility of introducing itself in the noumenal world without going beyond its own sphere of “reality”. However, it still does not show clearly whether it can move beyond the noumenal domain. What the transcendental ideas lack when they are to be conceived is the possibility of coming under the form of an intuition. There can be no sense-experience of a transcendental idea. However, what was forbidden for pure theoretical reason might now be allowed for practical reason. The concept of praxis might be the means by which transcendental idea, which is a merely logical *possibility*, becomes actual. Transcendental freedom might become actual by the immediate nature of

²⁵ Rostbøll, *Op. Cit.*, 351

action, and therefore it may as well be conceived not hypothetically but assertorically, in its intelligible, noumenal dimension.

2.b Kant's Noumenal Causality

For this to be possible, it is necessary that the super-sensible dimension of rational beings alone affects the sensible domain by which and in which they act. But, how is it possible that a moral law (universal and created by reason alone) causally affects the empirical realm of actions? Even more: how can reason be, in fact, the legislator of the behavior in rational beings if they remain subject to the causality of nature? Here Kant claims to have uncovered another kind of causality. As he states in the first Critique:

Only two kinds of causality can be conceived in regard to what occurs, viz., either a causality according to nature or one from freedom. The causality according to nature is the connection, in the world of sense, of one state with a previous state upon which the state follows according to a rule. Now the causality of appearances rests on conditions of time... By freedom, on the other hand, in the cosmological sense of the term, I mean the power to begin a state on one's own. (A533/ B561)

Kant seeks validity for a freedom by which the law may be determining. He has to find a law-abiding will capable of following universal maxims and of affecting life independent of natural causes. Moreover, “this (moral) law is to furnish the sensible world, as a sensible nature (in what concerns rational beings), with the form of a world of the understanding, that is, super-sensible nature, though without infringing upon the mechanism of the former.”²⁶

²⁶ (5:43)

And yet, with the discovery of autonomy,²⁷ human consciousness is aware of itself, in “another side, as a being in itself –conscious of its existence as determinable in an intelligible order of things –and conscious of this not, indeed, by special self-intuition but according to certain dynamic laws that can determine its causality in the sensible world.”²⁸ The causality that comes from the fact that pure practical reason can produce an effect on the will, can be thought as able to determine the will absolutely, precisely because the grounds for belief and action are deemed as rational and good — even if Kant denies that we can ever know it. This is why Henrich states the second condition for autonomy thus:

Even if reason were capable of this [designate the *what* of the will purely J.G.] , it could still lack autonomy —if it were not at the same time that which gives binding force to the “rectitudo” of the will. Were reason only to know the right, without at the same time being the ground of its obligatory power, then it would indeed be the faculty through which we came to know the right, but not its origin. There would be *autognosy* of reason without regard to rational action, but in no sense an *autonomy* of reason. Autonomy requires that reason have the power to effect actions which take place solely because they are rational. Just as the insight of someone acting on the basis of autonomous reason is not directed to facts distinct from reason itself, the energy of his will does not derive from sources other than this insight itself. He does what is right, not on the basis of an interest which his insight serves, but for the sake of the being of the right into which he has insight. Reason must therefore be supplied with a *principium executionis bonitatis*.²⁹

The binding force and effectiveness of this *principium executionis bonitatis* is explained in the dialectical relationship between freedom —this metaphysical idea—, and the moral law. The *ratio cognoscendi* of freedom is the moral law, and Kant’s argumentation goes as follows: “pure reason can be practical —that is, can of itself, independently of anything empirical, determine the will –

²⁷ With the presence of an internal moral law, with the strength of the categorical imperative, its rational necessity, its independence of any matter, and also with the autonomy of the pure will.

²⁸ (5:42) The noumenal/phenomenal split within the self is the price of freedom, as it were. If this division were to collapse, we would not be able to talk of a causality of freedom, independent of the causality of nature. Recall that Kant had to limit knowledge in order to make room for faith (and freedom).

²⁹ Dieter Henrich, *Op. Cit.*, 94-95

and it does so by a fact in which pure reason in us proves itself actually practical, namely autonomy in the principle of morality by which reason determines the will to deeds”.³⁰

The solution to this problem implies the warrant to understand the ends of reason metaphysically, but only for practical purposes, not theoretical. This allows him to supply a deeper understanding of the distinction between phenomenon and noumenon, in order to solve the contradiction between natural necessity and freedom. The law of causality is applied to things insofar their existence is determined in time and space. If there were no other way to represent the existence of things, or nothing that could be represented differently, it would be necessary to reject freedom as an impossible idea.

Consequently, the consciousness of this other causality, if it is to be shown as both necessary and existent, must come from beyond natural causality, meaning from another realm besides the phenomenal, namely, the noumenal. “For insofar as this subject is noumenon, nothing occurs in it and there is found no change requiring dynamical time determination and hence no connection with appearances as causes” (A541/ B569)

The necessary relationship between this metaphysics of reason and autonomy is apparent. For Kant, if we did not have such a double perspective from which we could observe human beings and their actions; in other words, if the human being existed only as a being in itself, and was incapable of determining his actions and his self *qua* phenomenon, there would be no room to

³⁰ (5:42) It is in this context that Kant offers a platonic analogy. The super-sensible nature can be thought of as if it were “natura archetypa” ruled by the law of reason, whereas sensible nature can be expressed as a “natura ectypa”, a copy of the former. This distinction means that the moral law can be understood as a kind of natural law and, yet, at the same time, it cannot be a subject of such a natural law. Kant is here trying to state the validity of the idea of freedom in its practical use. For that he needs to find not only a type of maxims that challenge sensible inclinations and grounded independently of all empirical condition, but also to justify a power -freedom of the will- that must be beyond the type of causality by which a specific kind of relation between phenomena is explained. Therefore, this pure practical lawgiving faculty cannot be grounded in any phenomenal nature.

speak of a causality of freedom either. Time and space would be determinations belonging to things in themselves, instead of pure intuitions of the human sensibility, and transcendental fatalism would be unavoidable. Placed by God in a world determined by natural causality as a link in a chain of causes – in which the creator itself would appear as the first ring – human beings could not envision more freedom than that of a reasoning automaton. The only singularity rational beings would have, compared to all other things in nature, would be their consciousness of such an external law. Heteronomy would be the name of their practical realm.

Conversely, if human beings were only understood as *noumena*, and their will could not affect the phenomenal sensible world, there would not be any freedom either.³¹ What is necessary for this *causa noumenon* to exist is a conception of human beings according to which their noumenal selves can be responsible for their phenomenal selves. Both sides of the self are necessary to understand how transcendental freedom works as an unconditioned causation of the will. Kant states:

How a *causa noumenon* is possible; this it [reason] cannot do, but as practical it does not even concern itself with this inasmuch as it only puts the determining ground of the causality of the human being as a sensible being (which is given) in pure reason (which is therefore called practical), and accordingly uses the concept of cause itself -from whose application to objects for theoretical cognition it can here abstract altogether (since this concept is always found a priori in the understanding, even independently of any intuition) not in order to cognize objects but to determine causality with respect to objects in general, and so for none other than practical purpose; and thus it can transfer the determining ground of the will into the intelligible order of things inasmuch as it readily admits at the same time that it does not understand how the concept of cause might be determined for cognition of these things.³²

³¹ Or at least, it would be freedom in a vacuum. Not “the causality of freedom,” whereby acts, unconditioned by natural causes, initiate a casual chain.

³² (5:46)

The fact of the moral law, its presence in reason, as well as its complete independence of empirical conditions, shows us that freedom exists. The idea of this other causality, of this other order, itself different from the natural order, allows us to deduce the reality of freedom. Kant has unveiled through this deduction the possibility of the immanent reality of this transcendental idea, even if the attempts to deduce freedom failed. Through practical reason, freedom is articulated positively insofar as it is proved that it can determine the will: “The moral law...is able for the first time to give objective thought only practical reality to reason, which always became extravagant when it wanted to proceed speculatively with its ideas, and changes its transcendent use into an immanent use (in which reason is by means of ideas itself and efficient cause in the field of experience).”³³

Now the door is open to apply the concept of causality to super-sensible beings, as long as it has something to do with the moral law, the free will, transcendental freedom and practical reason. Kant is proud of being authorized to do so by virtue of a “pure, not empirical origin of the concept of cause, inasmuch as I consider myself authorized to make no other use of it than with regard to the moral law which determines its reality, that is, only a practical use”³⁴ So, everything, even super-sensible nature, is subject to causality, what matters here is its source. He insists once and again in the idea of the emptiness of this *causa nomenon* in a theoretical sense, but it nonetheless has a “real application which is exhibited in concreto in dispositions or maxims, that

³³ (5:48) Kant has established the validity of an objective moral principle with respect to its causality and by means of a law that cannot be regarded as natural. Hence, he claims to have extended our cognition beyond the boundaries of the experience, inasmuch we can design positively freedom, but not what it is, or what its content is, since we do not have any intuition of it. He will not claim to have discovered a way to prove the actuality of freedom in all ways, as an existing object, but only as a transcendental ground for practical purposes, that is, as an unmoved mover of human will.

³⁴ (5:56)

is, it has practical reality which can be specified; and this is sufficient to justify it even with regard to noumena.”³⁵

This distinction will open the possibility for grounding his metaphysics of morals. A grounding of a metaphysical system through the moral law that seemed to be lost forever at the end of the first *Critique*. Through the securing of the concept of *causa noumenon* he has created a possibility for the securing of all other transcendental categories and, by analogy, of all super-sensible beings and transcendental ideas including God and immortality; that is, a securing of a metaphysics of morals.

This explains something about the nature of liberalism and its relationship with metaphysics, which is the central focus of our discussion. The idea of freedom that Kant secures for the moral sphere, and that is so closely tied with the capacity of the agent to rationally pursue the good, is not only a discovery of a different causality, it is the escape from natural determinism, or natural reductionism. At the same time, it shows how the only way to secure freedom, once the presence of natural science and the blows of skepticisms are taken into account, is the development of a metaphysics of reason for practical purposes only, that can explain our moral choices as something that happens rationally outside the back-and-forths of history and circumstance, as much as outside the force of nature. The idea of freedom that is present in the accounts of autonomy that neutralist authors defend continues the strand that Kant initiated by putting limits on our theoretical commitments to metaphysics. But since even the noumenal nature of the self, and the metaphysics of ends, cannot be acknowledged by these contemporary authors, they need another strategy, another sphere where this freedom can be effective and intelligible. They will

³⁵ (5:56)

call it the sphere of the political and will posit the capacity of autonomy as a normative center of it. This move conceals, I claim, a similar type of metaphysics as the noumenal world of Kant when it wants to be a cause in the phenomenal world.

2.c Emmanuel Levinas: Autonomy and the Theology of Freedom

Emmanuel Levinas believes that the metaphysical foundation of liberalism is expressed in the power that freedom, understood as autonomy, is presumed to have over nature, history, and social contingencies. Levinas identifies this metaphysical aspect as a secular translation of the “Judeo-Christian” understanding of the soul, to which the “philosophy of Hitlerism,” with its anchoring in the destiny inscribed in the body (race), and to a lesser extent, Marxism, with its emphasis in material determination,³⁶ are a rebuke.³⁷

As opposed to the Greek fate, or the Heraclitean irremediable flow of things, in Christianity, the past is subordinate to the present through remorse, forgiveness, and redemption: “through this, Christianity proclaims freedom and makes such freedom fully possible.”³⁸

This freedom is not only spiritual or religious, it implies a political framework that links the idea of the sacred covenant with God with that of the social contract. This is relevant for two reasons.

³⁶ “If the basic intuition of Marxism consists in perceiving the spirit to have a inevitable relation to a determined situation, this link is in no way a radical one. Individual consciousness determined by being is not sufficiently impotent not to retain, at least in principle, the power to shake off the social bewitchment that then appears foreign to its essence. To become conscious of one’s social situation is, even for Marx, to free oneself of the fatalism entailed by that situation.” Emmanuel Levinas, “Reflections on the Philosophy of Hitlerism,” trans. Seán Hand, *Critical Inquiry* 17:1 (1990) 67

³⁷ As Simon Chritchley explains in his “Levinas and Hitlerism,” *Op. Cit.*: “The term “Hitlerism” is not Levinas’ own. It is a term that was imposed by the editors of the journal *Esprit*.”, 233

³⁸ Levinas, *Op. Cit.*, 65

First, because liberal neutralists see neutrality as a tool not only to respect and articulate different comprehensive views, they also see that respect as a means to protect the possibility of (rationally) revising and changing one's system of beliefs. This capacity for revision is the point of being taken as a 'citizen' and not a person or a member of a group, to obtain rights. Citizens get their dignity independently of their beliefs, or their changes of heart with respect to those beliefs, as long as they are reasonable (more on this below). And, as we saw in chapter IV, by being reasonable, liberalism also understands both the possibility of revising and changing one's beliefs, and, also, having a detached relationship with one's comprehensive view, for political purposes.

Second, because the social contract expresses an inherent contingency: a contract can, in theory, be dissolved, as even Hobbes recognizes. In truth, the understanding of monarchy as a covenant, and the possibility of breaking that covenant for moral reasons is present in many Christian theologians, Thomas Aquinas paradigmatically, in both his *Summa Theologica* and, less so, in *On Kinship*.³⁹ For Levinas, these two characteristics, individual revision and dissolution of covenants, are precisely the core content of the supernatural idea of freedom in Jewish and Christian theology.

Not only is the choice of destiny a free one. Once the choice is made, it does not form a chain. Man retains the possibility —a supernatural possibility, certainly, but also a concrete and graspable one— of terminating the contract into which he freely entered. At any moment he can regain the nudity he had during the first days of creation.⁴⁰

The key of the Christian understanding of the soul is the contrast between the contingency of the world and its bodies, and the eternal nature of the soul. Freedom is at the “base” of that

³⁹ See PAul E. Sigmund (Ed and Trans), *St. Thomas Aquinas on Politics and Ethics*, (New York: W. W. Norton, 1988), p 22 and 58.

⁴⁰ Levinas, *Op. Cit.*, 64

notion of the soul, and it is “infinite with regard to any attachment and through which no attachment is ultimately definitive.”⁴¹ That contrast is “the ultimate foundation of the individual,” and the “purity” of that freedom “comes from a transcendent inspiration.”⁴² Simon Critchley has placed this analysis in the broader context of Levinas’s work, in which Levinas looks for a new path that can escape the entrapment of idealism without giving in to the particular kind of Heideggerian facticity that can be used to ground a “philosophy of Hitlerism.” In the context of this essay, the underlying question is, as Critchley puts it, “how might one respond to Hitlerism beyond liberalism?”⁴³ His analysis and critique of Hitlerism is thus predicated upon the recognition that National Socialism’s appeal in its time stems from a rejection of the flaws of liberalism. These flaws are the expression of the limitations of the idealist tradition that liberalism embodies politically. Critchley states:

If what Levinas is always criticizing is an idealist notion of the subject, namely the subject as the sameness of mind and world, the subject as self-constituting, self-positing, self-legislating, and constituted by reflection, then the political analogue to idealism is liberalism.⁴⁴

From this perspective, Luther’s understanding of the indestructible freedom of belief⁴⁵, Descartes’ understanding of the infinite in the *Meditations*, or the radical freedom that the *cogito* can be interpreted to contain,⁴⁶ Spinoza’s assertion that it is impossible for the government to

⁴¹ Levinas, *Op. Cit.*, 65

⁴² Levinas, *Op. Cit.*, 65

⁴³ Critchley, *Op. Cit.*, 224

⁴⁴ *Ibid.*

⁴⁵ See Martin Luther, “The Freedom of the Christian” in Martin Luther, John F. Thornton, and Susan B. Varenne, *Faith and freedom: an invitation to the writings of Martin Luther* (New York: Vintage, 2002), 3-17

⁴⁶ As Omri Boehm does.

change one's beliefs,⁴⁷ and ultimately, Kant's understanding of autonomy as the cornerstone of his moral and political philosophy, all form a metaphysical history of the idea of freedom in liberalism, in a process of de-sacralization that has continued until its current post-metaphysical claims. From that idea of freedom come this idea of autonomy and the debates it causes around the problematic of its metaphysical content. Levinas states it thus:

If the liberalism of these last few centuries evades the dramatic aspects of such a liberation, it does retain one of its essential elements in the form of the sovereign freedom of reason. The whole philosophical and political thought of modern times tends to place the human spirit on a plane that is superior to reality, and so creates a gulf between man and the world. It makes it impossible to apply the categories of the physical world to the spirituality of reason, and so locates the ultimate foundation of the spirit outside the brutal world and the implacable history of concrete existence. It replaces the blind world of common sense with the world rebuilt by idealist philosophy, one that is steeped in reason and subject to reason. In place of liberation through grace there is autonomy, but the Judeo-Christian leitmotif of freedom pervades autonomy.⁴⁸

It should be noted that this secularization of spiritual freedom and its transformation into the sovereign freedom of reason involves in a reshaping of the reasons available to the subject, meaning they become *choices*. As reasons to be chosen, they express a logical space in which both the options and the self that weighs them share a particular structure that conforms to rationality. This structure is the core of the human mind, and since this core is sovereign, it actually defines the essence of human nature. As Korsgaard puts it:

This means that the space of reflective distance presents us with both the possibility and the necessity of exerting a kind of control over our beliefs and actions that the other animals probably do not have. We are active, self-directing, with respect to our beliefs and actions to a greater extent than they are. And it is the same fact that we now both can have, and absolutely require, *reasons* to believe and act as we do.⁴⁹

⁴⁷ Baruch Spinoza, *Theologico-Political Treatise*, trans. R. H. M. Elwes, (New York: Barnes & Noble Inc., 2009), chap XX

⁴⁸ Levinas, *Op. Cit.*, 68

⁴⁹ Korsgaard, *Op. Cit.*, 4-5

That essence not only constitutes the unique and distinct ontology of what it means to be human, it also represents the power of women and men over the world. The modern *topos* of the mastering of nature, both in the sense of natural world of things outside and the nature of our impulses and lesser forces of our self, is shining here in a subtle way. It presupposes thus the separation between subject and object, self and world, inner and outer, spiritual and material, transcendent and contingent, that articulates the modern view.

In the world of liberalism, man is not weighed down by a History in choosing his destiny. He does not experience the possibilities open to him as a series of restless powers that seethe within him and already push him down a determined path. For him, they are only logical possibilities that present themselves to a dispassionate reason that makes choices while forever keeping its distance.⁵⁰

As we will see in detail below, one of the most important connections that so-called post-metaphysical liberals try to argue for is the one that makes autonomy, or one of its surrogates, the formal precondition for the substantive moral content that articulates the system: equal respect. The fact of autonomy, understood as that capacity of abstracting oneself from social and historical circumstances, but also from one's commitments, through rational revision and choice, is what makes every individual a subject of equal dignity. The shared human nature that makes us equal is the power of the self to move beyond group identification, for being responsible for one's beliefs and actions.

In an analogous way in which the Christian act of confession acknowledges that the determining factor of one's actions is the will of the self, and in this way this self becomes morally powerful and the circumstances of an action become secondary and contingent, the possibility of being autonomous is what makes the space in which individuals deal with their political arrangements a

⁵⁰ Levinas, *Op. Cit.*, 66

moral space where the dignity of each self becomes non-negotiable. We all deserve equal respect because we have the power to be moral through reason. If moral maturity was a question of getting it right, respect would depend on content in the same way that, in today's liberal systems, being mentally immature due to young age, mental illness, or constitution, implies a reduction of political rights, voting in particular. Similarly, under the worldview of any sort of determinist person —Levinas uses the examples of race-based determinism and materialism—, responsibility diminishes or disappears altogether, and with it, all possibility of moral progress, compromise, and agreement. Individuals cease to exist as subjects of dignity, and without that dignity, respect goes as well. Levinas links this respect with the spiritual power to regain virginity that comes from the understanding of the self as *noumenon*:

Throughout the vicissitudes of the world's real history, the power of renewal gives the soul a noumenal nature that is protected from the attacks launched by a world in which concrete man nonetheless is placed. This is only apparently a paradox. The soul's detachment is not an abstract state; it is the concrete and positive power to become detached and abstract. The equal dignity of each and every soul, which is independent of the material or social conditions of people, does not flow from a theory that affirms, beneath individual differences, an analogy based on a 'psychological constitution.' It is due to the power given to the soul to free itself from *what has been*, from everything that linked it to something or engaged it with something [*engage*], so it can regain its first virginity.⁵¹

Levinas's text is from 1934, and however prescient it was with respect to “the philosophy of Hitlerism” and its anti-liberalism, it did not entirely foresee the evolution of liberalism that the destruction brought about by Nazism (and totalitarianism in general) would produce. If the first theorizations of liberalism can be linked to the need for tolerance caused by the wars of religion of Europe's seventeenth century and the general topic of religious persecution, as it is time and

⁵¹ Levinas, *Op. Cit.*, 66

again made explicit in the works of today's liberals,⁵² the post-metaphysical or anti-metaphysical strand of liberalism that the liberal neutralists have theorized since the second half of the twentieth century cannot be separated, it seems, from the political and philosophical questions that arise after the horrors of WWII and around the growing threat of nuclear war.⁵³

In a sense, however, Levinas did show the ideological space open for new path, one that could overcome the problem that liberalism could not, namely, as Critchley asks: "How are the legitimate aspirations of liberalism and idealism consistent with the fact that, ontologically, they're completely wrong? Can there be some new path?"⁵⁴ But if the critique of liberalism highlights that "National Socialism is right in its basic intention, it is right in its critique of disembodied liberalism. What it sees is something obscured by liberalism, it sees the elemental enrootedness of the human being,"⁵⁵ at the same time it needs to overcome also the dangers of the other option, that Critchley identifies as Heidegger's facticity when Hitlerism takes it all the way to the politics of the body in the form of racism. This need for a new path that Levinas argues for, while, (according to Critchley), failing in finding, must be searched in the space between the critique of the metaphysics of liberalism and the embodiment of Heidegger's facticity. This is why Critchley says that "Levinas' philosophy, from the beginning to the end, is dominated by the shadow of Heidegger's thought."⁵⁶ And then he explains it thus:

It is Heidegger who permits Levinas the space for a critique of traditional philosophy, specifically, of traditional philosophy understood as intellectualism, theoreticism, or

⁵² For example, in John Rawls, introduction to *Political Liberalism*, *Op. Cit.*

⁵³ This can probably be said of Plato's Republic and the Peloponnesian War as well.

⁵⁴ Critchley, *Op. Cit.*, 245

⁵⁵ *Ibid.*, 226

⁵⁶ *Ibid.*, 224

idealism. The category in Heidegger that permits this critique is facticity, which is the key category for Levinas. The larger task, of which this is a part, is the overcoming of the tragic self-enclosure of Dasein in its finitude. How does one do that? That finitude is precisely what Levinas describes in 1934 as the philosophy of Hitlerism.⁵⁷

For the purposes of our critique, however, it is not necessary to discuss how that path may be possible. But it is relevant to note that Levinas's argument not only explains a metaphysical genealogy of the freedom of liberalism that can still be seen in the later formulations of autonomy that we are analyzing. Levinas also shows that the unrootedness that liberalism promotes must be answered from either another metaphysics or a critique of metaphysics. The attempt at neutralizing its metaphysical thickness that we are discussing fails in solving the problem of liberalism because it still maintains the same underlying metaphysical dangers. Critchley sees in the abstract approach to an autonomous individual the seed of these dangers that Levinas wants to combat:

Liberalism harbors a danger within itself, a danger that's really powerful: that of reducing everything to a game, a spectacle, something from which I can aesthetically distance myself. Liberalism is the political life of the aesthete. I choose my political party as if I were choosing dessert or my taste in works of art. Liberalism runs the risk of aestheticism. The flip side of aestheticism, as Durkheim showed a long time ago, is *anomie*— a society that is anonymous, atomized, alienated. Liberalism risks accepting what Levinas calls “degenerate forms of the ideal.”⁵⁸

The reasons that lead Levinas to think that we need a new path, namely, that liberalism “simply extends the unthinking privilege of theoretical consciousness into the political domain, producing a subject of rights,” that is, that it translates idealist metaphysics into a political philosophy that cannot answer to the appeal of Hitlerism, are, in a certain sense, the same reasons that lead neutralist authors to try and avoid metaphysics. Like Levinas, liberal neutralists need a way to block the advance of Hitlerism, but without calling it a “fundamental evil,” as Levinas can do

⁵⁷ *Ibid.*

⁵⁸ *Ibid.*, 230-231

from his ‘comprehensive view.’ The project of post-metaphysical liberalism cannot do what Levinas’s asks, namely that, in Critchley’s words:

we need to embrace Heidegger’s thinking of facticity against the abstractions of liberalism and also be able to push that thinking in another direction, namely beyond the tragedy of finitude toward responsibility for the other in their height, in their election, and ultimately in their divinity, *à-Dieu*.⁵⁹

However, while the reason that post-metaphysical liberals offer to reject that position is their appeal to neutrality, my claim is that they still hold on to a metaphysics of reason that is incompatible with philosophical positions that question that metaphysics. In other words, the conflict with liberalism is a metaphysical conflict. While liberalism, even in its alleged post-metaphysical formulation, sees its conflict with Hitlerism as that of openness against parochialism, Levinas sees it as the clash of two universalisms, one theoretical, the other forceful, through territorial expansion. This is how Critchley describes this final “extraordinary” move in Levinas’s essay:

From a liberal or Christian universality that is formal or spiritual, to a Hitlerist, embodied, materialist universality based on expansion through force. This is why the contrast between liberal universality and racist particularism is so fatuous. The conflict is really between two notions of universality: formal universality and forceful universality.⁶⁰

What Levinas’s essay and Critchley’s interpretation of it show is that the problem that made liberalism vulnerable to the National-Socialist appeal to authenticity has to do with an emptiness of the subject of rights that liberalism has in its center; an emptiness that produces an “anonymous, atomized, alienated” society. This emptiness is not the outcome of the absence of a metaphysical view, but quite the opposite. The autonomous subject that is demanded by neutrality expresses an idea of freedom with metaphysical content. The move that post-

⁵⁹ *Ibid.*, 226

⁶⁰ *Ibid.*, 231

metaphysical thinkers try to do in response to the dangers of totalitarianism does not *escape* the problem that both Levinas and Hitlerism saw in the 1930s.

2.d The Post-metaphysical Project and Autonomy

From a certain reading, not particularly controversial, Habermas's project can be seen as an attempt at salvaging Enlightenment from the ruins of the World Wars, and Rawls's explicit aim of resuscitating social contractalism in opposition to utilitarianism can also be seen as an attempt at bringing back certain moral conceptions to the political sphere.⁶¹ Both authors, and the wave of new theory they originated, however, were trying at the same time to avoid the problems that enlightenment and the old contractalism tradition presented. Within the rational project of modernity, and its political counterpart, there seemed to be a seed for sectarian politics. The failure of reason as a strong idealist notion and the failure of early liberal institutions in the hands of sectarian politics were seen as one and the same problem. Theoretically, the excess of metaphysics, as a philosophical analogue of fanaticism and religion, was the vanishing point where both strands converged. It would take us too far afield in this narrative to consider the importance of positivism and of analytical philosophy. But following Rorty's second introduction to his *The Linguistic Turn*,⁶²(after he changed his mind and started explaining Philosophy as just another literary genre or an expression of cultural politics),⁶³ it suffices to say that this same problem with metaphysics, and the acknowledgement of the fallibilistic nature of reason,⁶⁴ explains in part the revival of pragmatism. As Richard J. Bernstein

⁶¹ I want to suggest that this might also be in response to the war: the articulation of the American alternative to the soviet philosophy on the eastern block during the cold war.

⁶² Richard Rorty, "Twenty-Five Years After" in *The Linguistic Turn, Essays on Philosophical Method*, Ed. Richard Rorty, (Chicago: The University of Chicago Press, 1967, 1992)

⁶³ See Richard Rorty, "Philosophy as a transitional genre" in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 89-104

⁶⁴ What Rawls calls, "the burdens of reason," as seen in chapter 4

has explained,⁶⁵ stemming from the particularly American reading of Kant and Hegel during the late nineteenth and early twentieth centuries, pragmatism has become more and more prevalent due in part to the fact that it bridges some of these contemporary divides and has offered a theoretical standpoint that can deal with the problem of metaphysics without being too entangled in its disputes. Following Louise Menand,⁶⁶ the origins of American Pragmatism must be traced, not only to the readings of Kant and Hegel in the “metaphysical clubs” of Harvard, but also to two determining facts: the reception of Darwinism and the outbreak of Civil War. Both represent the tendency of detranscendentalization: in the case of Darwinism, a detranscendentalization of the teleological horizon shared by reason and nature proper of pre-Darwinian understandings of nature as either designed by a superior being or as the expression of some underlying rationality; and in the case of the Civil War, of the definitive character of ideological commitments, considering the bloodshed it can bring about, even when its good faith is difficult to dispute. Most telling in this story is the emergence of the ‘reasonable man’ as a standard for evaluating the acceptance of behavior, stemming from the fallibilistic science of probability, in particular, from theories of sociological statistical analysis, and championed by Supreme Court Justice, Oliver Wendell Holmes, one of the original founders of Pragmatism, or from Dewey’s original reformulation of the Hegelian tradition independently of the Frankfurt School. Since the beginning, the problem of the sociality of reason and the political problem of religion and sectarianism were central to the preoccupations of the first pragmatists, as can be seen in Peirce’s concept of “community of inquirers” and William James’s philosophy of religion respectively.⁶⁷

⁶⁵ Richard J. Bernstein, *The pragmatic turn* (Cambridge, UK: Polity Press, 2015)

⁶⁶ Louis Menand, *The Metaphysical Club* (New York (N.Y.): Farrar, Straus, Giroux, 2001)

⁶⁷ See my paper “Dependencies between Logic and Community. Philosophical Implications of Peirce’s Categories for Praxis” *Ramon Llull Journal of Applied Ethics*, Issue 2, 179-194 (Barcelona, 2011)

The post-metaphysical project of liberal neutralists and the problem of the most apparent metaphysical commitments that autonomy entails cannot be separated from the constraints that this history suggests. One constraint is the impossibility of making nature a source of normative value while attempting to produce some sort of universal normativity. As we have seen, the only remaining ground for universal norms, it seems, is the subject, —his or her reason, in particular.⁶⁸ On liberal universalists' account, normativity is determined in relation to what makes human beings flourish according their (second) nature. But while a teleological metaphysics can ground an appeal to nature, (because worldly orders, creatures and objects are themselves infused with normativity, in an Aristotelian way), appealing to human nature, either explicitly (as perfectionists may do) or implicitly (as neutralists do), becomes either vacuous or false.⁶⁹ It appears vacuous if no teleological metaphysics is assumed in the background; and false if a denial of teleological metaphysics is assumed (as is the case with neutralists). If nature cannot be a normative standard, human *nature* cannot meaningfully provide a ground for norms, unless one develops a metaphysics of human agency beyond what neutralists claim to be willing to do.

The project of liberal neutralists consists of walking this fine line between falsity and vacuity without being trapped by either of them. In previous chapters, we have seen several strategies that don't seem to overcome the necessary appeal to some form of human nature to ground universal norms or the basic structure of society (be it the a priori structures of communicative action or the idealization of reasonableness). But with autonomy, the problem redoubles.

⁶⁸ I will develop this insight in section 6

⁶⁹ I am much indebted to Omri Boehm's ongoing project on metaphysics and political thinking, yet unpublished, for this argument. Some hints of it can already be seen in preface to Omri Boehm, *Kant's critique of Spinoza*. (New York: Oxford University Press, 2014) and in non academic pieces such as "Can Refugees Have Human Rights?" in *The Stone*, at nyt.com, on March 9th, 2015.

The problem with autonomy for post-metaphysical liberal neutralists is that it still presupposes that rational beings are ‘value-conferring.’ Norms are grounded in relation to reason rather than nature. The rational willing is regarded as the ground or source of value. For Kant, this is possible because he defended a metaphysics and teleology of reason, instead of nature. That reason can produce value because “man necessarily thinks of his own existence” such that his “rational nature exists as an end in itself.”⁷⁰ But any post-metaphysical project would have to deny such an end in itself, or any other equivalent. Again, if “value-conferring” is assumed as mere brute fact —instead of in a context of a broader metaphysics—, the entire process is vacuous, if it is denied that autonomy implies some sort of end in itself, it runs the risk of being simply false. Most importantly, if it is denied but implicitly assumed, as I claim, the entire system becomes uncritical insofar as a metaphysics of ends in themselves is covertly assumed but repressed from the realm of philosophical reflection.

Omri Boehm has conceptualized this paradox from another perspective.⁷¹ Boehm suggests that the distinction between Kantian normative ethics and Spinozist-Nietzschean ethics of the will to power is rather fragile. Spinoza’s Nietzschean proposition that we never “desire anything because we judge it to be good” but we “only judge it to be good because we desire it” (E4p9s) has the same formal structure of Kantian “value-conferring.” The difference is only apparent in relation to the metaphysical content: unlike Spinoza or Nietzsche, Kant defends a metaphysics of rational will as opposed to natural desire. By locating rational will in a noumenal rather than the phenomenal (natural) realm, Kant is prepared to defend its position as an end in itself: “The only

⁷⁰ (G429)

⁷¹ Boehm, *Op. Cit.*

meaningful difference between Kantian ethics—insofar as it doesn't merely purport to be normative—and a Nietzschean ethics of the will to power can be in the metaphysical background of the former."⁷² From this perspective, post-metaphysical thinkers end up closer to Nietzsche than to Kant, regardless of their intentions. They can only tame the self-interestedness implicit in the appeal to decision theory or prudential decision making (as Rawls does), by appealing to vague notions of reasonability and its social character.⁷³ Boehm states it thus:

The motivation behind this approach, I think, is to eliminate reference to what are considered dubious religious or meta-physical ("transcendent") ideas. But it must be admitted that the value of these needs and interests—whether one is a Mill or a Rawls makes no difference here—must itself be regarded as null unless one has reasons to believe that human beings, or rationality as such, are purposive to some non-accidental good, any anchor of value is as good as human needs and desires. Because this is where post- and anti-metaphysical thinkers turn a deaf ear, their theories are likely to come out as nihilistic as Nietzsche's or Spinoza's. Without justifying the assumption that humanity or rationality is of non-accidental value—an assumption, again, that most modern rationalists positively reject—there is nothing less cynical in Rawls's or Habermas's positions than in Spinoza's conclusion that we judge as good what we desire indeed, their theories become manifestations of doing just that.⁷⁴

The post-metaphysical problem with autonomy is carried over to the collective authorship that the state represents.⁷⁵ Maybe because Christine Korsgaard's recent work on Kantian ethics begins to give a new voice to a metaphysics of normativity, especially to the metaphysics of the self,⁷⁶ she does not shy away from reclaiming the Platonic city-soul analogy to explain this parallelism between the self and the state. This parallelism underlines the connection between an

⁷² *Ibid.*, xv

⁷³ Preface to *Ibid.*, footnote 14.

⁷⁴ *Ibid.*, xv.

⁷⁵ For an early account of the state as author see Hobbes' *Leviathan*, Part I, Ch XXXI, *passim*.

⁷⁶ Christine Korsgaard, *Self Constitution: Agency, Identity, and Integrity*. (New York: Oxford University Press, 2009)

anthropological, strong understanding of autonomy and the place this particular understanding of the works of reason occupies in the political domain of liberalism. Since “what makes a certain event or set of events count as an action attributable to the state is that the state has a set of deliberative procedures” then “the function of the constitution of the state is to unify a diverse group of citizens into a single agent.”⁷⁷ Consider the entire analogy:

According to Plato, a just constitution must be one in which the class of the wisest citizens, in the state — or reason, in the soul — rules over the other parts. A state that is ruled, not by the wise, but by the soldiers or by the wealthy or by the common people contains the seeds of civil war within it, because none of these groups can be relied on to govern, as the wisest do, for the good of the whole. (...) Thus, according to both Plato and Kant, just as the agency of the state is constituted by the adoption of deliberative procedures whose perfect realization depends upon political justice, so the agency of the soul, of the human individual, is constituted by the adoption of deliberative procedures whose perfect realization depends upon personal justice or morality. The unity that is essential to agency and moral integrity are one and the same thing.⁷⁸

When neutralists try to deflate the metaphysical content of liberalism, sometimes they do so by appealing to the deliberative process that grounds norms. With the city-soul analogy, Korsgaard shows that the process of deliberation aims at creating an illusion of unity that grounds action as rational. Citizens “can be said to constitute themselves as a unified agent” with a collective deliberative procedure that mirrors the way in which “an individual human being constitutes himself as an agent, by adopting a procedure for the making of the laws.”⁷⁹ In the same way, the self-awareness that is at the base of rational autonomy is expressed in the “procedure that he [the individual human being] requires because of the reflective distance that makes it necessary for him to act for reasons.”⁸⁰ That reflective distance, that space of self-awareness, is presupposed in the individual by the set-up of the liberal state that aims at being neutral among conceptions of

⁷⁷ Christine Korsgaard, M. *The Constitution of Agency, Essays on Practical Reason and Moral Psychology*, (Oxford: Oxford University Press, 2008.), 13

⁷⁸ *Ibid.*, 14

⁷⁹ *Ibid.*

⁸⁰ *Ibid.*

the good. It also demands of citizens that they adapt their commitments to the existence of that reflective distance and the rational power that it presupposes over them. But more importantly, following the analogy, the public domain becomes such a self-aware space, a space of rationality that is created by neutrality. The co-dependency between neutrality and autonomy is absolute.

In order to see how these metaphysical entailments are dealt with by liberal neutralists, in the next sections I am going to show the different strategies that these authors pursue to explain the relationship between neutrality and autonomy, and if and how they fail in avoiding the strongest metaphysical commitments that the standard just explained suggests.

3. Autonomy versus neutrality

3.a Raz's Perfectionism: Autonomy over neutrality

Joseph Raz believes that neutrality and autonomy are incompatible, and rejects the former in favor of the latter. As a perfectionist, Raz defends the legitimacy of the state in fostering the common good. For our purposes, however, it is not necessary to show how perfectionists argue for that legitimacy; it suffices to highlight two claims that Raz develops about the connection between neutrality and autonomy. The first is critical of Rawls, in which Raz argues that Rawls has not demonstrated the need for neutrality in his system—nor has he shown that it is actually present in his system—precisely because there exists a defense of autonomy at the core of the architecture of his idea of justice. The second defends the idea of autonomy as the underlying ideal of moral pluralism as it is understood by perfectionists.

Raz's critique, however, primarily targets neutrality as a form of restraint on state action, with a view to the problem of consequences of state action (as seen in chapter I.6.b). This is so because

Raz's main aim is to justify state intervention. A neutral arbiter, he claims, ultimately helps the stronger party in a competition. This is the example he uses:

Two children may each appeal to their father to intervene...in some dispute between them. Their father may know that if he simply 'refuses to intervene' the older one, stronger and more resourceful, is bound to come out on top...In other words, the decision to remain neutral, according to the terms of our present definition, would amount to a decision to allow the naturally strong child to prevail. But this may look like a very odd form of neutrality to the weaker child.⁸¹

As showed in Chapter I, the question of the consequences of neutrality is the main problem that liberal neutralists face, and almost nobody is ready to defend a neutrality of consequences either because doing is impossible or because it would call for to an unwelcome, because excessive, form of state interventionism. This is criticized not only by perfectionists that have social-economic policies in mind, but also by those who defend "the right to culture,"⁸² as Avishai Margalit and Moshe Halbertal do: "The right to culture demands that the state abandon its neutral position and actively assist needy cultures, even when these cultures preach their own view of the good life that conflicts with other views within the state."⁸³ Raz believes that autonomy provides a way out of this problem. Not because it grounds neutrality, but because if we accept autonomy as a good to be defended, we can protect weak individuals and groups by making the empowering of the

⁸¹ Raz, *Op. Cit.*, 114.

⁸² "Human Beings have the right to culture —Not just any culture, but their own." Avishai Margalit and Moshe Halbertal, "Liberalism and the Right to Culture" *Social Research*, Vol. 71, No. 3, (FALL 2004), 529

⁸³ *Ibid.*, 530

autonomous capacity to lead a good life the task of the state.⁸⁴ This is why the example in the quotation presents an implicit defense of the good of autonomy and moral self-determination. The example also speaks to the fact that Raz defends a revised form of consequentialist morality.⁸⁵ The ideal of autonomy is meant to address the main problem with the agent-neutral feature of consequentialism, namely, that “the comparable value of any two states of affairs is the same from the point of view of all agents.”⁸⁶ Since this has been criticized as the source of the inequality that pervades the utilitarian predicament of aggregate utility, Raz responds by including the ideal of personal autonomy to account both for the fact of pluralism and the separateness of persons.⁸⁷

Thus, Raz understands autonomy in terms of, and an ideal form of, well-being. In order for such well-being to be possible, several conditions are necessary. The mere willing of something, in the Spinozist-Nietzschean sense, i.e. desiring, does not suffice because “not everything we willingly embrace is something we have freely or deliberately chosen from among various alternatives open to us.”⁸⁸ Following, in part, from this realization, Raz distinguishes between personal and

⁸⁴ Avishai Margalit and Moshe Halbertal take up on this idea of Raz and argue that the state might have an obligation to protect the individual right to culture as a right to pursue one’s idea of the good, especially because the state might be offering a “civil religion,” “ethos,” “narrative” or “metanarrative,” or “any other appellation you may choose for the shared values and symbols of the state’s citizens” that individuals or groups reject because it contradicts their culture. “These shared values and symbols are meant to serve as the focus for citizens’ identification with the state, as well as the source of their willingness to defend it even at the risk of their lives.” Margalit and Halbertal do not focus on the metaphysical weight that these issues might have, but see it from a cultural perspective, which relates their work to that of Kimlyca. In a way, the cultural perspective is similar to my approach, but in another way, it’s different. For example, from Rorty’s “The priority of democracy to philosophy,” in which Rorty argues that Rawls position defends a system that may be neutral metaphysically but at the cost of making it substantially American, and asks that he do that explicitly, defending a particular system instead of approaching the topic from the demands of some neutral philosophy. I think this is in a sense a separate topic, at least it would take us too far afield to see the consequences of it, but it’s still worth noting that there is a game of inclusion-exclusion being played here that shows how neutral liberalism may be too strict in leaving outside the sphere of legitimacy some views that are unreasonable just because they are not liberal. In my reading, this stems from the metaphysical conflict. This connection will be taken up briefly in the conclusions.

⁸⁵ See Part V of Raz, *The morality of freedom (Op. Cit.)* for an extensive explanation.

⁸⁶ *Ibid.*, 268

⁸⁷ *Ibid.*, 268 and ss.

⁸⁸ *Ibid.*, 369

moral autonomy, which, in a sense, is his way of avoiding (much like the Rawlsian method) the metaphysical question of autonomy. Moral autonomy is Kantian morality, which Raz identifies as consisting of “self-enacted principles” in the sense that the will is not “merely subject to the law, but it is so subject that it must be considered as also making the law for itself,” as Kant puts it.⁸⁹ Personal autonomy, “by contrast, is essentially about the freedom of persons to choose their own lives.”⁹⁰ Whether this is ultimately a distinction without a difference or rather something more significant is not important for our purposes; however, it is worth noting that the ideal of personal autonomy is what allows Raz to claim that the conditions necessary for the “capacity for autonomy” consist of three “distinct components: appropriate mental abilities, an adequate range of options, and independence.”⁹¹ Among the mental abilities, Raz includes a minimum of ‘rationality,’ which relates his position with theories of universal rationality, and its attendant problems. The second and third components are relevant for our purposes:

- a) “There must be adequate options available for him to choose from.”
- b) “His choice must be free from coercion and manipulation by others, he must be independent.”⁹²

From both conditions, Raz deduces that “personal freedom...can and should be promoted by political action.”⁹³ The gist of that promotion is both moral and economical:

⁸⁹ Raz quotes this passage in *Ibid.*, footnote 3 of page 370. The passage is from *Fundamental Principles of the Metaphysics of Morals*, tr. H. J. Paton, London, 1956, pp. 98–9

⁹⁰ *Ibid.*, 370

⁹¹ *Ibid.*, 372

⁹² *Ibid.*, 373

⁹³ *Ibid.*, 367

If having an autonomous life is an ultimate value, then having a sufficient range of acceptable options is of intrinsic value, for it is constitutive of an autonomous life that it is lived in circumstances where acceptable alternatives are present. The alternatives must be acceptable if the life is to be autonomous...The existence of many options consists in part in the existence of certain social conditions.⁹⁴

Even if the promotion of these social conditions that lead to an autonomy “presupposing value pluralism,” imposes on governments “the sort of constraints advocated by the reinterpreted harm principle,” Raz rejects frontally “the interpretation of the political protection of freedom as purely a doctrine of limited government.”⁹⁵ In order for this warrant for intervention to make sense, Raz needs to endorse an idea of pluralism that distinguishes between the “forms of the good” that are “admitted to be so many valuable expressions of people’s nature” from “certain conceptions of the good” that “are worthless and demeaning, and that political action may and should be taken to eradicate or at least curtail them.”⁹⁶

3.b Raz’s Critique of Rawls

It is in accordance with this standard that Raz not only rejects neutrality, but also questions the idea that neutralists who endorse some form autonomy, especially Rawls, are really defending a cogent conception of neutrality. In reality, he claims, Rawls’ theory defends the same kind of moral pluralism that perfectionists do because it only protects an “equal ability to pursue ideals of the good only in so far as that ability depends on the principle of equal liberty,”⁹⁷ that is, “neutrality between those conceptions of the good which greatly value an autonomous development of one’s life in accordance with one’s rational nature.”⁹⁸ In a way, Raz is here

⁹⁴ *Ibid.*, 205

⁹⁵ *Ibid.*, 67

⁹⁶ *Ibid.*, 133

⁹⁷ *Ibid.*, 17

⁹⁸ *Ibid.*, 133

criticizing Rawls as Rorty criticizes Rawls in his *The priority of Democracy to Philosophy*; only he puts Rorty on his head. Where Rorty said that Rawls is just defending American liberalism and doesn't pretend he does more than that (and hence isn't culturally neutral) Raz shows that it's not metaphysically neutral. For Rorty, those conceptions of the good which greatly value an autonomous development are American. For us, taking up Raz's critique, they are metaphysical conceptions. According to Raz, there are three characteristics of the Rawlsian variation on the "Kantian insight" regarding moral autonomy that exhibit this:

First, the fundamental idea which enjoys universal validity is that morality is the free expression of a person's rational nature. The claim that principles chosen in the original position express this nature depends on a certain conception of the person which is among the deep common presuppositions of our culture but no more.⁹⁹ Second, Rawls is anxious to make clear that the choice of principles because they express human nature 'is not a so-called "radical" choice: that is, a choice not based on reasons...The notion of radical choice...finds no place in justice as fairness.' Furthermore: 'The ideals of the person and of social cooperation . . . are not ideals that, at some moment in life, citizens are said simply to choose. One is to imagine that, for the most part, they find on examination that they hold these ideals, that they have taken them in part from the culture of their society. The conception of morality as an expression of the rational nature of people is consistent with the view that people's nature is socially determined, thus rendering the concrete manifestation of morality equally socially determined. Third, Rawls' conception of the person does not lead to unanimity of moral views.¹⁰⁰

These three characteristics — rationality as human nature, social contingency as common sense, and pluralism as a brute fact— lie at the core of most liberal neutralist appeals to autonomy, but, unlike Raz, they believe they are compatible with neutrality. Their strategies vary, and with them, the strength of their commitment to autonomy. However, if Raz occupies the position that rejects neutrality because he is in favor of autonomy, at the other end of the spectrum, there exists a strand of liberal neutralism that rejects autonomy in the name of neutrality. These liberal neutralists are receptive to the idea that a commitment to autonomy defeats the purpose of

⁹⁹ John Rawls, "Kantian Constructivism in Moral Theory" *The Journal of Philosophy*, Vol. 77, No. 9, pp. 515-572 (Sep. 9, 1980), 568

¹⁰⁰ Raz, *Op. Cit.*, 131

neutrality, and they try to find a way to support neutrality that will avoid the implication that the self is an autonomous agent. I believe that they fail in their attempt, but their arguments against autonomy show, in part, why I think my argument about mutual dependency is right.

3.c Barry's Neutralism: neutrality over Autonomy

One of the clearest examples of this found in the work of Brian Barry.¹⁰¹ Barry claims that autonomy and neutrality are incompatible because a commitment to autonomy would entail an obligation on the part of the state to favor institutions and policy, such as those related to public funding of education, that defend and promote autonomy, —and this would violate neutrality. “Does a conception of the good as autonomy lead to...neutrality between divergent conceptions of the good? It has been argued that it does. I do not believe that the case can be sustained.”¹⁰²

Barry understands autonomy as a second-order good: the rational, individual choice of a certain first-order good, which implies a certain relationship between reason and beliefs that Barry rightly identifies as a “trespassing” neutrality. According to this non-neutral conception of the good, “what is of central importance in human life is that people should make up their own minds about how to live and what to think, and that they should be able to express their beliefs freely and act on their conclusions about the best way to live.”¹⁰³

A state committed to neutrality, Barry claims, may fund policies and institutions conducive to autonomy, but, unlike a state committed to autonomy, it does not necessarily have to do so. It can find a way to design policies that will not discriminate against other ideas of the good that reject

¹⁰¹ *Op. Cit.*

¹⁰² Barry, *Op. Cit.*, 129

¹⁰³ *Ibid.*

autonomy. neutrality must account for the possibility of choosing not to be autonomous.¹⁰⁴ What should be neutral, then, is not the policy itself, but the procedures of policy-making: if the outcome of that process is the enforcing of what the partisans of autonomy desire, it is inconsequential to neutrally.

How, then, does it [‘justice as impartiality’] put different conceptions of the good on an equal footing while leaving their status as conceptions of the good intact? It does so by insisting that, at the point where the basic principles and rules are being drawn up, no conception of the good should be given a privileged position. Putting the same idea less abstractly, we may say that nobody is to be allowed to assert the superiority of his own conception of the good over those of other people as a reason for building into the framework for social cooperation special advantages for it. This will not prevent from pursuing their own conception of the good, in public life as well as in private, but will mean that they can do so only by means —casting votes, spending money, and exercising rights, for example— that are available to them on the same terms within the neutral constitutional framework.¹⁰⁵

The limit appears when a conception of the good involves “reshaping the life of the rest of the community in accordance with it,” in MacIntyre’s words.¹⁰⁶ If it were the government’s duty to promote the moral education of its citizens, so that they came to live out that conception of the good, or “any serious attempt to embody it in public life,” that would go beyond what Barry calls, “justice as impartiality,” and it would be “proscribed.”¹⁰⁷

All these reasons make apparent the problem of autonomy when it comes to building a neutral state. There needs to be a sphere in which being autonomous is not relevant to determine which laws and policies will be enacted and enforced. Barry tries to claim that if the procedure is

¹⁰⁴ Whatever that “choice” might mean, in this context, is unclear since it appears to be blatantly contradictory, which would show that autonomy is the central appeal of the liberal system, at least in a pre-reflective way.

¹⁰⁵ *Ibid.*, 160-161

¹⁰⁶ MacIntyre, *Whose Justice? Which Rationality?*, 326-48 as quoted in Barry, 125-126 (A “quite accurate account” of justice as impartiality, according to Barry)

¹⁰⁷ *Ibid.*

neutral, in the sense that there is not an advantage or a requirement to be autonomous, or any other conception of the good, and if there are checks in place that protect people who do not believe in autonomy, then the rest is fair game, including pro-autonomy policies. However, that does not solve the problem that the principle of neutrality creates. The respect for diversity implies that in the decision-making process, all views are regarded as equally valid. Barry does not claim that this is simply due to the fact of their existence. If a given idea of the good was the result of coercion or manipulation it would be invalid; after all, adherents to such an idea would not have autonomously ‘conceived’ or ‘assented’ of them.

In other words, for Barry, “justice as impartiality” is a variant of social contract theory in the sense that “the essence of a contract is that each of the contracting parties *voluntarily* accepts constraints on the pursuit of his own ends. And this is precisely what we are looking for.”¹⁰⁸ So while Barry accepts that one may see the pursuit of one’s own ends as non-autonomous, one must autonomously decide to constrain that pursuit for the sake of the social good. Although it is true that the word “voluntarily” can imply forms of decision-making that are less Kantian than the one described above, it is clear that the very purpose of neutrality is not simply to acknowledge that there exists a certain degree of pluralism, but to ensure that both the fact of pluralism and the acceptance of coercion proper of the state against some positions that are present in that plurality of views are the outcome of individual autonomy: that they can be seen as voluntarily accepted. Or, in Barry’s words: “The basic idea here is that just rules are those that can be freely endorsed by people on a footing of equality.”¹⁰⁹

¹⁰⁸ Barry, *Op. Cit.*, 31 (emphasis added)

¹⁰⁹ *Ibid.*, 52

The only difference that Barry sees between his and other versions of neutrality is that “justice as impartiality” does not see autonomy as “what is of central importance in human life.” But his theory does assume that thinking of people as behaving autonomously is the best way to reach an agreement to solve the problem that arises given a plurality of conceptions of the good. He says that justice as impartiality does not judge if the origin of a conception of the good is autonomous or not, but the constitutional framework that he proposes does embody the idea that the procedure that it puts forth for policy-making and political decisions should be neutral in the sense of treating all conceptions of the good as free decisions made by people. Politically, it treats people as autonomous beings. And since justice as impartiality presupposes that living in such a society is something centrally important to human life, because it protects life and freedom, it does believe that treating people as if they were autonomous is central to human life.

When Barry argues in favor of a procedure that guarantees that all conceptions of the good be treated equally, he insists that each group should be allowed to pursue their political goals through the normal channels of free speech and political advocacy, including spending in campaigns and, especially, electoral politics.¹¹⁰ There is little attention to voting in theoretical discussions of neutrality and also little attention to the tension between neutrality and autonomy.

Raz addresses this in a footnote:

This interpretation of the doctrine [the usual neutrality regarding ideals of the good life] is silent on whether individual political action (voting in elections, etc.) may rightly aim at the promotion of the good. Little attention has been paid to this as a separate issue in ‘neutralist’ writings.¹¹¹

¹¹⁰ *Ibid.*, 130

¹¹¹ Raz, *Op. Cit.*, 110, footnote 2.

This is not the place for a phenomenological analysis of voting, but it is worth noting that the act of voting is the single most relevant political action in liberal system, precisely because it signals and expresses the end of the conversation – moving to count the number of people in agreement and disagreement, beyond the subtleties of argument and communication. The entire liturgy of voting is based on individuality, protecting the privacy of decisions. The coercion of voters, in particular attempts at voter suppression, is one of the most salient controversies that arise in liberal democracies. The conditions for electoral rights, that is, voting, (age, mental health, freedom from coercion), the polling protections (booths, privacy of the voter), and the binding character of the results, all speak to the tension between neutrality and individual autonomy. The entire process of an election captures the idea that nothing external should condition the independence of elections. Similar protections are in place when we enter into contracts, or when we establish the rights of minors with respect to their legal guardians. The system is designed to be neutral because it protects the capacity for autonomous decisions in the strongest of senses. Barry’s account of impartiality is the most self-reflective version of this tension, but it is ultimately still unable to reconcile the fact that his understanding of neutrality assumes that the state must respect the choice to lead a non-autonomous way of life with the fact that such a choice is not, properly speaking autonomous. What we see here is not merely an inconsistency in his theory but a paradox at the heart of the concept of autonomy. Is the decision to not be autonomous itself an autonomous decision? Conversely, is the choice to be autonomous not itself a “forced choice” and hence heteronomous? Can we opt out of autonomy? Let us now see other strategies that try to deal with this same problem.

4. Rawls and Larmore: Non-metaphysical Autonomy

4.a Rawls's Political Domain

Rawls's argument in favor of autonomy displays a set of distinctions, or dichotomies, that aim at securing neutrality and freestandingness, to deflect the criticism that it entails a metaphysical position. These distinctions mostly follow the more general dichotomy between the political and the metaphysical, a distinction that we examined closely in Chapter IV. These distinctions are: citizen/person, rational autonomy/full autonomy, pure procedural justice/perfect procedural justice, formal interests/material interests. Each side of the divide is consistent with its "counterpart" in every pair, so that, to the political domain belong citizens with rational autonomy, and a pure procedural justice in defense of formal interests. On the other side, to the metaphysical perspective belong persons with full autonomy, maybe conditioned by perfect procedural justice in defense of material interests. Let us unpack the meaning of these notions.

In keeping with his method of avoidance and his "political, not metaphysical,"¹¹² Rawls's circumscribes what he calls the "conception of the person" to a "normative" and "political" sphere, "not metaphysical or psychological."¹¹³ With this, Rawls tries to exclude his views on the person from the realm of natural description. By saying that it is not metaphysical or psychological, he is saying that his understanding of the moral powers of the person expressed in the idea of autonomy do not represent an account of the ultimate reality of things, nor an empirical, and hence fallible, positive description of the workings of the mind. The normative and political realm as articulated by Rawls, then, *avoids* being criticized by means of denying that

¹¹² reference

¹¹³ Justice as Fairness, p 19

it has any connection with the relevant discourses on the real. It is a normative and political conception of the person that is authoritative precisely because it is meant to exclude scientific or metaphysical arguments from it. It purges those views.

And yet, it is not only a negation: from that normative and political perspective, a conception of the person is affirmed. By creating a separation between the normative and the political, on one side, and the metaphysical and scientific on the other, Rawls opens as space for his theory where he can affirm a conception of the person without having to demonstrate in any way the adequacy, or the epistemological status of that conception. The possibility of such a separation between the normative and the descriptive, between the force of the value and the dispute over the fact, is a severing of notions that is nowhere explained or justified, besides being stated as a brute normative fact and a necessity of the system. This is, again, vacuous or false, as we have seen above, but it shields the conception of the person from all criticism, because it rejects any substance to the claims. Once the separation is affirmed, as if it proved the “avoidance” of metaphysics, a set of concepts is unveiled inside the realm created *ad hoc* for that purpose. Those concepts belong to a very concrete metaphysical tradition, namely Kantianism. But then, this traditional concepts are qualified as ‘political,’ signaling their belonging to this new sphere, as if that assertion magically turned their metaphysical undertones off. While Kant’s separation of the practical and theoretical is justified in terms of their metaphysical import, Rawls simply rejects the conversation. All this is more or less apparent in the entire structure of ‘justice as fairness,’ but it is nowhere more visible than when it comes to autonomy, precisely because autonomy is the keystone of Kantian moral theory, and connects it with his metaphysics of reason.

Autonomy of political life must be distinguished from the ethical values of autonomy and individuality, which may apply to the whole life, both social and individual, as expressed

by the comprehensive liberalisms of Kant and Mill. Justice as fairness emphasizes this contrast: it affirms political autonomy but leaves the weight of ethical autonomy to be decided by citizens severally in light of their comprehensive doctrines.¹¹⁴

Let us not dwell on the fact that Rawls says here that the weight of the Kantian, Millian, non-political autonomy is “to be decided” by citizens, presumably through their very autonomy, and note that persons are assumed and required to be autonomous when acting as citizens in the political domain, independently of any scientific fact of the matter, and of whether it is consistent or not with several comprehensive views.

4.b *Citoyen versus Homme*

This distinction is predicated upon the difference between the notions of citizen and that of the whole person. The notion of citizenship is a normative fiction designed to establish principles of justice. Citizens are seen as equal before the law, and within the principles of justice that articulate the basic structure of society. In order for them to merit such treatment, citizens must possess certain qualities. This is one of the purposes of the ‘device of representation’ that the ‘original position’ is supposed to be:

Citizens are equal in virtue of possessing... the two moral powers. (...) In establishing the fair terms of social cooperation...the only relevant feature of persons is their possessing the moral powers. Features relating to social position, native endowment, and historical accident, as well as to the content of persons’ to determinate conceptions of the good, are irrelevant, politically speaking, and hence placed behind the veil of ignorance.¹¹⁵

As we have seen in the previous chapter, the two moral powers include the “power to form, revise, and rationally to pursue a conception of the good”¹¹⁶ on “reasonable and rational

¹¹⁴ Rawls, political liberalism, 78

¹¹⁵ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 80

¹¹⁶ *Ibid.*, 72

grounds.”¹¹⁷ This capacity, assumed normatively for citizens, is included in their moral powers so that there is “no loss of what we may call their public, or legal, identity,” if they “convert from one religion to another, or no longer affirm an established religious faith.”¹¹⁸ Citizens remain the same after having changed their reasonable comprehensive views because what defines them is their autonomy in doing so —their moral powers— and not the content of their beliefs. If there were no autonomy, and their beliefs were heteronomous or part of an irreducible identity or caused by some forces different from their will, then the change in the content of their beliefs could have an impact on their nature as citizens. While this is said to protect the private sphere from the intervention of sectarian politics, it also forces them to partake in the public domain following the anthropology that values the capacity and virtue of autonomy above all else. And, of course, the fact that we are talking about citizens implies a separation of the natures of the self: one public, one private.

4.c Larmore’s ‘modus vivendi’ liberalism

The conception of the political domain that separates it from the domain of life is, so to speak, a variant of the one that Larmore uses to defend autonomy. Although he denounces the fact that many of the adherents of neutrality, such as Ackerman, fail to achieve genuine neutrality, and strives himself for a ‘neutral justification of neutrality,’ he readily admits that in truth, his own justification, based on the ideas of equal respect and ideal conversation (a mix of Rawls and Habermas), also ultimately falls short of neutrality:

The argument does not aim at complete moral neutrality. It intends to be neutral only with regard to controversial conceptions of the good life and not to all values

¹¹⁷ John Rawls, *Justice as Fairness. A restatement*, (Cambridge: The Belknap Press of Harvard University Press, 2001), 21

¹¹⁸ *Ibid.*, 21-22

and norms whatsoever. Although, it is not completely neutral in this regard either, it is very nearly so, and certainly neutral enough for practical purposes.¹¹⁹

This moderate approach to neutrality,¹²⁰ along with an awareness of its limits, allows him to criticize those versions of neutrality that are significantly less neutral than his own. In particular, it is exceedingly important to him to show that upholding neutrality and autonomy at the same time ends up defeating the purpose of neutrality. He believes that his theory is as strong as it gets because his argument is neutral among ‘controversial ideals of the person,’ such as the ones that stem from the assertion of skepticism, experimentalism, and autonomy.¹²¹

Larmore identifies two kinds of liberalism, which he labels ‘modus vivendi’ and ‘expressivist’. Expressivist liberalism, which he associates with Kant and his followers, demands that the highest political ideal be “mirrored” in our highest personal ideal and for this reason is not neutral.¹²² ‘Modus vivendi’ liberalism, on the other hand, is the one he sees as consistent with neutrality. The expressivist view of liberalism implies a correspondence between state neutrality and the autonomy of the individual “as our highest personal ideal.”¹²³

The autonomous person sustains certain distance toward his substantial ideal of the good. Because this substantial ideal is his only because he has freely made it his, it cannot be constitutive of his understanding of what is of paramount importance to him. His highest ideal remains his autonomy, not the substantial ideal he autonomously chooses.¹²⁴

Larmore makes explicit here something that Rawls seems to reject, namely, that our deepest convictions might not be rationally chosen, but accepted brutally (so to speak), so that if we believe

¹¹⁹ Charles Larmore, *Patterns of Moral Complexity* (Cambridge: Cambridge University Press, 1987), 55

¹²⁰ This moderate approach is, as we saw in the first chapter, sufficient for Larmore’s position.

¹²¹ *Ibid.*, 55

¹²² Which could be labeled Platonic to the extent that it looks like a variant of the city-sould analogy.

¹²³ Larmore, *Op. Cit.*, 76

¹²⁴ *Ibid.*

that the content of our beliefs is rationally chosen it means that the true core of our ideal is not that content, but rather rational choice, or autonomy.¹²⁵ This explains why the reason that ‘modus vivendi’ liberalism does not transgress the limits of neutrality is that he believes it does not demand that we “repudiate” the conception of the good that we hold (by endorsing autonomy, for example), or that we relinquish our attachment to it.¹²⁶ Under this view, “neutrality is simply a means of accommodation...It establishes a modus vivendi between persons whose ultimate ideals do not coincide.”¹²⁷ This is predicated upon two things: separation of the two spheres (right and good or political and non-political domains), and, when in conflict, retiring to neutral ground and explaining our reasons following the “norms of rational conversation.”¹²⁸ These two predicaments are consistent with the claim that ‘modus vivendi’ liberalism can view neutrality as allowing “our substantial notion of the good to remain paramount outside the political realm.”¹²⁹

This divergence between *citoyen* and *homme*, between the “public” (the political) and the private (the nonpolitical), lies at the heart of the modus vivendi view. Liberalism is seen as an ‘art of separation,’ opposed to the idea of society as an organic whole.¹³⁰

Larmore claims that he avoids the trap of autonomy because he does not require that the highest ideals of the person that each of us holds and that explain the fact of plurality—the lack of consensus in political affairs—be grounded in an a priori, Kantian, principle. His ‘modus vivendi’ perspective only demands that in the sphere of the state—where we may retreat to

¹²⁵ This is what Larmore understands by autonomy: “An ideal that demands that we undertake a specific way of life, making its purposes and significances our own, only in the light of critically evaluating on our own its strengths and weaknesses, comparing them with those of other available ways of life.” *Op. Cit.*, 74

¹²⁶ *Ibid.*

¹²⁷ *Ibid.*, 75

¹²⁸ *Ibid.*, 75

¹²⁹ *Ibid.*

¹³⁰ *Ibid.*, 75-76

neutral ground— we conduct ourselves consistently within the limits of rational conversation. We are able to separate our moral ideals from our political ideals because in endorsing some political ideals neither our moral ideals nor our commitments to them are affected. Then, in our lives as *hommes* and *femmes* we can hold ideals that neither defend nor presuppose the ideal of autonomy.

However, this amounts to saying that the motivations behind our moral and political behavior might be rooted in something else, something antecedent to our autonomous choices. That is, either heteronomy, (being empirically motivated), or determinism, (being empirically coerced). neutrality prevents state intervention in favor of one view or another in the name of ‘modus vivendi,’ that is, normative autonomy, but by the same token, coercion or heteronomy in the realm of the “private,” when it comes to moral beliefs, cannot be neutralized by the state. This is something that Raz targets in his criticisms of neutrality theory. The state is then the only domain where we are free from coercion at the hands of other people or certain non-state forces. In the political domain, we operate according to the rules of neutrality, by deciding as if independently of the constraints of our conceptions of the good (in particular when they are not consistent with autonomy), and treating others as not bound by any particular constraint.

The paradox lurking below this architecture is that, on the one hand, we are autonomous in our private lives with respect to the state that renounces coercion, so that we can be free not to regard autonomy as a moral good and disregard individuals as the main unit of moral life. On the other hand, when we enter the public domain as citizens, we are coerced to behave “autonomously” and adopt political principles that presuppose autonomy in us and others, as individuals who, for that reason, deserve equal respect. After all, what is the practical reason for neutrality if not what

Larmore calls “internal circumstances of justice,”¹³¹ that is, the moral requirement of safeguarding people’s personal convictions from the reach of the state? The reason might not be entirely moral, as implied by the expression ‘modus vivendi.’ It might all be a pragmatic attempt at securing peace, in the sense that pluralistic ‘modus vivendi’ is what avoids war and violent sectarian politics. That would amount to saying that there is no argument against a successful dictatorship, that is, one that secured peace —and even, was able to indoctrinate its populace. Larmore and Rawls ask that we treat people as autonomous citizens who merit equal respect, so that every path to moral perfection is open and held to be equally valid.¹³²

4.d The Limits of the Strategy

Rawls’s distinction between rational and full autonomy and Larmore’s endorsement of the substantial moral principle of equal respect both show the limits of this strategy. For Rawls, rational autonomy is the capacity presupposed for the parties in the design of the ‘original position.’ By placing behind the ‘veil of ignorance’ all the contingent circumstances of the person, what remains is only the “rational” capacity to decide on which political principles are the most just. But the existence of those conditions that determine the principles of justice also determine the way in which citizens will actually live under those rules. This is the distinction between rational and full autonomy: the parties in ‘the original position’ are rationally autonomous when deciding the principles of justice, and the citizens that live under those principles are fully autonomous. Rational autonomy is the capacity, while full autonomy is the actualization thereof. The achievement of full autonomy is “modeled by the reasonable

¹³¹ *Ibid.*, 75

¹³² Otherwise, they would be libertarians.

conditions imposed on the parties as rationally autonomous.”¹³³ The principles of justice that appear as the outcome of the ‘original position,’ precisely because they come out of a pure form of autonomy, become the content on the basis of which citizens act politically, and this is due to the fact that acting from those principles is part of their moral powers:

This means that in their conduct citizens not only comply with the principles of justice, but they also act from these principles as just. [...] It is in their public recognition and informed application of the principles of justice in their political life, and as their effective sense of justice directs, that citizens achieve full autonomy.¹³⁴

Full autonomy is the expression of the plenitude of the individual’s moral powers. It includes a form of self-understanding —Rawls says that when fully autonomous, “citizens think of themselves as free”¹³⁵—which contains three self-centered qualities. First, a rational stance vis-à-vis our conceptions of the good —the distance defended by Korsgaard; second, a ‘value-conferring’ capacity: citizens think of themselves as “self-authenticating” sources of claims; third, as responsible for our ends.¹³⁶ This value conferring status is not exactly posited brutly. As I said above: from the capacity to “act from these principles”, we can derive autonomy; but thereby we also derive the idea that human beings are ends in themselves. Since they are ends in themselves, they are value conferring: to be an end is to be value conferring. As an end, what you will has value in relation to you. That makes the argument not at all brute. But, indeed, metaphysical. These three characteristics corresponding to full autonomy are derived from the procedure that is laid out in the ‘original position,’ which represents the pure nature of the moral powers. “Pure

¹³³ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), 78-79

¹³⁴ *Ibid.*, 77.

¹³⁵ *Ibid.*, 72

¹³⁶ *Ibid.*

procedure”, in the ‘original position’, as opposed to he calls “perfect procedure” (with external forces in play), is a mirror of pure practical reason, that is, based on a metaphysics of reason.

Pure procedural justice means that in their rational deliberations the parties... recognize no standpoint external to their own point of view as rational representatives from which they are constrained by prior and independent principles of justice.¹³⁷

The very distinction between external and internal forces represents a very particular and tendentious understanding of human nature, which speaks to the relationship between the situation of the self with respect to the world. The ‘original position’ factors in this distinction by means of ‘the veil of *ignorance*’ (emphasis added), an epistemological notion. The veil marks the border between inner and outer: between the self that only knows itself and its rational content, and the rest, the contingencies and vicissitudes of the world. Ignorance grants us purity, and it is so by design.

This border also separates material interests from formal interests. Comprehensive views are regard in a dual manner. Formally, they are required for a reasonable structure, which implies a certain degree of skepticism about their truth and a certain belief in a particular structure of rationality (as we have seen in the previous chapter). Once that is established, the material content of those views is seen as equal, in the sense that each content has the same value from a formal perspective. The content does not make a difference. Within the constraints of reasonableness, all content is the same. When all that is taken into account, what remains is just the rationally autonomous self, pure as a Kantian *noumenon*, that is, the rational *ground* of justice. The role of the parties is to secure for the moral powers of autonomy an “adequate development and full exercise” following the very concrete understanding of that “exercise” demanded by a metaphysics of reason—this metaphysics is mostly concealed but we know it to be, in the main,

¹³⁷ *Ibid.*, 73

of Kantian persuasion.¹³⁸ The idea of the self that underlies the ‘original position’ is consistent with a conception of the nature of moral goods as abstract containers that, rendered in adequate form, can harbor almost any material content. The interests that parties in the ‘original position’ defend are shaped only metaphysically:

These interests are purely formal: for example, the sense of justice is the highest-order interest in developing and exercising the capacity to understand, to apply, and to act from whatever principles of justice are rationally adopted by the parties.¹³⁹

neutrality and autonomy are here mutually dependent because the idea of respect that is supposedly at the base of neutrality means respect for autonomy, and that respect becomes a duty to protect and foster the rational capacity and the virtue of autonomy,— two sides of the same coin—, as a higher order good. The parties must adopt principles of justice that “enable” persons to advance “some determined (but yet to be specified) conceptions of the good.”¹⁴⁰

Larmore believes this is problematic because it gives too much weight to autonomy, and he realizes that without a metaphysics of reason, that affirmation of the will is indistinguishable from desire and brute individualism. As we have seen, he acknowledges the need for autonomy, and he argues that constitutional principles are legitimate “in virtue of being objects of reasonable agreement, and thus their authority is taken to derive from our collective will as citizens.”¹⁴¹ However, as Christian F. Rostbøll has criticized, he tends to recur to a form of Platonic moral realism when he insists that the principle of respect must have a “deeper kind of

¹³⁸ *Ibid.*, 74

¹³⁹ *Ibid.*, 75

¹⁴⁰ *Ibid.*, 74

¹⁴¹ Charles Larmore, *The autonomy of morality* (New York: Cambridge University Press, 2009), 150

validity”; it “must be considered as a norm binding on us independent of our will as citizens, enjoying a moral authority that we have not fashioned ourselves.”¹⁴²

Rostbøll finds Larmore’s position too substantial, and thus violating neutrality. However, he advocates for a more open defense of autonomy, against Larmore’s view that autonomy-based systems open the door for anti-liberal critiques. Rostbøll believes that Kant can be interpreted as defending a notion of autonomy that only has instrumental value, without metaphysical connotations. Be that as it may, his main point is that because something has instrumental value, or, as he claims, is a necessary condition for a good (as autonomy is for equal respect), that implies that it has no value in itself, and therefore cannot be considered a good (but rather a “means of production”).¹⁴³ In this way, we can defend autonomy as an instrument of the system without opening ourselves to the charge of defending a certain good, and thereby inviting controversy.¹⁴⁴ From this perspective, we can distinguish between autonomy as a principle for regulating interpersonal relations, as a norm to treat others *as if* “they have chosen (or were able to choose and revise) their conception of the good self-reflectively” and not treating them depending on whether “they refuse to see a life of critical self-reflection as the most valuable form of life for themselves.”¹⁴⁵ The former does not violate neutrality, the latter would, he claims. And yet, this not only does not solve the same problems that Rawls’s and Larmore’s incur when trying to justify a non-metaphysical autonomy, but it actually shows that the problem is at the core of the

¹⁴² *Ibid.*

¹⁴³ Rostbøll, Christian F, “Kantian Autonomy and Political Liberalism” *Social Theory and Practice*, Vol. 37, No. 3 (July 2011), 345

¹⁴⁴ *Ibid.*

¹⁴⁵ *Ibid.*

very notion, especially when, as he says, “liberalism needs a stronger justification than the contingencies of a particular culture.”¹⁴⁶

The separation between normative and non-normative —the “*as if*” of the public domain— wants to validate the claim that the way we see others in the public sphere is independent of the way we see the world and ourselves. This way, by behaving autonomously in the public domain, even if instrumentally, we do not necessarily commit to autonomy in our own lives. But, simultaneously, the separation assumes that we are able to relate, with some distance, to our own convictions and comprehensive views in such a way that we are to be able to adapt them, if so required, so they can be expressed in the same terms that justify the separation itself. In order to do both, we must treat ourselves and others as autonomous because the entire system is predicated upon us being capable of doing so.

The fact that Kantian autonomy can be interpreted practically, and not as expressing a metaphysical truth, only makes sense because reason is seen from a metaphysical perspective. Without that subject, there is no such instrument. And it is that subject that sustains the necessary dignity for pluralism to make sense morally. It is the end-in-itself character of the self that grounds it rationally, and moral. That is why Rostbøll must conclude that, from a Kantian perspective, “taking autonomy as a practical principle is a condition of possibility for a regime of mutual respect”.¹⁴⁷ The entire architecture of liberalism, of the freedom it claims to safeguard, and the rights it establishes, is based on a notion of the individual, socially rooted or not, who has the power to make decisions that are regarded as moral not because of their content, but because its *a priori* form. That “value-conferring” capacity of the self makes it stronger not only than the

¹⁴⁶ *Ibid.*, 347

¹⁴⁷ *Ibid.*

concrete historical and material contingencies of its situated circumstance, but of contingency in itself. For that tension to make sense, the very distinction between the metaphysical realm and the non-metaphysical realm must be assumed and upheld. To say that it is normative, or instrumental, does not defuse its metaphysical content; indeed it affirms it.

5. Nagel's paradox

5.a Higher Order Impartiality

What sets Thomas Nagel's view of autonomy and neutrality apart is that he is skeptical on both accounts. Nagel devotes considerable effort to showing that both notions are problematic in themselves. For this reason, he has more hopes of eradicating destructive ideologies than he does of philosophically legitimizing the liberal vision: "The most one can hope is that it [the ideal of reasonable unanimity] will be able to contain most of the disagreement that divides democratic societies, and that the forms of fanaticism which it cannot accommodate will gradually die out."¹⁴⁸ This hope contains something that Nagel assumes as inevitable, namely, that the content of the ideal that is chosen as a vanishing point of one system (reasonable unanimity in the face of coercion, in his case), has some effects that imply the favoring of certain modes of life over others, including the possibility of the disappearance of other forms of life. This is "unavoidable," he claims. Liberal toleration may discourage religious orthodoxy and promote secularism, for example.¹⁴⁹ "Hence it will not *be neutral in effect* among conceptions of the good."¹⁵⁰

¹⁴⁸ Thomas Nagel *Equality and Partiality*. (Oxford: Oxford University Press, 1991), 168

¹⁴⁹ *Ibid.*, 166

¹⁵⁰ *Ibid.*

The result of that restraint [neutrality] appears with suspicious frequency to favor precisely the controversial moral conceptions that liberals usually hold. [...] This leads to the suspicion that the escalation to a higher level of impartiality is a sham, and that all pleas for toleration and restraint really disguise a campaign to put the state behind secular, individualistic, and libertine morality.¹⁵¹

This preoccupation is the basis for Nagel's early critique of Rawls, to which Rawls responded with the modifications we highlighted in chapter IV. Nagel accuses Rawls of presupposing an "individualistic conception according to which the best that can be wished for someone is the unimpeded pursuit of his own path."¹⁵² Therefore, the original position is only neutral among "different life plans of this general type," but not with respect of conceptions of the good that are not as individualistic: "how can this be described as a fair choice situation for principles of justice?"¹⁵³

Nagel, then, tries to offer a version of liberal toleration that "aspires" to be compatible with the views of those who "disagree deeply over many matters of the first importance—including the value of autonomy."¹⁵⁴ In order to aspire to such compatibility, Nagel believes that we need a higher order impartiality, one that goes beyond recognizing the equal value of everyone's life. Being impartial because everyone's life is of equal value "could take its content from a specific conception of the good," defeating the purpose of value neutrality. So the higher-order impartiality he has in mind "operates precisely on the conflicts between different first-order impartialities informed by conflicting conceptions of the good."¹⁵⁵ This way he believes he can

¹⁵¹ *Ibid.*, 156

¹⁵² Cited in Raz, *Op. Cit.*, 117, from Nagel, *Reading Rawls*, ed. N. Daniels, Oxford, 1975, p 10.

¹⁵³ *Ibid.*

¹⁵⁴ Nagel, *Op. Cit.*, 157

¹⁵⁵ *Ibid.*, 155

overcome the problem that arises when the autonomous subject is posited as the ground for value-conferring, respect, and neutrality. Even if a higher order impartiality were able to solve this problem somehow, Nagel still believes it is questionable, because it “seems to require us to subordinate our concern for people’s good to something else, but it is obscure both what that is and why it should carry such weight.”¹⁵⁶

However, when Nagel's specifies the details of this higher-order impartiality, inevitably the problems that we have seen with rationality and autonomy appear nevertheless. Nagel identifies four different grounds for coercion, since coercion is the action that best shows where the limits of state intervention are according to each ideology:

(1) grounds which the victim would acknowledge as valid; (2) grounds which the victim does not acknowledge, but which are nevertheless admissible because he is grossly unreasonable or irrational not to acknowledge them; (3) grounds which the victim does not acknowledge, without being irrational, but which are admissible under a higher-order principle which he does acknowledge, or would be unreasonable not to; and (4) grounds which the victim does not acknowledge—either reasonably or even somewhat unreasonably—and which are such that he cannot be required to accept a higher-order principle admitting them into political justification even if most others disagree with him. It is type (4) that determines the extent of toleration essential to a legitimate system, and the problem is to explain why its supposed instances do not fall under (2) or (3) instead.¹⁵⁷

It is apparent that of these four grounds, two (1 and 3) presuppose a form of autonomy: the victim acknowledges the validity of coercion, either because the ground is accepted as valid or because it depends on a higher order principle that he or she acknowledges as valid. Note that the operative word—acknowledging—is epistemological, rooted in the concept knowledge.

¹⁵⁶ *Ibid.*

¹⁵⁷ *Ibid.*, 159-160

According to dictionaries, it implies *admitting* the reality or truth of something, or recognizing the authority or rights of someone or something.¹⁵⁸ This somewhat rational quality is reinforced by ground number 2, which is the only one that eschews autonomy by appealing to irrationality or ‘unreasonableness.’ Later on, Nagel calls this ground a “basic form of paternalism,” and “the enforcement of criminal law against the *willfully* antisocial” (emphasis added),¹⁵⁹ which implies either a lack of mental capacity or a hostile use of autonomy. The last ground, the touchstone of liberal toleration, illustrates the paradox of autonomy when coupled with neutrality. On this ground, one “cannot be required to accept” either the ground or the higher-order principle, namely, either autonomy or neutrality; which is to say that one can reject them: the former because one must be free to eschew autonomy, the latter because one must be free to pursue a politics of the good. The possibility of rejecting both is based on the combination of autonomy and neutrality.

5.b Autonomy as Belief

On the other hand, in *The View from Nowhere*¹⁶⁰, Nagel offers an epistemological analysis, which is not directly political, of the problems associated with the notion of autonomy. He admits in the beginning of the section on autonomy that he seems unable to take a clear position on the nature and validity of autonomy, because “when we look at the idea closely, it collapses,” but, at the same time, “the alternative form of explanation doesn’t really explain action at all;”¹⁶¹ which leads him to explain why autonomy cannot be thought as anything but a belief, and an inconsistent one, at that:

¹⁵⁸ See definitions in the Merriam-Webster dictionary, for example.

¹⁵⁹ *Ibid.*, 160

¹⁶⁰ Thomas Nagel, *The View From Nowhere*. (Oxford: Oxford University Press, 1986)

¹⁶¹ *Ibid.*, 115

The sense that we are the authors of our actions is not just a feeling but a belief, and we can't come to regard it as a pure appearance without giving it up altogether. [...] It presents itself initially as the belief that antecedent circumstances, including the condition of the agent, leave some of the things we will do undetermined: they are determined only by our choices, which are motivationally explicable but not themselves causally determined.¹⁶²

For Nagel, this belief is inconsistent because to be explained it needs both the absolute freedom that comes from the external view —the view from nowhere— that is the source of our sense of objectivity, and, at the same time, the totally subjective understanding of the causality of action. So while a free action “should not be determined by antecedent conditions, and should be fully explained intentionally,”¹⁶³ which would imply an explanation that “does not take us outside the point of view of the agent,” also there cannot be any other explanation once “all available reasons have been given.”¹⁶⁴ That leaves the question of autonomy at the same solipsistic level where all questions about the validity of subjective impressions are, and “this seems to mean that an autonomous intentional explanation cannot explain precisely what is supposed to explain, namely *why I did what I did rather than the alternative that was causally open to me.*”¹⁶⁵

The philosophical question about the nature of motivation must be asked from a point of view external to the agent, even the belief in autonomy, as an event of the world, must be explained thus, but “why aren't these autonomous subjective explanations really just descriptions of how it seemed to the agent... to do what he did; why are they something more than impressions?”¹⁶⁶ That is a cul-de-sac, to the extent that it is an explanation “insulated from the external view,” the same

¹⁶² *Ibid.*, 114

¹⁶³ *Ibid.*, 115

¹⁶⁴ *Ibid.*, 116 It is worth noting briefly that all available reasons are not necessarily all reasons. What's the problem with the fact that some reasons are not available? This is open to Kant's precise claim in the antinomies: the issue is with the “all”. All reasons cannot be given, because the notion of totality cannot be coherent here. Hence, two perspectives are necessitated.

¹⁶⁵ *Ibid.*

¹⁶⁶ *Ibid.*, 118

external view that “renders illegitimate all further requests for explanation of my action as an event in the world.”¹⁶⁷ And yet, it is precisely because we are capable of that external view, that we believe that our reason has the kind of power, the kind of independence, that can free us from circumstance and contingency. Reason, including Rawlsian reasonability, requires us to adopt the “external view,” to become alienated from ourselves through reason. But does not doing so undermine our belief in autonomy, once we have to look upon ourselves (our beliefs and actions) as mere links in a causal chain?

I have to conclude that what we want is something impossible, and that the desire for it is evoked by the objective view of ourselves that reveals it to be impossible. [...] The very capacity that is the source of the trouble — our capacity to view ourselves from outside— encourages our aspirations of autonomy by giving us the sense that we ought to be able to encompass ourselves completely, and thus become the absolute source of what we do. [...] So the external standpoint at once holds out the hope of genuine autonomy, and snatches it away. By increasing our objectivity and self-awareness, we seem to acquire control over what will influence our actions, and thus to take our lives into our own hands. Yet the logical goal of these ambitions is incoherent, for to be able to be really free we would have to act from a standpoint completely outside ourselves, choosing everything about ourselves, including our principles of choice —creating ourselves from nothing, so to speak.¹⁶⁸

Believing in autonomy implies believing in a kind of freedom that cannot be seen or demonstrated by any means. If we try and look at it objectively from the outside, we also see the works of determinism and the fact that our experience of autonomy is an experience of impressions, with no real clues as to whether our decisions and actions are spontaneous (self-created). This is especially so because the experience of the belief, from the inside, is revealed from the outside as our belonging to “a world we have not created and of which we are the products; it is the external view which both reveals this and makes us wish for more.”¹⁶⁹ We view it

¹⁶⁷ *Ibid.*, 117

¹⁶⁸ *Ibid.*

¹⁶⁹ *Ibid.*, 118-119

from the inside, because “we can only act from inside the world.” We never scape the subjective trap, and the autonomy we experience “appears an illusion, and we who are looking from outside cannot act at all.”¹⁷⁰

This means, to an extent, that the very notion of autonomy presupposes the modern debate around the distinction objective-subjective, inner-world outer-world, the problem of correspondence and the gap between the self and the world, the problem of skepticism, and, of course, the question of how to make sense of noumenal-phenomenal dualism. At its core, the problem is “the literally unintelligible ambition of intervening in the world from outside (an ambition expressed by Kant in the unintelligible idea of the noumenal self which is outside time and causality)”.¹⁷¹ Autonomy is an account of the cause of our actions in a moral sense, it is the answer to the question posed by the affirmation of an inner self with certain powers grounded in itself, the infinite self of some interpretations of Descartes, or “the moral law within” of Kant, which leads us to a never-ending metaphysical regression:

However objective a standpoint we succeed in making part of the basis of our actions and beliefs, we continue to be threatened by the idea of a still more external and comprehensive view of ourselves that we cannot incorporate, but that would reveal the unchosen sources of our most autonomous efforts. The objectivity that seems to offer greater control also reveals the ultimate givenness of the self.¹⁷²

The intrinsic incoherence of this belief clarifies why Nagel is reluctant to endorse autonomy as the vanishing point of toleration, but it also shows that the descriptions of his Kantian reasonable unanimity and the requirement to treat people as ends and not as merely means is another way of describing autonomy and the debate it creates.

¹⁷⁰ *Ibid.*, 120

¹⁷¹ *Ibid.*, 119

¹⁷² *Ibid.*, 118-119

5.c Consequences of the Paradox

Autonomy as a requirement of the public sphere assumes that we can know and rationally explain the reasons for our actions, as coming from the independent, Cartesian self. At least, it presupposes that there is a struggle between material conditions and options (as Raz says), which means that in an ideal world of abundant resources or of multiple “dignified” options, our reasons for acting would suffice to make our moral choices and choose a path of life worthy our dignity as human beings. Thus, it presupposes the self as a source of meaning and action. But in order to escape the subjectivism of the moderns, neutralists need an account of the transcendental reach of the self outside itself, as much as an account of its powers to resist or evaluate external sources, understood as different from the self. All this precludes views of human nature that reject that distinction between our moral reasons and the material constraints, or the self and the external sources over which the self exercises its moral powers, views that describe our moral sphere as maybe determined only by social, or biological, or species-driven reasons.

Hypothetically, those precluded views could ground ideas of, say, biologically grounded roles for genders, or the non-legitimacy of LGTBI rights, as much as opposite views. It could ground an understanding of the violation of the rights of women, LGTBI, or racial minorities, as something rooted not in the tension between self-creation and external imposition, but rather in a dialectic of forces of history collapsing or realizing themselves in the existential concreteness of our lives. The same could be said of collective belongings, from family or group identity, to a radically social understanding of our political life. The appeal of all these values can be witnessed in many accounts of our political life, in appeals to a dignity that is not rooted in self-originating claims. These political narratives could refuse to accept the assumption of autonomous actions

because it is in itself enslaving. It could be seen as imposing a certain individualistic view, a view that may see social or historical conditions as the product of our free choices, thus grounding the prevalence of the status-quo. A critique of the liberal system from certain materialistic, culturalistic, communitarian, feminist, or social views is tamed if the anthropological assumption of individual autonomy governs the system. It may be desirable, from certain conception of the good, that is.

Other alternatives that reject this metaphysical self could include those views that reject grounding altogether, and conceive of action in terms of fate to which we are puppets. We can presuppose teleology without autonomy. Or we could be totally skeptical about our moral powers beyond the half-blind act of empathizing with others as equals. In these alternatives we see the impetus of post-enlightenment critiques: from the limits of situated reason, Heidegger's thrownness in the world and its facticity, or post-war existentialism, to the politics of the body articulated in gender studies, critical race theory, or communitarian cultural politics—to name a few. This leads us back to Levinas's critique of liberalism seen above, and the loss of shared substantive conception of community, Hegel's *Sittlichkeit*.¹⁷³ It is precisely from these critiques that can be seen how it may be true that “liberalism leads to fragmentation and *anomie*,” because it lacks “a kind of authenticity, both in terms of social life and personal commitment of the kind formerly found in religion, especially Christianity,”¹⁷⁴ but also the opposite, in the sense that liberalism puts forth a conception of the good and of human nature that is too thick to really leave room for other views equally thick, both when they are fanatical (e.g., Hitlerism) and when a

¹⁷³ Critchley, *Op. Cit.*, 223

¹⁷⁴ *Ibid.*

escape from the fragmentation of liberalism is longed for (e.g., Levinas's Heideggerianism).

Consider Critchley's take on the fundamental ontology of Hitlerism and Heideggerianism:

The philosophy of Hitlerism is fundamental ontology and its new insight is an insight into the fundamentality of facticity. Hitlerism and Heideggerianism are elemental: they renounce the spiritual, rational, or formal universality of Christianity and liberalism in the name of the identity of the self and the body, the unity of Dasein with world, Dasein with *Mitsein*, and ultimately the people (*das Volk*) and history as heritage (*Erbe*). Such an elemental philosophy is not particularist, but advances a radically new conception of universality based on expansion, force, war, and conquest, rooted in the identity of the people united by blood and race.¹⁷⁵

Levinas, believes that there is a way to use that facticity to unveil liberalism's fragmentation without falling prey to racism. According to Critchley, "Heidegger, Agamben, and Levinas, despite their differences, still insist, in different ways, on the stability of the distinction between facticity and biology."¹⁷⁶ For us, this is but one example of the potential critiques of liberal neutralism that shows how the architecture that forms the structure of the liberal system blocks certain critiques (from the right as much as from the left) precisely because at the gate of the public domain there is a sphinx that asks metaphysical questions.

The indifference of liberal theories to these questions make them blind not only to their own metaphysical commitments, as I have been arguing throughout this dissertation, but also to problems tied to specific liberal institutions. The liberal judicial system is a good example that Nagel uses to problematize autonomy. The rendering of judgments by court officials requires the form of public reason, which presupposes reasonableness and neutrality. Nagel sees the act of judging in court as an act of empathy of reasons and choices, not only of actions, and that presupposes autonomy and transcendental intersubjectivity. Nagel explains it thus:

¹⁷⁵ *Ibid.*

¹⁷⁶ *Ibid.*, 234

The judge's sense of the defendant's alternatives is revealed as an illusion which derives from the judge's projection of his own illusory —indeed unintelligible— sense of autonomy into the defendant. [...] In a sense, the radically external standpoint is not a standpoint of choice at all. It is only when I forget about it and return to my status as fellow creature that I can project myself into the point of view of another agent in the way required for judgement of responsibility. Only then can I evaluate the alternatives facing *him*, and thereby judge him for what he did. [italics in the original]¹⁷⁷

At its root, autonomy assumes both that the self is powerful enough because it is a force in itself, but also that there is a way out of its subjective prism. This way out can only be metaphysical, as Descartes's God or Kant's noumenal causality are. The attempt at finding a way-out in, say, Rawls's use of common sense and background culture, fails because it just assumes the truth of impressions (in Nagel's sense), while 'avoiding' the metaphysical questions that it raises. In this sense, it is no way-out at all.

neutrality can be seen as the view from nowhere in the sense that it is an outer perspective that presupposes actions as free in this particular sense. Liberalism regards commitments as valid if and only if they are reasonable, that is, if they can be rationally explained in the public sphere; and they can be rationally explained because they can be rationally understood in the inner sphere and/or the private sphere of community. To become political, these beliefs are to be translated into the language of autonomy, of decision-making and judgment, in the precise sense in which the self is said to be able to exist and reason, on the one hand, and escape its own prison, by means of practical and social reason, on the other. The expression of reasonability and intersubjectivity, then, includes the understanding every self as if it were autonomous, and could accept the truths and methods of science that stem from it. Or as Jürgen Habermas refers to the 'dogmatic core' of his discourse theory of democracy: 'the idea of autonomy according to which

¹⁷⁷ Nagel, *Op. Cit.*, 123

human beings act as free subjects only insofar as they obey just those laws they give themselves in accordance with insights they have acquired intersubjectively.”¹⁷⁸

6. Markus Gabriel: Liberal Naturalism and the Defeat of Freedom

6.a Metaphysical Quietism

The limits of the avoidance strategy that liberal thinkers try to spell out can be seen from the old tension between Kant and the German Idealists (particularly Hegel and Schelling), but restated in a hypermodern fashion that takes into account the development of philosophy in the twentieth and early twenty-first centuries. Markus Gabriel rehashes this debate with respect to what he calls “liberal naturalism.”¹⁷⁹ In the same sense that Hegel’s critique of Kant tries to show that Kant does not apply to itself the thought that subjectivity can be understood as “constituting forms of reference (categories) outside of which nothing determinate can be apprehended;”¹⁸⁰ and thus, Kant’s metaphysics of intentionality “*is dialectically contradictory under self-application precisely because it does not reflect on its own position, on its own constitution,*”¹⁸¹ liberal naturalism suffers from a similar dialectical contradiction at the level of its constitution.

Here is a brief summary of Gabriel’s argument:

As opposed to the reductionist idea of naturalism, according to which science defines what “counts” as “natural,” so that “in truth only those objects that square with the necessary

¹⁷⁸ Jürgen Habermas *Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy*, (Cambridge, MA: The MIT Press, 1996), quoted in Christian F. Rostbøll, “Kantian Autonomy and Political Liberalism” *Social Theory and Practice*, Vol. 37, No. 3 (July 2011), pp. 341-364

¹⁷⁹ See Markus Gabriel, *Transcendental ontology: essays in German idealism* (London: Bloomsbury, 2013) in particular the chapter on “The Ontology of Knowledge.”

¹⁸⁰ *Ibid.*, xvi

¹⁸¹ *Ibid.*, xvii

ontological assumptions of our best scientific theories can be properly said to exist,” liberal naturalism “admits that there is a human second nature.” This second nature is continuous with the first, which is thereby expanded by “including a particular *human* nature.” Liberal naturalism offers an account of the place of man in the cosmos “without positing any supernatural dimension” or committing itself to metaphysical doctrines “in the pejorative sense of a weirdly premodern onto-theology.”¹⁸² If metaphysics can be understood as the attempt at explaining the world as world, as Gabriel claims, “reductive naturalism is unquestionably a *metaphysics*, in that it claims that everything that actually exists, all content of the world, is describable in natural scientific terms.”¹⁸³ Is in this way that reductive naturalism becomes an easy target for skepticism.

Reductive naturalism presupposes that our empirical data processing is reliable and that the anticipation of the whole underlying the process of data collection is in turn justified by the successful collection of data. Skepticism famously disputes that this circular epistemological structure can or ever could be grounded in anything external to it, precisely because the circle itself constitutes the very possibility of reference to a natural order.¹⁸⁴

Liberal naturalism’s human, second nature is a strategy that is meant to offer an escape from total reductionism, “so as to preserve the space wherein we negotiate and manage our beliefs,” and thereby showing that “skeptical doubts are mere pseudo problems that arise only for a theoretically alienated subjectivity, that is, subjectivity that has reflected itself out of the world.” This position leads to a form of “anti-metaphysical quietism” that can account for skepticism as part of the same temptation that leads to “false, metaphysical transcendence.”¹⁸⁵

¹⁸² *Ibid.*, 2-3

¹⁸³ *Ibid.*, 3

¹⁸⁴ *Ibid.*, 4

¹⁸⁵ *Ibid.*, 41

Gabriel believes that there is a dialectical contradiction in liberal naturalism in the sense that the motivational structure of the theory is incompatible with its manifest propositions, even if there is not direct, logical contradiction within the structure of the theory.¹⁸⁶ The dialectical contradiction of liberal naturalism emerges from the fact that the escape from metaphysics is predicated on a particular conception of human nature as fundamentally free, which liberal naturalism uses to avoid the metaphysical trap of reductive naturalism. However, “the liberal naturalist is, *according to his own theory*, in no position to lay claim to such a concept of freedom.”¹⁸⁷ Why the liberal naturalist is, according to his own theory, in no position to lay claim to such a concept of freedom is where I disagree with Gabriel and the theme of the following epigraph.

6.2. The Defeat of Anti-metaphysics

Following Hegel and Schelling, Gabriel believes that because liberal naturalists utilize skeptical arguments to defuse reductionist naturalism, they end up in a form a quietism that precludes the very possibility of a human freedom as the foundation of that second, human nature.

From this vantage point, it turns out that liberal naturalism violates human freedom, which as Hegel and Schelling see it, consists in the fact that we are above and beyond all that is given at any given time and that things *überhaupt* can only be grasped within the horizon of an unconditioned, infinite, or absolute. In its attempt to purify the human condition of its tendency toward transcendence, since this potentially leads to a skeptical loss of the world, liberal naturalism deprives us of the specifically human dimension of an unconditional freedom, that to which Schelling and Hegel in their utterly antireductionist attitude consistently lay claim.¹⁸⁸

My claim in this chapter (and elsewhere in this dissertation) is, in a sense, the opposite. The contradiction between an appeal to freedom and the anti-metaphysical aim seen in liberal

¹⁸⁶ *Ibid.*, 52

¹⁸⁷ *Ibid.*, 61

¹⁸⁸ *Ibid.*, 61

naturalism is resolved, according to Gabriel, in the defeat of freedom; whereas I have been trying to prove that what results defeated is the attempt at being anti-metaphysical or a-metaphysical. In a way, Gabriel's underlying question is whether it is possible to "integrate skepticism into the construction of metaphysical theories, while at the same time rendering quietism impossible," which, as he says, is "essentially the question of whether or not metaphysics in a post-Kantian idealist sense is feasible."¹⁸⁹ My underlying question, moving in the opposite direction, is whether one can integrate the idea of freedom in politics, in the different forms of respect, consensus, social reason or autonomy proper of liberal neutralists, without assuming and defending institutionally a very strong set of metaphysical commitments, such as those that occupy the dialectical space between Descartes, Kant, and Hegel.

Autonomy is the concept that most clearly shows this tension, and this is why it is so problematic for liberal neutralists. If it is true what Gabriel notes about Hegel and liberal naturalism, namely, that, following Hegel, there exists a "dialectical instability" in their position, an incompatibility between its motivation, "the freedom of self-consciousness," and its conclusion "that the skeptic should simply content himself with what is given, so long as one does not seek to establish the given through philosophical claims,"¹⁹⁰ if this is true, then, it does not follow that the assumption of the given as a common sense fact is a quietist trap that renders freedom untenable, but quite the opposite. Not seeking "to establish the given through philosophical claims" conceals the fact that "the freedom of self-consciousness" is the anthropological center of all liberal theories.

neutrality is of course a moral, regulative ideal, but that does not deflate its metaphysical character at all. On the contrary, its normative force derives from its necessity. As with Nagel's

¹⁸⁹ *Ibid.*, 71

¹⁹⁰ *Ibid.*, 78

paradox, seen above, neutrality offers an outermost perspective from which to contemplate freedom, intersubjective communication, and reasonability as facts of the world pertaining to our second, human nature. neutrality is a necessary condition to accomplish these ideas in reality. Autonomy, intersubjectiveness, and reasonableness require explanation in such a way that can accommodate the external and the internal simultaneously, and, at the same time, account for skepticism and the need for action. neutrality is the perspective that allows for such explanation. It claims for itself the total void as a brute fact, and once that view from nowhere is articulated, then it can claim to see the political reality of freedom, intersubjectiveness, and burdened forms of social idealism. In a sense, it does not *see* them: it *creates* them.

The domain of the political turns that normative force into a requirement for partaking in institutional power. By demanding of the individuals that must be part of the ongoing social contract to accept and live after the political value of that particular metaphysical anthropology, the resulting institutions and policies permanently foster that same understanding of human nature as true. The fact that liberal neutralists pretend to separate the political domain from other domains, as if that made them avoid the claim that those apparently metaphysical commitments are true in a theoretical sense, has the opposite effect. Beyond the theoretical realm and its many controversies, a particular comprehensive view, to use their vocabulary, is enforced in the practical realm.

Conclusions

The Virtue of an Awareness

Contrary to what its defenders claim, the neutral state, as described in contemporary liberal theories, is not neutral. The claim that liberal state theories are non-metaphysical in their commitments and presuppositions also cannot be sustained. In fact, state neutrality theory presupposes and fosters a particular idea of human nature derived from certain metaphysical assumptions. It assumes, at the very least, that the modern subject that lies at its basis is able to avoid solipsism through the sharing of reasons because there are universal rules of communication that express a universal rational structure of the mind, a structure that is ahistorical and non-contingent, even if it transpires amidst the vagaries of the social space. This subject is also capable of being reasonable, which means, according to Rawls and others, that it has the possibility of disengaging itself from its beliefs, and of adapting them to the standards of reasonability present in all other subjects and the intersubjective space they cohabit. In liberal theory, this idealized form of reason is stronger than any “metaphysical” commitments or attachments one may have to their particular social identity

The force of reason is paradigmatically seen in concept of autonomy, which is the cornerstone of the system. A spiritual force, the purest of all, able to create an independent domain of self-awareness and self-consciousness, separate from all contingencies, thanks to the rational powers of the mind. The structure of its rationality, and the freedom it presupposes, constitute a universal, minimal, a priori moral content that forms the basis for political agreement. This moral content, the presence of which is what neutrality is supposed to avoid, is expressed in the ideas of political freedom and equality. The constituent freedom and equality are the reason for

subjects to be treated with equal dignity and why they deserve equal respect in the eyes of the state.

The formal dignity conferred upon the individual by the state does not take into account the content of their beliefs, which must be left outside the political sphere unless one is willing to adapt them. The form of the beliefs has to be rational and/or reasonable in the precise sense the system requires. That dignity does not account for their bodies, their historical situatedness, their social position, their culture, or their belonging to a minority or a majority group within the neutral state. If any of these particularities deserve some sort of state intervention to assure that the subjects are treated fairly, it is so because their formal, a priori, individual qualities merit it, and only insofar as this intervention allows them to be taken in the political domain as equal and free citizens in the sense here stated, so that they can partake in the relevant conversation following the rules derived from these qualities.

This liberal subject is the one admitted in the public sphere. She or he can propose, make, justify, apply, and judge laws from the content of that reason presupposed in her or him. She or he is the bearer of rights, rights that express her nature. And she or he can exclude from that same political sphere those individuals or groups who do not accept one or more of the following: 1) the morality that liberalism endorses (as if it were minimal), 2) the presuppositions that this moral content requires (as if they were only political and not metaphysical), 3) the concept of the subject that is supposed to embody them (as if it were only a predicate of the concept “citizen”), or the view of the world that posits a clear distinction between the world of nature and the world of man, one determined by natural causality, the other by individual, anonymous, formal human freedom (that the system hides).

neutrality is a necessary condition for all these characteristics of the liberal state. neutrality is an outermost perspective that shapes the understanding of the individual as the indivisible building block, the atom of the social molecules. And it is the actual *Cerberus* of the public domain, the gatekeeper of the public sphere. In order for one to partake in the public square, one must adapt one's beliefs and actions to the conditions of inter-subjectivity, reasonableness and autonomy that the system presupposes. If one is willing to believe in the existence and efficacy of those qualities, if one manages to adapt the priority of one's values and views of the world to the presence of those qualities, then one can be successful in partaking in and even controlling the relevant political mechanisms of the liberal state.

Neutrality is what explains the metaphysical status of liberalism. It is not, as one might claim, that it falls short of being neutral and that we find the holes in the theory through which metaphysics creeps in. On the contrary: it is the presence, the need, the intrinsic goodness, and justice of neutrality that are metaphysical and opens up a vocabulary to simultaneously affirm and hide metaphysical commitments.

It may be the case that this picture that liberalism offers is exactly how things are, or at least, approximately how things are, especially in contrast with other comprehensive views. Maybe it is not true, but it's good nevertheless. Maybe it is the only way to exclude from the sphere of power those ideologies and views that lead to a more unstable, unequal, and violent world. Or maybe it is a midpoint in an evolution towards a better system, which could provide a basis upon which to conceive of progress in some unthought-of way, and the limits and constraints it poses are the only path for this new way of looking at the political problem of justice to emerge. Maybe the insufficiencies, cruelties, and oppressions that the neutral system seems to allow are the price we pay for the best of possible systems in our time. But even if one of the above possibilities is right, some things must still be said. First, it becomes undeniable that it is anchored in a particular idea

of the relationship between the self and the world, even that the very idea that these two names—self and world— designate something meaningful and separable, that must be bridged somehow by the works of reason. And anchored in an idea of the relationship between self and self that can be bridged by the shared structure of reason that every self possesses. In other words: regardless of it being good and true or bad and untrue, still is a metaphysical position.

Second, the fact that the principle of state neutrality is a strategy to masquerade the metaphysical positions of liberalism makes it the ideal device to make those positions immune to critique—at least, at first sight—, normalizing them, turning them into the content of common sense.

Politically, neutrality's dependency on methods such as the Method of Intuition, or the method of avoidance (Rawls), or the separation of the political and the *non-political* spheres of life, together with the role of gatekeeper that neutrality plays in that separation, turns liberalism into an essentially conservative position: it creates a trench around the core functions and restraints of the state that lies outside the reach of both critical theories and unthought-of ideas and ideologies.

Moreover, the capacity to exclude views, and the power to preserve the hegemony of certain views, sanctions the status-quo of the time, both its liberal characteristics—as stated—, and the characteristics of other forces, whatever they may be, to which the system turns a blind eye. The national identity that the state espouses, the hierarchies derived from class, gender, ethnic, or cultural differences become much harder, if not impossible, to overthrow within the limits of the liberal system imposed by neutrality. The anonymous, formal, metaphysical subject that lies at its core cannot account for the inequalities that come from the historical, social, or cultural qualities of the concrete person because neutrality forces the state to be blind to them, to deny their legitimacy as political standpoints from where to legislate or judge.

All the while, the system legislates from a situated perspective, which is indeed metaphysical, but also cultural, religious, social, and historical, and to which the system is also blind due to its own metaphysical position, particularly in its understanding of human nature and reason. The conflicts between minority groups and the majority, the struggle against inequalities and oppressions, the claim that the very freedoms and rights that the system purports to endorse is reserved for those that enjoy the power given by their situated selves, —all these political battles cannot be recognized, much less arbitrated, because the principle of neutrality denies their epistemic value as reasonable conflicts. If any of these battles wants to be recognized, it has to adapt their views to the metaphysical subject of liberalism, but also to all the other content that neutrality hides: nation, religion, gender, ethnicity, language, culture, and history. The paradox is that by excluding these qualities from the process of justification —as can be seen in, for example, the ‘original position’—, the neutral state endorses them as they are.

If neutralist liberalism were aware of the metaphysical content that the theory presupposes and the state imposes, two desirable things could happen:

First, in the long run, some of the claims that aforementioned critical views direct against liberal democracies could be welcomed into the public sphere with more guarantees. The awareness of the content that the state harbors could make the sphere of the political more porous and open, due to the intrinsic humility contained in the realization that the liberal state is not neutral.

Second, in the short run, I believe that an awareness of the status-quo bias of liberalism, together with its content, allows us to better understand the rise of populism and nationalism that seems to be threatening to overturn some of the most important victories of liberal democracies, and also to better respond to it. The former, because in the claims of certain populist, far-right, and

nationalistic movements there exists an echo of what Levinas saw in the critique of liberalism that Hitlerism and Marxism put forth. The liberal system excludes views and protects the status quo at a fundamental level, and these movements exploit the space that neutrality ignores to organize a response to the powers that be. Paradoxically, then, the attempt at excluding fanaticism makes fanaticism the only open avenue to criticize liberalism successfully. But the latter, that is, how to offer a better response to this, is also benefited from the awareness of the metaphysical (cultural, historical, etc.) import of neutralist liberalism precisely because it allows for an explicit defense of what it might be good about it. In the same way that neutrality prevents criticism from really happening at the fundamental level, it also prevents the rights that are secured through liberalism to be defended at a fundamental level, that is, beyond the equilibrium of forces that today make it hegemonic.

This last point opens the door for further avenues of research: identifying the precise way in which this metaphysical position prevents certain views from successfully advancing their case might offer a new impulse to some of these critical discourses. In my case, this will be pursued within the context of the studies of collective self-determination and the problem of the *demos* in liberal democracies. In the following years, I will devote time and effort to see both how the current, legitimate *demos* of liberal democracies is described, enforced, and protected by means of neutrality and other characteristics of the liberal state that are explicitly used in law and public discourse, and, how this description may be the source of discrimination against minorities of all kinds.

WORKS CITED

- ACKERMAN, Bruce *Social Justice in the Liberal state* (New Haven, Conn: Yale University Press, 1980)
- AN-NA'IM, Abdullahi A. "Islamic Politics and the Neutral state, A Friendly Amendment to Rawls?" in, *Rawls and Religion*, Tom Bailey and Valentina Gentile (Ed) (New York: Columbia University Press, 2014)
- AUDI, Robert and WOLTERSTORFF, Nicholas, *Religion in the Public Square: The Place of Religious Convictions in Political Debate*, (New York: Rowman & Littlefield Publishers, Inc, 1997)
- BARRY, Brian, *Justice as Impartiality* (Oxford: Oxford University Press, 1995)
- BAILEY, Tom and GENTILE, Valentina (Ed) *Rawls and Religion*, (New York: Columbia University Press, 2014)
- BENHABIB, Seyla. *Critique, norm, and Utopia: a Study of the Foundations of Critical Theory* (New York: Columbia, 1986)
"In the Shadow of Aristotle and Hegel: Communicative Ethics and Current Controversies in Practical Philosophy" in KELLY, Michael Kelly (Ed.) *Hermeneutics and Critical Theory in Ethics and Politics*, (Cambridge: MIT Press, 1990)
- BERLIN, Isaiah. 'Two Concepts of Liberty', in *Four Essays on Liberty*, (London: Oxford University Press. New ed. in Berlin 2002.)
- BERNSTEIN, Richard J. *The Reconstructing of Social and Political Theory*, (USA: University of Pennsylvania Press, 1976)
The pragmatic turn (Cambridge, UK: Polity Press, 2015)
- BLUMENBERG, Hans. *The Genesis of the Copernican World* (Cambridge: MIT Press, 1987)
- BOEHM, OMRI. *Kant's critique of Spinoza*. (New York: Oxford University Press, 2014)
- BRANDON, Robert (Ed). *Rorty and his critics*. (Malden: Blackwell, 2008)
- CALHOUN, Craig MENDIETA, Eduardo, VAN ANTWERPEN, Jonathan (Eds.), *Habermas and Religion*, Ed. et al., (Cambridge, UK: Polity Press, 2013)
- CHRISTMAN, John. "Liberalism, Autonomy, and Self-determination" in *Social Theory and Practice*, Vol. 27, No. 2 (April 2001), pp. 185-206

CRITCHLEY Simon. "Levinas and Hitlerism" *Graduate Faculty Philosophy Journal*, Num 1., (2014), 223-249

DAVIDSON, Donald, "A Nice Derangement of Epitaphs" in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. Ernest LePore (Oxford: Blackwell, 1986)

DELLA ROCCA, Michael, "The Taming of Philosophy," in *Philosophy and its History*, Ed. Mogens Laerke et al., (Oxford: Oxford University Press, 2013)

DWORKIN, Ronald "The Original Position," *University of Chicago Law Review*. (500 1972-1973)

FREEMAN, Samuel (Ed.), *The Cambridge Companion to Rawls*, Samuel Freeman Ed., (Cambridge: Cambridge University Press, 2003)

GABRIEL, Markus, *Transcendental Ontology Essays in German Idealism*, (New York: Continuum International Publishing Group, 2011)

GALSTON, W., *Liberal Pluralism: The Implications of Value Pluralism for Political Theory and Practice*, (Cambridge: Cambridge University Press, 2002)

GAUS, Gerald F., "Liberal neutrality: a Compelling and Radical Principle", in *Perfectionism and neutrality* Ed., Steven Wall and George Klosko, (Oxford: Rowman and Littlefield, 2003) 138
Justificatory Liberalism: An Essay on Epistemology and Political Theory, Oxford: Oxford University Press, 1996)

GEUSS, Raymond, *The Idea of a Critical Theory: Habermas and the Frankfurt School*, (Cambridge: Cambridge University Press, 1981)

GRAUPERA, Jordi, "Dependencies between Logic and Community. Philosophical Implications of Peirce's Categories for Praxis" *Ramon Llull Journal of Applied Ethics*, Issue 2, 179-194 (Barcelona, 2011)

GROFF, Ruth, *Ontology Revisited: Metaphysics in Social and Political Philosophy*. (London: Routledge, 2012)

HABERMAS, Jürgen, *Between Naturalism and Religion*, trans. Ciaran Cronin (Cambridge, UK: Polity Press, 2008)
The Theory of Communicative Action, Vol.1: *Reason and The Rationalization of Society*, trans. Thomas McCarty, (Boston: Beacon Press 1984)
Reconciliation through the use of reason: remarks on John Rawls Political Liberalism *The Journal of Philosophy*, Vol. 92, No. 3. (Mar., 1995), pp. 109-131
LENOX, Sara, LENOX, Frank "The Public Sphere: An Encyclopedia Article" *New German Critique*, No. 3. (Autumn, 1974), pp. 49-55
and Maeve Cooke, *On the pragmatics of communication* (Oxford: Polity, 2003)

Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy, (Cambridge, MA: The MIT Press, 1996)

HEGEL, G. W. F. *Phenomenology of Spirit*. Trans. A. V. Miller. Ed. J. N. Findlay. (Oxford: Oxford University Press, 2013)

System of Ethical Life (1802/3) and *First Philosophy of Spirit* (part III of the System of Speculative Philosophy 1803/4). Trans. Henry Siltou Harris and Thomas Malcolm Knox. Albany: state U of New York, 1979;

Elements of the Philosophy of Right. Trans. Allen W. Wood and Hugh Barr Nisbet. Cambridge: Cambridge UP, 2012.

HEIDEGGER, Martin. "What is Metaphysics?" in *Pathmarks*. Edited by William McNeill (Cambridge: Cambridge University Press, 1998)

HENRICH, Dieter. *The Unity of Reason, Essays on Kant's Philosophy, essay: Ethics of autonomy*, Trans Louis Huny, (Cambridge MA: Harvard U press, 1994)

HOBBS, Thomas, *Leviathan*, Ed. Edwin Curley (Cambridge: Hackett Publishing Company, Inc., 1994)

HOY, David C. and MCCARTHY, Thomas, *Critical Theory*, (Cambridge, MA: Blackwell, 1994)

JENNINGS, Ian, *Against state neutrality: Raz, Rawls, and Philosophical Perfectionism*, (Saarbrücken, Germany: Südwestdeutscher Verlag für Hochschulschriften GmbH & Co. KG, 2009)

KANT, Immanuel. *Critique of Pure Reason*, translated and edited by Werner S. Pluhar (Cambridge: Hackett Publishing Company, Inc., 1996)

Critique of Practical Reason, translated and edited by Mary Gregor (Cambridge, Cambridge University Press, 1997)

KELLY, Michael (Ed.) *Hermeneutics and Critical Theory in Ethics and Politics*, (Cambridge: MIT Press, 1990)

KLEIN, Ernest, *A Comprehensive Etymological Dictionary of the English Language*, (New York: Elsevier Publishing Company, 1971)

KOLAKOWSKI, Leszek, *Modernity on Endless Trial*, (Chicago: The University of Chicago Press, 1990)

Metaphysical Horror, revised edition (Chicago and London: The University of Chicago Press and Penguin, 2001)

KORSGAARD, Christine M. *The Constitution of Agency, Essays on Practical Reason and Moral Psychology*, (Oxford: Oxford University Press, 2008.)

Self Constitution: Agency, Identity, and Integrity. (New York: Oxford University Press, 2009)

KOSELLECK, Reinhart, *Critique and crisis enlightenment and the pathogenesis of modern society*(Cambridge, MA: MIT Press, 2015)

LARMORE, Charles, *Patterns of Moral Complexity* (Cambridge: Cambridge University Press, 1987)
Equality and Partiality. (Oxford: Oxford University Press, 1991)
The autonomy of morality (New York: Cambridge University Press, 2009)

LEVINAS, Emmanuel, “Reflections on the Philosophy of Hitlerism,” trans. Seán Hand, *Critical Inquiry* 17:1 (1990), pp. 62–71.

LUTHER, Martin. “The Freedom of the Christian” in Martin Luther, John F. Thornton, and Susan B. Varenne, *Faith and freedom: an invitation to the writings of Martin Luther* (New York: Vintage, 2002), 3-17

MARGALIT, Avishai and HALBERTAL, Moshe “Liberalism and the Right to Culture” *Social Research*, Vol. 71, No. 3, (FALL 2004), pp. 529-529

MENAND, Louis, *The Metaphysical Club* (New York (N.Y.): Farrar, Straus, Giroux, 2001)

MCCARTHY, Thomas “The Burdens of Modernized Faith and Postmetaphysical Reason in Habermas' Unfinished Project of Enlightenment” in *Habermas and Religion*, Ed. Craig Calhoun et al., (Cambridge, UK: Polity Press, 2013)
“Kantian Constructivism and Reconstructivism: Rawls and Habermas in Dialogue” *Ethics*, Vol. 105, No. 1, pp. 44-63 (University of Chicago Press: Oct., 1994)

NAGEL, Thomas, *Equality and Partiality* (Oxford: Oxford University Press, 1991)
The View From Nowhere. (Oxford: Oxford University Press, 1986)

NOELLE-NEUMANN, Elisabeth (1974), “The spiral of silence: a theory of public opinion,” *Journal of Communication*, (2): 43–51

NOZICK, Robert, *Anarchy, state, and Utopia* (USA: Basic Books, 1974).

PRODI, Paolo, *Una historia de la justicia: de la pluralidad de fueros al dualismo moderno entre conciencia y derecho* (Madrid, Spain: Katz, 2008).

RAWLS, John, *A Theory of Justice* “Original Edition” (Cambridge: The Belknap Press of Harvard University Press, 1971)
A Theory of Justice “Revised Edition” (Cambridge: The Belknap Press of Harvard University Press, 1999)

Political Liberalism (New York: Columbia University Press, 1993)

“The Priority of Right and Ideas of the Good,” *Philosophy & Public Affairs* Vol 17, No. 4 (Autumn, 1988), 251-276

“Justice as Fairness: Political not Metaphysical”, *Philosophy and Public Affairs*, Vol. 14, No. 3. pp. 223-251 (Summer, 1985)

“The Idea of an Overlapping Consensus” *Oxford Journal of Legal Studies* Vol. 7, No. 1 (Spring, 1987), pp. 1-25

“Basic Structure as a Subject” *American Philosophical Quarterly* Vol. 14, No. 2 (Apr., 1977), pp. 159-165

“Kantian Constructivism in Moral Theory” *The Journal of Philosophy*, Vol. 77, No. 9, pp. 515-572 (Sep. 9, 1980)

“The Domain of the Political and Overlapping Consensus” *New York University Law Review* 64 (1989): 233–55

“The Idea of Public Reason Revisited,” *The University of Chicago Law Review*, Vol. 64, No. 3 (Summer, 1997), 765-807

Justice as Fairness. A restatement, (Cambridge: The Belknap Press of Harvard University Press, 2001)

RORTY, Richard, “Metaphilosophical Difficulties of Linguistic Philosophy” Introduction to *The Linguistic Turn, Essays on Philosophical Method*, Ed. Richard Rorty, (Chicago: The University of Chicago Press, 1967, 1992)

“Cultural politics and the question of the existence of God” in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007) 3-26

“Justice as Larger Loyalty” in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007)

“Philosophy as a transitional genre” in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007), 89-104

“Grandeur, profundity, and finitude”, in *Philosophy as Cultural Politics*, Philosophical Papers Vol. 4 (Cambridge: Cambridge University Press, 2007)

“Wittgenstein, Heidegger, and the reification of language” in *Essays on Heidegger and others: philosophical papers*. (Cambridge, Eng.: Cambridge University Press, 1999)

ROSTBØLL, Christian F., “Kantian Autonomy and Political Liberalism” *Social Theory and Practice*, Vol. 37, No. 3 (July 2011), pp. 341-364

SELLARS, Wilfrid, “Philosophy and the Scientific Image of Man” in *Empiricism and the Philosophy of Mind*, (London: Routledge & Kegan Paul Ltd, 1963)

SPINOZA, Baruch, *Theologico-Political Treatise*, trans. R. H. M. Elwes, (New York: Barnes & Noble Inc., 2009)

WALL, Steve and KOSLO, George , ed., *Perfectionism and neutrality*, (Oxford: Rowman and Littlefield, 2003)

WALZER, Michael *Thick and Thin*, (Notre Dame: University of Notre Dam Press, 1994)

“A Critique of Philosophical Conversation,” in Michael Kelly (Ed.)
Hermeneutics and Critical Theory in Ethics and Politics, (Cambridge: MIT Press,
1990)

WENAR, L., 1995, “Political Liberalism: An Internal Critique,” *Ethics*, 106(1), 1995: 41–48.

WILLIAMS, Howard, “Kant, Rawls, Habermas and the Metaphysics of Justice”, *Kantian Review*
Vol. 3. pp 1 - 17 (March 1999)

