UNIVERSITY OF RIJEKA
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
DEPARTMENT OF PHILOSOPHY

Iva Martinić

JUSTICE AND PLURAL CAPABILITIES FOR NON-REASONABLE AND NON-RATIONAL BEINGS

DOCTORAL THESIS

Rijeka, 2025.

UNIVERSITY OF RIJEKA
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
DEPARTMENT OF PHILOSOPHY

Iva Martinić

JUSTICE AND PLURAL CAPABILITIES FOR NON-REASONABLE AND NON-RATIONAL BEINGS

DOCTORAL THESIS

Supervisor: Full Professor, Elvio Baccarini, Faculty of humanities and social sciences in Rijeka

Rijeka, 2025.

SVEUČILIŠTE U RIJECI
FILOZOFSKI FAKULTET
ODSJEK ZA FILOZOFIJU

Iva Martinić

PRAVEDNOST I PLURALNE SPOSOBNOSTI ZA NERAZLOŽNE I NERACIONALNE SUBJEKTE

DOKTORSKI RAD

Mentor: Redoviti profesor u trajnom zvanju, Elvio Baccarini, Filozofski fakultet u Rijeci

Rijeka, 2025.

Mentor: Redoviti profesor u trajnom zvanju, Elvio Baccarini, Filozofski fakultet u Rijeci

Doktorski rad obranjen je dana _____ u Rijeci,

Pred povjerenstvom u sastavu:

1. _____
2. _____
3. _____

ACKNOWLEDGMENTS

First of all, I would like to thank my mentor, Professor Elvio Baccarini, for his invaluable guidance and support in the preparation of this thesis. I am also grateful for the lessons he shared with me on networking, food recipes and Erasmus travelling, as well as for the opportunity to be his "office roommate" during my time working with him.

I would also like to thank the Croatian Science Foundation (HRZZ) for funding my doctoral thesis as part of the project Young researchers' career development project - Training of new doctoral students (Grant: DOK-2021-02-2742) and the project Public justification and the pluralism of capabilities (Grant: HRZZ-IP-2020-02-8073).

I would like to thank all my colleagues, the philosophy department in Rijeka and all the wonderful people I met during my Erasmus exchanges, who undoubtedly enriched my work with their comments, constructive criticism and discussions.

I am infinitely grateful to my parents Ines and Ivan. First and foremost, for making it possible for me to move to Rijeka, for their unwavering support and for always encouraging me. I will always be grateful to them for the upbringing that taught me true values. I love you.

To my sisters Karla and Tihana: Being your middle sister is a true delight, and you are both a constant source of inspiration for me. *Bubek, thank you for being the stand-up comedian who never fails to make me laugh, even in life's most uncomfortable moments.*

I would like to thank Marina "Lifts" for introducing me to the gym and truly changing my life. What started as a simple routine has evolved into something deeply meaningful. With her guidance and encouragement, I have learnt the importance of taking care of my body as well as my mind and I feel stronger in every way. Marina, your support has given me a second chance when I was at my lowest point. I am very grateful to you for opening the door to this journey for me.

I would also like to thank Ozzy, the biggest attention-seeking cat, whose behaviour has allowed me to take breaks from writing.

My deepest gratitude goes to "my people," my dearest friends: Dora, Ante, Ana, and Paola—I adore you. My suns in the darkness, my guiding light. You made me realise that friendship is the greatest and most enduring treasure in life. You have embraced me at every stage of my life, always listened to me, comforted me and cheered me up. Your friendship has shown me what it means to experience real love and to reciprocate it. You understand me, support my growth, and encourage me to go further. *To be*

loved is to be seen, and with you I feel most seen. I will never be able to thank you enough, but I know that you are the greatest blessing in my life, and I do not know what I have done to deserve such jewels. The four of you are the reason I am alive.

And last but not least, I thank my infinite source of joy, Matej. *Volim te.*

Finally, I would like to quote Snoop Dogg[1] and thank myself:

"I wanna thank me for believing in me. I wanna thank me for doing all this hard work. I wanna thank me for having no days off. I wanna thank me for never quitting. I wanna thank me for always being a giver and tryna give more than I receive. I wanna thank me for tryna do more right than wrong. I wanna thank me for just being me at all times."

[1] Snoop Dogg. *Hollywood Walk of Fame Acceptance Speech*. 19 Nov. 2018, Hollywood, California.

SAŽETAK

Ova disertacija istražuje implikacije koncepta pravednosti u uključivanju pojedinaca koji nemaju sposobnost racionalnosti i razložnosti, s ciljem predlaganja inkluzivnijih okvira. Polazeći od paradigme koju je postavio John Rawls, rad će biti strukturiran u skladu s tim temeljnim okvirom, analizirajući njegove mogućnosti i ograničenja u kontekstu inkluzije. Rawlsovo središnje pitanje pravednosti jest kako odgovoriti na izazove pravičnosti među pojedincima koji sudjeluju u produktivnim odnosima te sudjeluju u oblikovanju sustava pravednosti. Moja disertacija proširuje ovo pitanje na pravednost prema osobama koje nemaju sposobnost sudjelovanja u produktivnim odnosima temeljenima na reciprocitetu. Nadalje, pokazujem da je takva ekstenzija nužna. Model reinterpretacije i odgovarajuće nadogradnje pravednosti koji nudim, logikom rasuđivanja vodi i do uključivanja neljudskih životinja u krug pravednosti. U skladu s tim istraživačkim okvirom, iako polazim od Rawlsove teorije, kritički je razmatram i nadopunjujem, osobito s obzirom na njezine temelje u teoriji društvenog ugovora. Ključan dio analize čini usporedba s glavnim suprotstavljenim teorijskim modelom – „pristup sposobnostima“ Marthe Nussbaum. Potencijale rawlsijanske društvene ugovorne teorije u ekstenziji pravednosti izvan skupa osoba koje su produktivni članovi društva temeljem reciprociteta, predstavljam preformulacijom modela rasuđivanja koju je ponudio Rawls. Kao i u izvornoj formulaciji, nudim misaoni eksperiment koji ilustrira ispravan model rasuđivanja o pravednosti. U misaonom eksperimentu zamišljam agente koji ispravno rasuđuju o pravednosti, koje nazivam idealnim razložnim agentima (IRA). Međutim, naglašavam da njihova razložnost implicira da se ne koriste privilegiranom pozicijom u zamišljenom procesu rasuđivanja o pravednosti te da su obvezni proširiti sadržaje pravednosti koje su konstruirali na sva bića koja s njima dijele relevantna svojstva. No, takva šira uključenost pravednosti neminovno povećava sukobe među pravima i rivalitet za resurse u realnom svijetu. Stoga, osim što razmatram pravednost na idealnoj razini, također istražujem kako rješavati te sukobe i rivalitete u praktičnom, društvenom kontekstu.

Drugi dio disertacije usmjeren je na specifičnu skupinu pojedinaca obuhvaćenih ekstenzijom pravednosti, s posebnim naglaskom na pitanja pravednosti u području psihijatrije. Analiziram kritike prema kojima je psihijatrija podložna subjektivnim prosudbama moćnih, što može narušiti njezinu objektivnost i otvoriti prostor za represiju te nepoštivanje autonomije u dijagnosticiranju mentalnih poremećaja. Oslanjajući se na modele opravdanja vrijednosti, razvijam pluralistički pristup koji istovremeno poštuje individualne izbore i osigurava pravednost u dijagnostičkim standardima. Disertacija obuhvaća kako teorijske ideale, tako i praktična ograničenja, zagovarajući psihijatrijsku praksu koja promiče autonomiju, raznolikost i jednakost, dok se istovremeno suočava sa strukturnim nejednakostima.

Doprinos disertacije leži u razvoju pravednijeg i suosjećajnijeg pristupa, s ciljem stvaranja inkluzivnog društva koje bolje odgovara na izazove suvremenog svijeta.

Ključne riječi: kognitivne poteškoće; pravednost; razložnost; mentalni poremećaji; ne-ljudske životinje.

ABSTRACT

This dissertation explores the implications of the concept of justice in the inclusion of individuals who do not have the capacities to be rational and reasonable, with the aim of proposing a more inclusive framework. Based on the paradigm established by John Rawls, the study is structured in accordance with this fundamental framework, analysing its possibilities and limitations in the context of inclusion.

Rawls's central question of justice is how to respond to the challenges of fairness among individuals who participate in productive relationships and contribute to the shaping of a just system. My dissertation extends this question to justice for individuals who lack the capacity to engage in productive relationships based on reciprocity. Furthermore, I demonstrate that such an extension is necessary. The model of reinterpretation and appropriate development of justice that I propose, through the logic of reasoning, also leads to the inclusion of non-human animals within the sphere of justice. In line with this research framework, although I take Rawls's theory as a starting point, I critically examine and supplement it, particularly concerning its foundations in social contract theory. A key part of the analysis is the comparison with the main opposing theoretical model—the "capabilities approach" of Martha Nussbaum. I present the potential of Rawlsian social contract theory in extending justice beyond the group of individuals who are productive members of society based on reciprocity by reformulating the model of reasoning proposed by Rawls. As in the original formulation, I offer a thought experiment that illustrates the correct model of reasoning about justice. In this thought experiment, I imagine agents who reason correctly about justice, whom I call ideal reasonable agents (IRA). However, I emphasise that their reasonableness implies that they do not exploit a privileged position in the imagined process of reasoning about justice and that they are obliged to extend the contents of justice they have constructed to all beings that share relevant characteristics with them. However, such a broader inclusion of justice inevitably increases conflicts of rights and competition for resources in the real world. Therefore, in addition to considering justice at an ideal level, I also explore how to address these conflicts and rivalries in a practical, societal context.

The second part of the dissertation focuses on a specific group of individuals encompassed by the extension of justice, with particular emphasis on issues of justice in psychiatry. I analyse criticisms suggesting that psychiatry is susceptible to subjective judgments of the powerful, which can undermine its objectivity and create space for repression and disregard for autonomy in diagnosing mental disorders. Relying on models of value justification, I develop a pluralistic approach that simultaneously respects individual choices and ensures fairness in diagnostic standards. The dissertation encompasses both theoretical ideals and practical constraints, advocating for a psychiatric practice that promotes autonomy, diversity, and equality while also addressing structural inequalities.

The contribution of this dissertation lies in the development of a more compassionate approach, aiming to create an inclusive society that is better equipped to respond to the challenges of the contemporary world.

Keywords: cognitive disabilities; justice; reasonableness; mental disorders; non-human animals.

PROŠIRENI SAŽETAK

Ova disertacija istražuje implikacije koncepta pravednosti u uključivanju pojedinaca koji nemaju sposobnost racionalnosti i razložnosti, s ciljem predlaganja inkluzivnijih okvira. Polazeći od paradigme koju je postavio John Rawls, rad analizira njegove mogućnosti i ograničenja u kontekstu inkluzije, te predlaže reinterpretaciju pravednosti kako bi obuhvatila prethodno isključene skupine pojedinaca poput osoba s teškim kognitivnim poteškoćama. Cilj rada je razviti teorijski utemeljene i praktično primjenjive pristupe pravednosti koji odgovaraju na potrebe marginaliziranih skupina, s posebnim naglaskom na osobe koje nisu sposobne sudjelovati u produktivnim društvenim odnosima temeljenima na reciprocitetu. Ovo proširenje uključivosti označavam kao prvu ekstenziju Rawlsove teorije pravednosti. Analizom logike argumentacije, ustanovljavam da je potrebna i druga ekstenzija. Ta ekstenzija vodi do uključenosti i neljudskih životinja u sklopu pravednosti. Nakon ovog određenja općeg okvira i obuhvatnosti pravednosti, prelazim na specifična pitanja pravednosti koja se tiču osoba s kognitivnim poteškoćama i mentalnim poremećajima. S obzirom na kritike koje su upućene psihijatriji, prema kojima je riječ o disciplini u kojoj moćni i predstavnici mainstreama nameću vrijednosti, pokazujem kako se u psihijatrijskim praksama može osigurati autonomija, jednakost i objektivnost te spriječiti sustavna marginalizacija pojedinaca. Pravednost se u ovom kontekstu ne odnosi samo na distributivnu pravednost, već i na reguliranje ljudskog ponašanja, rješavanje konflikata i održavanje socijalne kohezije. Stoga, cilj je razviti sveobuhvatnije razumijevanje pravednosti koje pridonosi stvaranju pravednijeg i suosjećajnijeg društva.

U skladu s ranije predstavljenim opisom, disertacija je podijeljena u dva dijela, od kojih svaki odgovara na jedan od glavnih izazova.

Prvi dio rada bavi se kritičkom analizom postojećih teorija pravednosti, s posebnim naglaskom na Rawlsovu teoriju pravednosti i „pristup sposobnostima“ Marthe Nussbaum. Istraživanje pokazuje da Rawlsov model, unatoč svojoj normativnoj snazi, u izvornoj formulaciji ima ograničenja u pogledu inkluzivnosti. Naime, njegov model počiva na racionalnoj deliberaciji i uzajamnoj suradnji te se pokazuje nedostatnim za uključivanje osoba koje nisu sposobne za racionalnu deliberaciju i uzajamnu suradnju. Usporedba s Nussbaum pokazuje da „pristup sposobnostima“ nastoji dati potencijalno širi okvir, ali se suočava s problemom univerzalnih kriterija za određivanje sposobnosti te pitanjem pluralizma. Kao odgovor na te teorijske nedostatke, disertacija razvija model idealnog rasuđivanja o pravednosti. Ovaj model temelji se na misaonom eksperimentu s idealnim razložnim agentima (IRA), hipotetskim djelatnicima koji donose odluke o pitanjima pravednosti ne koristeći se privilegiranim pozicijama. Njihovo rasuđivanje vodi do proširenja obveze pravednosti na sva bića koja dijele relevantna svojstva, čime se uspostavlja princip pravednosti koji nije isključivo vezan za reciprocitet. Međutim, takva šira uključenost pravednosti neminovno povećava sukobe među pravima i rivalitet za resurse u

realnom svijetu. Stoga se, osim idealne razine pravednosti, u disertaciji analizira i način rješavanja konflikata u praktičnom društvenom kontekstu.

Drugi dio disertacije fokusira se na specifičnu skupinu pojedinaca obuhvaćenih ekstenzijom pravednosti – osobe s mentalnim poremećajima i kognitivnim poteškoćama. Kritike Thomasa Szasza i Michela Foucaulta ukazuju na opasnosti koje proizlaze iz subjektivnih dijagnostičkih kriterija i moguće zloupotrebe moći. Kao rješenje, disertacija predlaže modele opravdanja vrijednosti koji omogućuju razvoj pluralističkog epistemološkog okvira. Ovaj okvir nastoji uspostaviti dijagnostičke standarde koji ne samo da poštuju autonomiju pojedinaca, već i osiguravaju transparentnost i objektivnost u procjeni mentalnih poremećaja. Time se izbjegava arbitrarnost u dijagnostičkim procesima i smanjuje opasnost nametanja subjektivnih vrijednosnih sudova, čime se osigurava pravedniji i inkluzivniji pristup u psihijatrijskoj praksi. Kao važan aspekt pristupa, ističem činjenicu da se preporučuje psihijatrijska praksa u kojoj se poteškoće i poremećaji ne analiziraju samo osobinama pojedinaca kao jedinih potencijalnih uzroka tih stanja. Pored osobina pojedinaca, analiziraju se i društveni, ambijentalni i drugi konteksti koji mogu biti nepravični te, kao takvi, zaslužuju biti prepoznati kao primarni uzroci poteškoća ili poremećaja. Drugim riječima, u određenim situacijama potrebno je prihvatiti različitost i atipičnost pojedinaca te ukazati na nepravičnost okoline koja treba biti izmijenjena. Primjena ovog modela razmatrana je analizom fenomena poput suicida, „sindroma uljeza“ te mentalnih stanja depresije, anksioznosti, opsesivno-kompulzivnog poremećaja i poremećaja hranjenja. Pokazalo se da primjena inkluzivnih dijagnostičkih standarda može dovesti do pravednijeg i suosjećajnijeg tretmana pacijenata, smanjujući stigmatizaciju i poboljšavajući njihovu dobrobit.

Ukratko, disertacija pokazuje da koncept pravednosti, ako se dosljedno primjeni, zahtjeva proširenje njenih granica izvan tradicionalnih okvira racionalne suradnje i reciprociteta. Model idealnih razložnih agenata nudi novi pristup promišljanju pravednosti, omogućujući inkluziju osoba s kognitivnim poteškoćama i neljudskih životinja. U području psihijatrije, istraživanje naglašava potrebu za normativno opravdanim dijagnostičkim standardima koji izbjegavaju arbitrarnost i poštuju autonomiju pacijenata. Integracija pluralističkog modela u psihijatrijsku praksu može doprinijeti većoj pravednosti i smanjenju sustavnih nejednakosti.

Doprinos disertacije leži u razvoju teorijskog okvira koji omogućuje pravedniju raspodjelu obveza, kako u općem etičkom smislu, tako i u konkretnoj primjeni na psihijatrijsku praksu. Time se pridonosi stvaranju pravednog i suosjećajnog društva, sposobnog za adekvatno suočavanje s izazovima suvremenog svijeta.

Contents

INTRODUCTION..... 1

1. PART ONE: THE PROBLEM OF RAWLSIAN SCOPE OF JUSTICE 6

1.1. CHAPTER ONE: SECTION ONE: INTRODUCTION TO THE PROBLEM 6

1.2. Section Two: The significance of Rawls' ideas 8

1.3. Section Three: Alternative by Martha Nussbaum 16

1.4. Section Four: Why Martha Nussbaum's Alternative Doesn't Work 23

1.5. Section Five: Solution and replies to Nussbaum's alternative 27

A. Gabriele Badano: Minimal reasonableness and minimal rationality as constituents of personhood **28**

B. Henry Richardson's Approach: Adaptation of Original Position for Individuals with Severe Cognitive Disabilities **32**

C. Samuel Freeman: Contractualist Approach to Individuals with Severe Cognitive Disabilities **34**

D. Cynthia Stark's Approach: Beyond Productivity in Social Contracts **37**

2. CHAPTER TWO: SECTION ONE: NEW SOLUTION: IDEAL REASONABLE AGENTS (IRAS) MODEL FOR JUSTICE 40

2.1. Section Two: Extending the principle of justice: the case of non-human animals 47

A. Eva Feder Kittay Approach **48**

B. Expanding Justice to Non-Human Animals: Rawlsian Theoretical Approaches and Limitations **64**

C. Ideal Justice and Real-World Justice: The Case of Non-Human Animals in Society **71**

CONCLUSION OF PART ONE 78

3. PART TWO: OBJECTIVITY OF EVALUATIVE STANDARDS IN PSYCHIATRIC CLASSIFICATION OF MENTAL DISORDERS 81

3.1. Chapter Three: Section One: Introduction to the problem 81

A. Thomas Szasz: Values, objectivity and the dynamics of power in the diagnosis of mental disorders **82**

B. Michael Foucault: The ideological role of mental disorder and population regulation **89**

C. Towards Objectivity and Pluralism in Psychiatric Classification **91**

3.2. Section Two: Aristotelian replies 93

A. Christopher Megone: Understanding human nature, rationality and the concept of disorder **93**

B. Philippa Foot: Natural Goodness **96**

C. Objections to the Aristotelian answers: Towards a synthesis of objectivity and pluralism **98**

3.3. Section Three: Graham's Rawlsian strategy 105

3.4. Section Four: The solution: A new approach to defining mental disorders 109

A. Justification of General Classifications **109**

B. A weakly externalist model of justification for the determination of mental disorders **112**

3.5. Section Five: Conclusion of Chapter Three 116

4. CHAPTER FOUR: THE TEMPORAL DIMENSION OF UNRESPONSIVENESS TO REASONS IN MENTAL DISORDERS: A DYNAMIC APPROACH TO SYSTEMS OF REASONS 118

5. CHAPTER FIVE: APPLICATION OF THE WEAK EXTERNALIST JUSTIFICATION TO MENTAL DISORDERS 126

5.1. Section One: The Rationality of Suicide 126
5.2. Section Two: Impostor Syndrome 130
5.3. Section Three: Application of the model to depression 133
5.4. Section Four: Application of the model to anxiety disorders 136
5.5. Section Five: Application of the model to obsessive-compulsive disorder (OCD) 140
5.6. Section Six: Application of the model to eating disorders 141
5.7. Section Seven: Conclusion of Chapter Five 143
CONCLUSION OF PART TWO 145
CONCLUSION 147
LITERATURE 149

INTRODUCTION

Justice is a fundamental concept for all well-functioning societies. Following John Rawls (1971), justice should be regarded as the primary value of political institutions, just as truth is the cornerstone of science. My aim is to move beyond Rawls' central focus on cooperation between free and equal persons to examine the broader implications of justice, exploring its full scope to ensure a more extensive understanding and application. Thus, justice reflects the overarching concern to create a just society based on fair principles that prioritise the rights of all, including the most vulnerable. It also means that justice should serve as a framework for regulating human behaviour, resolving conflicts and maintaining social cohesion. It provides a set of norms that regulate the behaviour and status of individuals within a community. These rules help to create social order and prevent chaos by defining acceptable and unacceptable behaviour, giving each individual their due. My project to extend Rawls' theory deals with issues that he himself did not fully explore, but which he recognised as legitimate and necessary to address:

> *I put aside for the time being these temporary disabilities or mental disorders so severe as to prevent people from being cooperating members of society in the usual sense. Thus, while we begin with an idea of the person implicit in the public political culture, we idealize and simplify this idea in various ways in order to focus first on the main question. Other questions we can discuss later and how we answer them may require us to revise answers already reached. We may think of these other questions as problems of extension* (Rawls 1995: 20).

The concept of justice that I aim to address in this dissertation should ensure that individuals are treated fairly and that everyone has access to basic rights and opportunities. When disputes and conflicts arise, the principles of justice serve as a guideline for the resolution process. Legal systems, courts and other institutions are often established to distribute justice and restore injustices to ensure peaceful co-existence. In addition to safeguarding individual rights and freedoms, justice establishes a framework that shields individuals against prejudice, tyranny, and opression. It strikes a balance between the general welfare and individual freedoms. As society progresses, the interpretation and application of justice evolve to reflect changing perspectives, values, and societal needs. This ensures that justice remains a dynamic and adaptable force, capable of responding to contemporary challenges and aspirations.

In the past, theories of social contracts have often linked justice to reason and rationality, often not including those who were not considered reasonable or rational in the discussion. For figures such as John Locke, this connection led to the marginalization of individuals with disabilities. Following Stacy Clifford's (2014) interpretation of Locke, his writings suggest an association between physical defects

and deficiencies in the human mind, reflecting the social prejudices of his time. In his theory of the social contract, Locke introduced a 'capacity contract' that excluded 'lunatics and idiots.' This highlighted a historical bias that linked the capacity for reason to the definition of personhood, which in turn contributed to the exclusion of individuals with cognitive differences from being fully recognized as persons (Clifford 2014).

However, recent years have witnessed a significant shift in perspective, driven by a growing awareness of the diverse patterns of life. This shift acknowledges the rights of a broader group of beings, regardless of their cognitive capacities (Nussbaum 2006; Richardson 2006; Stark 2007; Hartley 2011; Freeman: 2018; Begon: 2023). Moreover, this change in foundational assumptions compels us to explore new dimensions of justice that transcend anthropocentrism and recognize capacities beyond the boundaries of species (Singer 2009; Nussbaum 2006; Kymlicka and Donaldson 2011; 2014).

In this dissertation, I aim to analyse the domains of justice, with a particular focus on ensuring the impartial and fair inclusion of individuals with disabilities and non-human animals within the scope of justice. This analysis aims to advance the discussion of justice and equality for all individuals while examining the justice system's responsibility to all species in our global community. Accordingly, this dissertation addresses two primary challenges.

The first challenge is to extend the principles of justice to a broad spectrum of individuals who are impaired or incapable of being rational and reasonable. To be reasonable means possessing a sense of justice, which implies the moral and political capacity to understand, apply and act upon the principles of social justice that define fair conditions of co-operation[2]. This implies recognising others as free and equal and seeking agreement despite differing comprehensive doctrines (legitimate differences of belief). In contrast, being rational refers to the capacity to develop, revise and pursue an idea of the good. It involves practical reasoning[3] aimed at determining what constitutes a good life for oneself and harmonising decisions with personal goals and values. While reasonableness ensures fair co-operation, rationality guides individuals in shaping and achieving their personal goals (Rawls 2001: 18-19). The challenge of extending the principles of justice to individuals who are not reasonable or rational

[2] Rawls sees co-operation not just as “working together” but as a system of fair interaction between free and equal individuals, guided by principles that ensure fairness and reciprocity. In *Justice as Fairness*, he argues (2001) that the terms of co-operation should be chosen behind a "veil of ignorance" to ensure impartiality. This co-operation takes place within the basic structure of society — its institutions and rules — which must be designed to provide fair opportunities and an equitable distribution of benefits. I will explain this in more detail in the next sections.

[3] Practical reasoning, in Rawls' paradigm, refers to the instrumental process of deliberating and deciding on actions that align with one’s personal preferences and conception of the good life, by evaluating means to achieve desired ends effectively (Rawls, A Theory of Justice, 1971).

arises from the pursuit of a future that ensures fairness for all. Justice must not only recognise but also embrace the diversity of life that coexists in the world. It must recognise that this diversity enriches our collective experience, rather than being a marginal aspect of it.

As an introduction to this challenge, I draw on Martha Nussbaum's (2006) critique of John Rawls' theory of justice, in which she emphasizes the need for greater inclusivity in traditional conceptions of justice. In other words, the first challenge of the dissertation is concerned with not including individuals with certain disabilities in Rawls's theory of justice because they cannot fulfill the criteria of reasonableness and rationality. I argue that Rawls' theory needs further examination to fully include them and ensure that their rights and needs are adequately addressed.

I propose a politically liberal solution, meaning it is based on the idea that a theory of justice should be independent of any particular comprehensive doctrine so that reasonable people can accept it. To illustrate what such acceptance entails, I construct a theoretical model called *Ideal Reasonable Agents* (IRAs) [4] that shows how reasonable deliberation should work. At the centre of this argument are IRAs as hypothetical individuals in a thought experiment, characterised by their capacities for both reasonableness and rationality, ensuring that the principles of justice emerge from fair and inclusive reasoning. These agents are idealised through a theoretical process that refines their attributes to represent fairness and impartiality. They are crucial to the development of principles of justice, which encompass fundamental rights and protections designed to ensure fairness and equality in society. The main argument focuses on a broader understanding of the term "reasonableness" and on the IRA's capacity to imagine analogous conditions in a thought experiment. It emphasises the universal extension of rights so that the needs of individuals who are unable to be reasonable and rational (such as individuals with severe cognitive disabilities) are taken into account. This approach is based on the principle of universalisation, emphasizing that the principles of justice should address shared, common characteristics rather than create distinctions based on specific individual or group differences. In particular, it rejects the exploitation of a privileged position that arises from possessing certain advantageous characteristics (Martinić and Baccarini, 2023).

Building on this foundation, I will broaden the scope of justice further by exploring its requirements concerning non-human animals. In doing so, I will distinguish between two levels of justice: ideal justice, which refer to to abstract principles, and real-world justice, which addresses practical considerations and constraints in the

[4] The term *Ideal Reasonable Agents* (IRA) used in this context corresponds to the concept of *Ideal Legislators* introduced in the joint article, 'Capabilities and Justice for People Who Lack the Capacity for Reason and Rationality,' published in *Filozofska istraživanja* 43.3 (2023): 495–507. This article was co-authored with my doctoral supervisor, Elvio Baccarini, as part of the JOPS research project.

context of existing societal conditions. Thus, while ideal justice advocates for inclusivity and equal consideration for all living beings, real-world justice accounts for practical constraints and limitations. Ultimately, the overarching goal of the dissertation's first challenge is to develop an inclusive perspective of justice that focuses on extending rights and protections to two key groups: individuals with severe cognitive disabilities and non-human animals. By addressing their needs through a politically liberal and universally applicable framework, I aim to contribute to a more equitable and comprehensive understanding of justice[5].

Following the first challenge described above, the second challenge, which I will analyse in my dissertation, is addressed by Thomas Szasz (1960; 1994; 2000) and Michel Foucault (1989). This challenge critiques psychiatry's approach, raising concerns about its ability to uphold true objectivity and to respect the individual as a free and equal agent within its classifications of mental disorders. In particular, the objectivity of psychiatric diagnoses is questioned, with both Szasz and Foucault arguing that these diagnoses are often based on value judgements and not on objective, naturalistic categories. This critique highlights the historical treatment of individuals within psychiatry, emphasising how subjective judgments may have influenced the classification of mental disorders. To address the second challenge of balancing fairness and personal relevance in defining mental disorders while avoiding sectarian impositions, I propose a weak externalist model of justification inspired by Gerald Gaus's theory (2011)[6]. This model has two primary aims. First, it addresses pluralism by creating a form of objectivity grounded in public justification, ensuring that standards for defining mental disorders are based on the convergence of diverse perspectives rather than the imposition of a single viewpoint. This approach affirms diversity within a political community, reducing the subjectivity often associated with psychiatric classifications. Second, the model establishes a framework for more objective evaluative standards[7] in psychiatric diagnoses. Drawing on Baccarini and Lekić Barunčić (2023), I will explain how public justification distinguishes disorders

---

[5] I would like to thank Tom Shakespeare for his valuable feedback at the "Disability and Justice" conference, MANCEPT 2024, in which he emphasised my primary intention — not to diminish the status of humans with cognitive disabilities, but rather to elevate the status of non-human animals.

[6] The ideas for this approach were developed in collaboration with my doctoral supervisor, Elvio Baccarini, and the JOPS research project. More specifically, it is part of an article co-authored with Baccarini and Shane Glackin. The model is also presented in the article by Baccarini and Lekić-Barunčić (2023).

[7] By evaluative standards, I mean criteria that are used to evaluate and judge something. In the context of psychiatry, these standards are used to assess mental health conditions and determine whether someone has a mental disorder. Essentially, they help professionals decide whether certain symptoms or behaviours meet the criteria for a particular diagnosis. These standards aim to create a uniform method of diagnosing and treating mental disorders while taking into account individual differences and needs.

from mere diversity, ensuring that criteria for mental disorders are inclusive and equitable. Thus, the model focuses on two key questions: defining mental disorders in a general sense and applying this definition to determine whether a specific condition qualifies as a disorder. Using a weak externalist justification inspired by Gaus (2011), I will assess when a person is unresponsive to reason, aligning with the general definition of a disorder. Recognising the pluralism of values and reasons that shape individuals' understanding of wellbeing, this approach aims to improve fairness and consistency in mental health assessments by addressing biases and value-laden assumptions inherent in psychiatric classifications. To demonstrate its practical application, I will apply this weak externalist model to case studies, showing how it can contribute to more consistent mental health assessments in practice.

In conclusion, this dissertation expands the principles of justice to include the rights and needs of beings traditionally excluded, focusing on individuals with cognitive disabilities and nonhuman animals. By taking up and extending Rawls' theory of justice, it argues for a more inclusive framework that recognises diversity as a central element of justice rather than a peripheral concern. This dissertation develops a weak externalist model of justification to address the challenges of objectivity and fairness in psychiatric assessment. Taken together, these explorations aim to advance a vision of justice that is both inclusive and adaptive, reflecting the evolving needs of a diverse and interconnected global community.

1. PART ONE: THE PROBLEM OF RAWLSIAN SCOPE OF JUSTICE

1.1. CHAPTER ONE: SECTION ONE: INTRODUCTION TO THE PROBLEM

The American political philosopher John Rawls is widely recognized as one of the most influential thinkers in the field of political philosophy. His seminal work *A Theory of Justice*, first published in 1971, established the fundamental framework for the discourse on the scope and reach of justice in society. This work was followed by his examination of political liberalism in *Political Liberalism*, first published in 1993. In this later work, he elaborates on the need for a political conception of justice that can be accepted by citizens with different moral, religious, and philosophical views.

Rawls' theory of justice, which aims to ensure fairness and equality in the formulation of principles of justice, has had a significant impact on contemporary political philosophy. However, despite its profound influence, Rawls' theory has been criticized for its applicability to disadvantaged groups, particularly individuals with severe cognitive disabilities. In this chapter, I will examine the critical alternative to Rawls' framework proposed by Martha Nussbaum (2006). In her seminal work *Frontiers of Justice*, she offers a profound critique of Rawls' philosophy, focusing on the limits of the social contract tradition to which he adheres. While Rawls seeks a universal justification for justice through the social contract, Nussbaum argues that this framework is inadequate for addressing the needs and rights of individuals with disabilities.

Nussbaum's critique (2006) of social contract theory focuses on three unresolved issues of social justice. First, she highlights the issue of justice for individuals with cognitive or physical impairments, arguing that they are not adequately included in society on equal terms in areas such as education, healthcare, political rights, and freedoms. Addressing these demands requires a new way of thinking about the foundations of social cooperation, moving beyond the notion of mutual advantage, which she deems an inadequate basis for justice. Nussbaum insists that what is needed is not merely a new perspective within the existing paradigm but a fundamental shift in approach. The second issue she raises concerns the expansion of justice beyond national borders to ensure fairness for all people worldwide, advocating for a conception of justice that is not constrained by the limits of the nation-state. The third set of concerns pertain to justice for non-human animals. Nussbaum argues that the suffering inflicted upon animals by humans is generally viewed as a moral concern but not as an issue of justice, a perspective she seeks to challenge (Nussbaum 2006).

A crucial aspect of Nussbaum's argument is her call for a departure not only from the model of justice based on mutual advantage but also from the model of the rational being as the sole subject of justice. Her central critique is directed at social contract theory, which, in her view, Rawls developed to its highest level, successfully demonstrating its superiority over utilitarianism. However, even Rawls (1995)

recognized that certain issues remained unresolved within this framework. Nussbaum contends that these challenges cannot be adequately addressed within the social contract tradition. She highlights a fundamental problem within society: the identification of some individuals as fully cooperative and others as parasitic, a distinction rooted in the idea of justice as contingent on mutual benefit. This, she argues, is an inherent limitation of social contract theory, necessitating a broader and more inclusive approach to justice (Nussbaum 2006).

In response, Nussbaum proposes the "capabilities approach", which focuses on the basic capabilities—or opportunities—that individuals need to lead a full life of dignity and agency (2006; 2011). This approach aims to ensure that all individuals, regardless of their physical or cognitive conditions, have the opportunity to achieve well-being and participate fully in society. In this context, Nussbaum (2006) criticises Rawls for assuming that only rational individuals motivated by self-interest are included in justice. She argues that this perspective does not apply to people with permanent disabilities or severe cognitive disabilities, who do not fit into the conventional notion of 'reasonable people'. She contends that Rawls' framework for designing principles of justice, which assumes participants are fundamentally capable of being cooperative members of society, fails to consider the needs of those who do not meet this standard. Furthermore, she asserts that this oversight arises from Rawls' assumption that individuals represented in the hypothetical negotiation are not characterized by "permanent disabilities or mental disorders so severe that they prevent people from being cooperative members of society in the usual sense" (PL, 20[8]). This is why Nussbaum (2006) argues that her approach is a more inclusive form of justice and emphasizes the need for laws and policies that benefit all members of society. While she acknowledges that her capability theory requires further analysis and comparison with Rawls' framework, she continues to advocate for refining our understanding of justice to better address the diverse needs of individuals in society.

This chapter will provide an examination of Nussbaum's critique and her capabilities approach, exploring how these insights aim to create a more equitable framework for justice. I will compare Nussbaum's views with Rawls' theory to assess how well they meet the needs of people with disabilities. Finally, I will examine how each approach supports fairness and inclusion and suggest how to improve them. The aim is to understand the strengths and weaknesses of both theories and explore how modern political philosophy can better promote inclusion in today's diverse society.

Therefore, the main aim of this chapter is to critically analyse Nussbaum's critique of Rawls' theory of justice, focusing on its limitations in relation to the inclusion of individuals with severe cognitive disabilities. To achieve this, I will first provide an account of Rawls' theory of justice, as this foundational explanation is essential for

[8] https://politicalnotmetaphysical.wordpress.com/2016/07/01/basic-issues-can-rawlsians-offer-a-plausible-account-of-disability-justice/ 08.07.2023.

understanding Nussbaum's critique. I will argue that while Nussbaum's perspective aims to offer a broader framework for justice by focusing on the capabilities essential for human well-being, her approach has limitations, especially in fully addressing the justice needs of individuals with severe cognitive disabilities. These limitations will be explored through an examination of key criticisms of Nussbaum's framework.

In the final chapter of Part One of this dissertation, I will propose a solution to address the limitations of Rawls's concept of justice and the challenges identified in Nussbaum's approach. This solution aims to combine insights from both theories while overcoming their respective shortcomings, ultimately contributing to a more inclusive framework for justice.

1.2. Section Two: The significance of Rawls' ideas

The introduction of Rawls' framework is important for several reasons. First, it provides essential context. A basic understanding of Rawls's core principles sets the stage for the analysis that follows, helping even readers unfamiliar with his work grasp the key concepts and terms used throughout the dissertation. Second, it allows for informed comparison. With a clear understanding of Rawls' theory, readers can more effectively engage with Nussbaum's critique, assessing the strengths and limitations of Rawls's framework, particularly in relation to individuals with severe cognitive disabilities. Third, it ensures clarity and coherence. A brief overview of Rawls' ideas reduces the risk of ambiguity, enabling readers to follow the arguments and counterarguments more systematically. Lastly, it enables critical engagement. Understanding Rawls' principles allows readers to evaluate Nussbaum's critique within his theoretical framework, fostering a deeper understanding of the issues discussed.

To fully understand the significance of Rawls' ideas, it is important to first recognise that his work is firmly rooted in the tradition of social contract theory. Before outlining Rawls' basic ideas and emphasizing his progressive stance on social contract theory, I will briefly highlight the key differences between Rawls and his predecessors, including figures such as Thomas Hobbes, John Locke, and Immanuel Kant[9]. I will also present the elaboration of the social contract theory itself. This comparison will help to illustrate how Rawls improves on and goes beyond the theories of his predecessors.

The theory of the social contract is a central concept in political philosophy that explores the legitimacy and origins of state power through a – hypothetical or actual – agreement between individuals. This tradition has evolved considerably from its classical origins to contemporary formulations. The classical theory of the social

[9] I base my reflections on the proponents and their differences primarily on the review of Nussbaum's book (2006), which I consider to be one of the clearest and most concise overviews of the proponents of the social contract theory.

contract is based on the concept of natural rights—rights that each person inherently possesses and that the social contract seeks to protect. In contrast, contemporary theories view the social contract as a mechanism for creating rights and principles of justice (Nussbaum, 2006). The primary predecessors of the classical social contract theory are Thomas Hobbes, John Locke, and Immanuel Kant.

Thomas Hobbes argued that without political authority, individuals would live in a state of constant uncertainty and chaos. To escape this state, people enter a social contract and, in return for security and order, surrender some of their freedoms to a sovereign authority. Hobbes emphasized the need for a strong, central authority to maintain peace and prevent conflict (Nussbaum, 2006: 10). In contrast, John Locke began with the premise that individuals are free, equal, and independent by nature. Locke's theory of the social contract is based on the idea that no one has an inherent right to rule over others and that everyone has the right to self-government. He emphasized the importance of mutual respect and the protection of property rights. Locke focused on moral duties such as self-preservation and the protection of the liberties of others. His theory integrates the concepts of individual dignity and mutual benefit, suggesting that the social contract should promote both personal freedom and collective advantage. In order to further develop the theory of the social contract, Immanuel Kant emphasised the moral necessity of joining a civil society governed by universal laws. Kant's approach, reflected in works such as *Groundwork of the Metaphysics of Morals* (2012) and *Critique of Practical Reason* (2002), centres on the idea that individuals, as rational beings, must abide by moral laws that are universally applicable. Kant's theory integrates the concept of the categorical imperative to act according to rules that one wishes to be universalised. His vision of the social contract is less about mutual advantage and more about creating a just and moral system in which individuals recognise each other as morally equal. This perspective also extends to the international level, proposing a "league of nations" and a kind of global law based on the ideas of public rights. He claims that a federation of states based on justice and morality can help to maintain eternal peace. The emphasis is on structuring political relations in a way that upholds the moral standards that guide people. According to Kant's theory of the social contract, a moral society must be created in which people recognise each other as moral equals and work together to develop a just political and legal system. The values guiding this community uphold the moral worth and dignity of every individual. Consequently, Kant believed that the social contract is a moral system based on the principles of free will and reason, rather than just a set of rules (Nussbaum 2006).

The main difference between Hobbes and Locke is that Hobbes emphasises the need for strong authority to avoid chaos, while Locke focuses on the preservation of individual rights and mutual respect. Hobbes' theory is centred on security and order, while Locke's theory is based on the dignity of the individual and mutual rights. On the other hand, the main difference between Locke and Kant is that Locke's theory

emphasises the importance of property rights and mutual advantage, while Kant's theory emphasises moral autonomy and universal principles. Kant's approach is more concerned with creating a just society through rational and moral agreements rather than focussing solely on property and mutual benefit.

Due to the different approaches to formulating the social contract mentioned above, theories of the social contract can be categorized into three main variants: egoistic, hybrid, and Kantian models (Nussbaum: 2006). The egoistic variant includes theories that assume that individuals act primarily out of self-interest. According to this view, the social contract is concluded based on mutual benefits in terms of property and security. This approach focuses on making agreements that maximise individual advantage. As we have seen above, Hobbes falls into the egoistic variant of social contract theory. His model is focused on self-interest and mutual benefit that emphasises the need for strong authority to maintain security and order. Hobbes' social contract describes individuals uniting to escape a state of nature defined by chaos and uncertainty, prioritising their own security and stability. Egoistic models typically exclude entities or relationships where there is no recognisable mutual benefit. Hybrid theories combine elements from different approaches and combine aspects of egoism with other considerations. They often include the idea of mutual advantage, but also recognise additional factors such as fairness or moral obligations in the social contract. Locke can be categorised in a hybrid model. Although he emphasises individual rights and mutual benefit (elements of egoism), he also integrates ideas about moral duties and the importance of a just society. Locke's social contract includes both personal freedom and collective benefits and combines self-interest with principles of individual dignity and mutual benefit. Kantian theories, on the other hand, focus on moral autonomy and universal principles and aim to create a just society through rational and moral agreements rather than self-interest or mutual advantage. They emphasise that justice arises from a fair process and not from pre-existing conditions (Nussbaum: 2006).

Rawls further develops the concept of justice arising from a fair process, rather than from pre-existing conditions, in his *A Theory of Justice*. This work builds on Kantian ideas by examining how fairness, social cooperation, and state authority are linked. *A Theory of Justice* represents a groundbreaking approach in the field of political philosophy, as it departs significantly from the mentioned traditional view of how societies function. In this modern view, the concept of justice is essential at every stage of the formation of social contract. In other words, this newer view assumes that the social contract is the source of our rights and principles, not just the defence of our pre-existing rights. By imagining people as being free, equal, and able to make their own choices, Rawls (1971; 1999; 2001; 2005) addresses big questions about fairness and why governments have authority. I will now discuss his framework in more detail.

Rawls' theory of justice is better known as "justice as fairness", which is characterised by a liberal-egalitarian doctrine. It is "egalitarian" because everything revolves around fairness and equality. Everyone should have the same basic rights and opportunities. It is "liberal" because it refers to a set of ideas about freedom and individual rights. In a liberal society, individuals have certain freedoms, such as freedom of speech and the freedom to choose how they organise their lives. Egalitarian liberalism combines this idea of freedom with the principles of fairness and equality. In other words, the goal of egalitarian liberalism is a society in which everyone has the same basic rights and opportunities, and these rights and opportunities are combined with the freedom to make choices about their lives. It is about creating a balance between individual freedom and a fair, equal society. In such a system, the government and rules are designed to protect individuals' freedom and ensure fairness for everyone. The main idea of Rawls' theory, therefore, is to create a framework for a just society that ensures the fair distribution of rights, responsibilities, and resources.

This theory consists of two central principles of justice: the principle of freedom and the principle of equality:

I. The first principle (Equal basic liberties): Each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties; which scheme is compatible with the same scheme of liberties for all?

II. The second principle: Social and economic inequalities must satisfy two conditions: They are to be attached to offices and positions open to all under conditions of fair equality of opportunity; They are to be to the greatest benefit of the least-advantaged members of society (the difference principle) (Rawls 2001: 42–43).[10]

The latter principle is further subdivided into fair equality of opportunity and the principle of difference. The principle of fair equality of opportunity states that social and economic positions should be open to all individuals under the conditions of fair equality of opportunity. This means that everyone should have a fair chance to achieve positions of power and prestige, regardless of their background. The difference principle means that social and economic inequalities are only acceptable if they benefit the least favoured members of society. In other words, inequalities are justified if they improve the situation of the most disadvantaged people and make their lives better than would be the case if resources were distributed more evenly. Further, the second principle argues that inequalities in society (like differences in income and opportunities) should be connected to jobs and positions that anyone can try for, and

[10] It is important to note that these two principles are articulated in both Rawls' works *Political Liberalism* and *A Theory of Justice*. However, I emphasize here the revised version presented in his work *Justice as Fairness: A Restatement* (2001).

everyone should have a fair shot at these positions. In the context of the first principle, “fully adequate” means that these basic liberties should be good enough to protect and promote what Rawls calls our two moral powers – the capacity to be *rational* and the capacity to be *reasonable* in our interactions with others.

In Rawls' framework (2005), being *rational* means having the capacity to pursue one's own conception of the good and to make decisions that are consistent with one's rational preferences and interests. Consistent preferences mean that rational individuals have preferences that are coherent. They can rank their preferences and make decisions based on this ranking. Furthermore, rational individuals choose means by which they can achieve their goals. They can determine which actions will best help them to achieve their goals. To explain the importance of the capacity to be *reasonable*, it is first necessary to elaborate more on rationality—the individual capacity to form and pursue one's own conception of the good. In his later work, *Political Liberalism*, Rawls himself was drawn to the question of how people with different and sometimes contradictory conceptions of the good can live together in a fair and just political framework. He argues that individuals inherently have different comprehensive doctrines (aspects of a person's worldview, including religious, moral, cultural and intellectual beliefs) about what constitutes a good life. He (1995; 2005) refers to this as reasonable pluralism. It would be unfair to free and equal individuals to base the public justice system on a single view that embraces certain values, beliefs, or principles. Instead, public justice must be grounded in political ideals that all rational and reasonable people can agree upon. This means that, since we do not voluntarily choose to be part of political society, it cannot be based on a comprehensive doctrine. Thus, to navigate this pluralism, Rawls introduces the idea of "public reason" and a concept of rationality centered on reasonableness. In this way, the concept of reasonableness serves as a broader and more encompassing moral capacity—it is a capacity for a sense of justice. Being reasonable means that you are willing to engage in fair and impartial political reasoning. Reasonable individuals are open to considering the perspectives and interests of others and are willing to find common ground and compromise to achieve a just and fair society. Reasonableness involves a willingness to engage in moral dialogue and to adhere to principles of justice that can be accepted by all, not just those who benefit oneself. According to Rawls' theory (1999; 1995; 2001; 2005), the basic structure of society should be designed to protect and promote the capacities for rationality and reasonableness in all individuals. While rationality serves personal conceptions of the good, reasonableness is crucial for public affairs and the construction of a just society. Rawls envisions a just society as one that is well-ordered, where individuals, both rational and reasonable, cooperate to create a fair social order. In such a society, reasonable and rational individuals should support shared principles of justice that can be justified to others based on universally acceptable reasons, ensuring that rationality prioritises common principles that respect diverse conceptions of the good rather than individual comprehensive doctrines.

Because Rawls' vision of a political society is that of a fair system of cooperation across generations between free and equal individuals, he emphasizes the need for equal concepts of justice that recognize the position of citizens as free and equal. To understand how to interpret the concepts of freedom and equality, we must look to the political culture of a democratic society and its tradition of interpreting its constitution and basic laws for some essential principles that contribute to the formation of a vision of political justice (Rawls 1999). In this task, he invokes fundamental ideas: free and equal people and a well-ordered society guided by a public understanding of justice. Fair terms of cooperation are terms that "each participant can reasonably accept, and sometimes should accept, provided that all others accept them," which is one of the fundamental aspects of a fair system of cooperation. Fair terms of cooperation define the concepts of "reciprocity and mutuality" (Rawls, 2001: 6).

In this context, the concept of reciprocity plays a crucial role in ensuring that these fair terms of cooperation are genuinely fair and just. Reciprocity, as defined in Rawls' *Political Liberalism*, essentially means that when individuals or groups propose terms of fair cooperation, they should assume that these terms are fair not only to themselves but also to others. These terms should be proposed based on free and equal citizenship, without domination or manipulation, and without taking advantage of the disadvantaged positions of certain individuals or groups in society. In other words, reciprocity emphasises the idea that fair terms of cooperation should be acceptable to all parties and should not favour any particular group or enforce unequal power dynamics (Rawls, 2005: 446). It is important to emphasise that Rawls' concept of reciprocity differs from the concept of mutual advantage. Rawls explains that reciprocity lies between two other ideas: impartiality, which is altruistic, and mutual benefit, which is often understood to mean that each party benefits in proportion to the other party's present or expected future circumstances (Rawls, 2005: 16-17). Put more simply, Rawls emphasises that reciprocity is a concept in its own right. It is not just about the pursuit of mutual benefit, where each seeks to improve their own situation. Nor is it just about altruism, where individuals act solely for the benefit of others. Instead, reciprocity is about a balance between these extremes, where individuals consider both their own interests and the interests of others in a fair and equitable way. This approach aims to create a fair and mutually beneficial social framework.

Another crucial component of justice is the concept of rational advantage, which explains what individuals aim to achieve through co-operation. Rawls distinguishes between the rational and the reasonable, emphasising that reasonable people are willing to propose, accept or consider fair ideas in order to create a cooperative system that is just and accepted by all. Reasonable people recognise the importance of adhering to these principles as long as others do the same, but they are released from this obligation if others do not reciprocate. This dynamic ensures that the principles

of justice remain fair and enforceable, even if some people do not abide by them. The principles of justice set out the basic rights and obligations that govern the distribution of the benefits of social co-operation. They promote a common understanding that publicly recognises the basic structure and strengthens the individual's sense of justice. This sense enables people to understand and apply these principles while fulfilling their corresponding obligations as far as possible. Fairness and enforceability are not fundamentally undermined even when some individuals reasonably take advantage of favourable circumstances.

A crucial criterion for evaluating concepts of justice is their ability to serve as a publicly recognised framework in a society conceived as a system of cooperation between free and equal persons. While an ordered society based on a comprehensive doctrine is not feasible due to reasonable pluralism, an ordered society can exist if it is based on a political conception of justice (Rawls, 2005). Rawls' ideal theory envisions a perfectly just society and proposes principles that guide its structure and provide a basis for fairness and co-operation amidst diversity. This framework consists of key political and social institutions that are built through social co-operation, assigning rights and responsibilities while distributing the benefits of co-operation. These include elements such as the political constitution, an independent judiciary, the property system, the economic structure and the family. The central challenge of political justice is to ensure the just organisation of this basic structure. Although political justice does not directly regulate the activities of organisations or groups, it sets limits to them through background institutions. These institutions maintain fairness in daily life by shaping the basic framework of society. Fair negotiation requires preventing position distortions and power imbalances, ensuring agreements made under truly fair conditions remain just. Therefore, any definition of justice must begin with the concept of society as a fair system of co-operation between free and equal individuals that forms the basis for lasting fairness and collective harmony.

The above discussion introduces Rawls' (1999) thought experiment of the original position, which operationalises the concept of a fair process. In the original position, individuals possessing the capacities for reasonableness and rationality are imagined in a symmetrical relationship. In this hypothetical scenario, they are deprived of knowledge about their personal characteristics—such as wealth, gender, or social status—and are aware only of general social facts, such as resource scarcity and the limits of altruism. This state symbolises the veil of ignorance, a fair and impartial standpoint from which individuals can reflect on principles of justice. Rawls presents this model as the ideal perspective for reaching agreement on the fundamental structure of society. The veil of ignorance ensures that individuals set aside personal bias and select principles of justice that are fair for all, rather than advantageous only to themselves. By neutralising personal circumstances, it leads to decisions that prioritise equality and fairness. Rawls argues that rational actors behind this veil of

ignorance would agree on two central principles of justice: equal basic freedoms for all, and the difference principle, which permits social and economic inequalities only if they benefit the least advantaged members of society (1999).

When Rawls asks us to imagine a hypothetical social contract, he is essentially inviting us to consider a scenario in which individuals are equal. This scenario removes any existing advantages or privileges that might distort decision-making in real societies. The thought experiment is crucial as it raises a fundamental question about the legitimacy of governments: *Would people who are free, equal, and able to make their own decisions voluntarily consent to a government with certain rules and powers*? Rawls explores whether individuals, under these just conditions, would choose to form a government. This perspective offers insight into the justification for the existence of governments and the authority they exercise over us. In essence, Rawls' model of the original position operationalises the concept of a fair process. It envisions a state in which all individuals, possessing the two fundamental capacities for reasoning about justice—reasonableness and rationality—are in a symmetrical relationship. Through this framework, the principles of justice are determined, and their interpretation and implementation are guided by what Rawls refers to as public reason. The outcomes derived from this procedure are considered legitimate, as they are rooted in fairness and equality.

To summarise, Rawls' sophisticated and foundational framework for understanding justice, combining the principles outlined in *A Theory of Justice* and *Political Liberalism*, provides a strong basis for creating a fair and just society. In *A Theory of Justice*, Rawls (1971; 1999) presents a compelling argument for a social contract grounded in reciprocity, where individuals agree to the principles of a just society based on mutual fairness. In contrast, *Political Liberalism* (1995; 2005) explores how to establish a well-ordered society, focusing on the need for a just and fair political framework that can accommodate a diversity of reasonable views and values.

Despite its significant influence and robust theoretical foundation, Rawlsian theory has limitations due to its idealised assumptions. Critics such as Nussbaum (2006) highlight challenges in applying Rawlsian principles to complex real-world scenarios. These critiques address issues such as distributive justice in international relations, the treatment of individuals with cognitive disabilities, and the rights of non-human animals. These limitations underscore the difficulty of addressing the full spectrum of individuals' needs and experiences within Rawls' framework.

Nevertheless, I argue that Rawls' theory remains a vital tool for understanding and addressing social problems. His emphasis on fairness, equality, and individual freedom provides an essential and robust foundation for creating a just society. The concepts of the original position and the two principles of justice remain valuable for ideal theory, particularly when adapted to contemporary challenges. By revising and building upon it, the theory could better address the challenges faced by people with

severe cognitive disabilities, ensuring their full inclusion in justice. Although alternative approaches such as that of Nussbaum (2006) offer valuable perspectives, in the next sections I will argue that the politically liberal theory of the social contract offers a more comprehensive and established framework for addressing issues of justice and inclusion. In the following section, I will first analyse Nussbaum's alternative response to these challenges and evaluate its strengths and limitations in comparison to Rawls' theory.

1.3. Section Three: Alternative by Martha Nussbaum

In her influential book *Frontiers of Justice*, Nussbaum (2006) offers a profound critique of Rawls' philosophy, highlighting the shortcomings of the social contract tradition he advocates. While Rawls seeks a universal justification of justice through the social contract, Nussbaum argues that this approach fails to adequately address the needs and rights of individuals with disabilities.

Her critique focuses primarily on three key issues inherent in social contract theory:

i. the idea of associating those who participate in the formation of the social contract with those who fully enter the domain of justice.
ii. the idea that primarily equality in terms of power and strength determines the status of moral equality; the emphasis on mutual advantage; and
iii. the consequent difficulty of including individuals with mental and physical impairments and deficiencies.

A key element of social contract theory is the basic importance that all participants attach to rationality[11]. It goes without saying that individuals entering contracts must be rational. Consequently, social contract theory assumes cooperation between generally capable and equal individuals. However, this perspective overlooks the critical challenges faced by people with disabilities, who may be unable to participate equally in social cooperation. Moreover, it fails to account for the stigmatization and alienation these individuals often experience in contemporary societies. Reflecting on the realities faced by people with disabilities reveals a significant flaw in social contract theory: it conflates those who develop principles of justice with those for whom these principles are intended. The problem lies in the assumption that the individuals designing principles of justice are the same as those expected to live by them (Nussbaum, 2006: 16). Because these principles are based on mutual advantage, others—such as individuals with disabilities—can only be included later and in an indirect manner.

According to Nussbaum, it is not necessary to distinguish between the group of beings for whom principles are developed and the group of beings who develop principles. Precisely because the theory of the social contract makes this distinction, those who

[11] This includes what Rawls sees as the capacities of reasonableness and rationality, as described earlier — i.e., both are included in the "rationality" mentioned here.

were not involved in the development of the principle of justice are excluded from the scope of justice. Nussbaum argues that within Rawls's framework, it is apparently denied that there are any questions of fairness between people who have Kantian moral capacities (reasonableness and rationality) and people (or non-human animals) who do not. Thus, if people who lack the capacity to be reasonable and rational have any rights at all, it must be because they are the objects of "the interest and concern of Kantian rational beings" (Nussbaum 2006: 138). Therefore, according to Nussbaum, nothing in Rawls's theory guarantees that the interests of persons with severe cognitive disabilities are valuable for their own sake or that they are fairly considered in the formulation or selection of principles of justice. She (2006) argues that there is no need to exclude them. However, thinking about social justice has evolved in a way that creates a weak connection between them and rational individuals - who formulate justice primarily for themselves. As a result, addressing the needs of these "others" is often seen as merely a matter of goodwill.

Given all the points above, Nussbaum offers a proposal—the capabilities approach—that she believes represents a significant advancement over the social contract tradition. She argues that her approach provides a distinct perspective on human welfare and justice, offering a fresh and innovative contribution to political philosophy. At its core, the capabilities approach focuses on enhancing individuals' capabilities and opportunities, rather than merely addressing notions of entitlements and obligations. Before delving into Nussbaum's specific interpretation of the capabilities approach and how it presents a promising alternative to the conventional social contract tradition, I will first provide a brief overview of the approach in more general terms.

The capability approach[12] is an approach that denotes a broad umbrella concept. Namely, due to its multidisciplinary nature, the approach is utilized across various disciplines. Its flexibility is particularly evident in its application to global public health, where it informs efforts to address inequalities in healthcare access and outcomes. In development ethics, it shifts the focus toward human flourishing rather than mere economic growth (Robeyns, 2017). Moreover, the approach plays a crucial role in environmental protection by examining how ecological sustainability intersects with human capabilities (Robeyns, 2017). In education, it promotes curricula designed to enhance students' ability to lead meaningful lives (Murray, 2024). It also informs technological design, advocating for innovations that expand rather than constrain human potential (Oosterlaken, 2009), and guides welfare state

[12] I would like to thank Ana Gavran Miloš for drawing my attention to the fact that the term "capabilities approach" is mainly used in the works of Martha Nussbaum, who relies on a specific list of capabilities and therefore refers to them in the plural. Other authors, such as Ingrid Robeyns, who builds on the work of Amartya Sen, usually use the term in the singular – "capability approach" – emphasising a broader theoretical framework. Given the broader context discussed in this part, I use the singular form, which is also the title of the book I am referring to.

policy by prioritizing the actual freedoms and opportunities available to individuals over mere resource distribution (Robeyns, 2005). For instance, in addressing poverty, the capability approach moves beyond measuring income levels to assess whether individuals have access to adequate healthcare, education, and the ability to participate fully in society (Sen, 1999). This broad applicability justifies its characterization as an "umbrella term," encompassing diverse yet interconnected issues of human development and justice (Robeyns, 2017).

The significance of the approach is centred on the capabilities that seek to provide a structure for evaluating an individual's well-being in daily functioning. Capabilities are defined as an individual's effective potential to be and do something. More specifically, these are plausible options for a person to be or do something if they so desire in given circumstances. Besides capabilities, another important term used by all capability approach (CA) scholars is functionings. Functionings are defined as accomplished capabilities. Whether someone can transform a set of means—resources and public goods—into a functioning (i.e., whether she has a particular capability) is critically dependent on certain personal, socio-political, and environmental conditions, which are referred to as "conversion factors" in the capability literature. Capabilities, as opposed to mere formal rights and freedoms, have also been referred to as real or substantive freedoms, because they represent freedoms that are not hindered by obstacles. The focus of the capability approach is not on the functions a person has already achieved, but on their real freedom—that is, their ability or capacity to function and pursue what they value (Robeyns 2017). The fundamental focus of the capability approach is therefore on the capabilities of the individual and their affective freedom to be and do what they choose.[13]

Different versions of the capability approach (CAs) vary in how they determine which capabilities are most important. Nussbaum's approach aims to define a specific list of essential capabilities. In *Women and Human Development* (2000), she defines human capabilities as "what people are actually able to do and be." Nussbaum identifies ten central capabilities that are considered fundamental rights; that is, they must not be violated in the pursuit of other forms of social justice. This means that these capabilities must be safeguarded up to a certain threshold level (Robeyns, 2017).

The ten central capabilities include:

1. *"Life. Being able to live to the end of a human life of normal length; not dying prematurely, or before one's life is so reduced as to be not worth living.*

[13] A similar challenge arises when deciding which capabilities should serve as evaluative standards in areas such as psychiatry. This topic will be explored in more detail later in the thesis.

2. *Bodily Health. Being able to have good health, including reproductive health; to be adequately nourished; to have adequate shelter.*
3. *Bodily Integrity. Being able to move freely from place to place; to be secure against violent assault, including sexual assault and domestic violence; having opportunities for sexual satisfaction and for choice in matters of reproduction.*
4. *Senses, Imagination, and Thought. Being able to use the senses[14], to imagine, think, and reason—and to do these things in a "truly human" way, a way informed and cultivated by an adequate education, including, but by no means limited to, literacy and basic mathematical and scientific training. Being able to use imagination and thought in connection with experiencing and producing works and events of one's own choice, religious, literary, musical, and so forth. Being able to use one's mind in ways protected by guarantees of freedom of expression with respect to both political and artistic speech, and freedom of religious exercise. Being able to have pleasurable experiences and to avoid non-beneficial pain.*
5. *Emotions. Being able to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; in general, to love, to grieve, to experience longing, gratitude, and justified anger. Not having one's emotional development blighted by fear and anxiety. (Supporting this capability means supporting forms of human association that can be shown to be crucial in their development.)*
6. *Practical Reason. Being able to form a conception of the good and to engage in critical reflection about the planning of one's life. (This entails protection for the liberty of conscience and religious observance.)*
7. *Affiliation. Being able to live with and toward others, to recognize and show concern for other humans, to engage in various forms of*

[14] It is important to note that Nussbaum's list of capabilities has evolved over time. In earlier versions, it included "use *all* of one's senses," which implied that a person who was deaf would be considered disabled, as all five senses were required. However, in later versions, the framework allows for more inclusive definitions, such as Jessica Begon's (2023) argument that one can be deaf without being disabled. See more in: Nussbaum, M. (1988). *Nature, function and capability: Aristotle on political distribution. Oxford Studies in Ancient Philosophy, Supplementary Volume* (Vol. 6), 145–184. Oxford: Clarendon Press. https://changingminds.org/explanations/needs/nussbaum_capabilities.htm

social interaction; to be able to imagine the situation of another. (Protecting this capability means protecting institutions that constitute and nourish such forms of affiliation, and also protecting the freedom of assembly and political speech.)

Having the social bases of self-respect and non-humiliation; being able to be treated as a dignified being whose worth is equal to that of others. This entails provisions of non-discrimination on the basis of race, sex, sexual orientation, ethnicity, caste, religion, national origin and species.

8. *Other Species. Being able to live with concern for and in relation to animals, plants, and the world of nature.*
9. *Play. Being able to laugh, to play, to enjoy recreational activities.*
10. *Control over one's Environment.*

Political. Being able to participate effectively in political choices that govern one's life; having the right of political participation, protections of free speech and association.

Material. Being able to hold property (both land and movable goods), and having property rights on an equal basis with others; having the right to seek employment on an equal basis with others; having the freedom from unwarranted search and seizure. In work, being able to work as a human, exercising practical reason and entering into meaningful relationships of mutual recognition with other workers." (Nussbaum 2000: 78-70).

The political liberal particularity is evident in the justification of the idea of human dignity. Nussbaum based this idea on overlapping consensus, which refers to an agreement among different moral or philosophical doctrines on certain principles of justice, despite their diverse foundational beliefs. In the idea of dignity, it attempts to provide a strong justification for a specific list of ten central capabilities that it considers essential for a life of dignity. For Nussbaum, the notion of human dignity is not an abstract or empty concept, but one that should be based on concrete and measurable aspects of human well-being. Nussbaum's concept of dignity is not grounded in the Kantian understanding of rationality; rather, it is Aristotelian. It is based on the recognition that, in addition to being rational beings, we are also needy and vulnerable, as we have bodies and other essential needs. The list of capabilities, therefore, emerges as a reflection of what is necessary to live a life worthy of human dignity. These ten core capabilities, outlined in her work, provide a concrete framework for understanding and promoting human well-being, making the concept of human dignity more tangible and achievable (Nussbaum, 2006: 75).

As noted above, Nussbaum's earlier thoughts on the capabilities approach were strongly influenced by Aristotle and were considered perfectionist[15]. However, she later tried to refine her theory and align it with political liberalism (2006). In this development, capabilities are presented as the foundation for political principles within a liberal, pluralistic society. The framework is rooted in political liberalism, and it deliberately avoids a metaphysical foundation (Nussbaum, 2006: 70). According to Nussbaum, capabilities should be guaranteed up to the threshold required for truly dignified human life. These capabilities are not mere philosophical abstractions; rather, they should serve as practical guidelines for societies and governments. Nussbaum (2006) argues that these capabilities represent a minimum threshold that must be respected and upheld by governments worldwide. In this way, her theory establishes a set of standards that should be universally recognised, as they express an overlapping consensus. They must be protected to ensure a dignified life for all individuals.

Capabilities are complex concepts, and Nussbaum distinguishes three concepts of human capabilities: basic capabilities, internal capabilities, and combined capabilities (Robeyns 2017). Basic capabilities are defined as the innate equipment of an individual, which is necessary for the development of more advanced capabilities such as speech and language. They need to be nurtured before these capabilities can develop into true capability. Internal capabilities represent the internal aspect of the capability, that is, the prerequisites for fulfilling that capability. If we have the skills and meet the physical prerequisites for running, we may or may not run a marathon. If suitable external conditions are established, then we are talking about combined capabilities (Robeyns 2017). To clarify further, I will provide an example for each of the types of capabilities mentioned. The capacity to perceive the five senses (sight, hearing, touch, taste, and smell) is a basic capability. Without this innate equipment, it would be impossible for individuals to develop more advanced capabilities such as recognizing and interpreting symbols and signs in language. A common example of this internal capability is a learned language. When I acquire it under the appropriate circumstances, it becomes activated or developed, similar to learning to walk. It is, therefore, a developed state ready to be activated as a functioning. For Nussbaum, capabilities in the truest sense are a combination of internal and external capabilities. When external circumstances are favorable for activating internal capabilities — meaning there are no barriers preventing or limiting them—then we can speak of external capabilities. For example, public speaking is a combined capability that requires both internal and external factors. Internally, it involves having the capability

[15] In this context, perfectionism refers to a normative approach in political philosophy that assume the state should promote objective values or human excellences for a good life. Nussbaum's early capabilities approach, influenced by Aristotelian ethics, was seen as perfectionist for prescribing essential capabilities for human flourishing. However, she later revised it to emphasize individual autonomy in exercising these capabilities.

to organize thoughts coherently, speak confidently, and possess knowledge of the language (all internal capabilities). Externally, it requires an appropriate audience, a platform to speak, and the freedom of speech, ensuring that, for example, a woman is not discriminated against while speaking (external conditions). Without both the internal and external elements coming together, one cannot effectively engage in public speaking.

There are clear parallels with political liberal form of theory of the social contract, but unlike the social contract theory, Nussbaum (2006) argues that her theory makes a significant contribution by broadening the scope of fairness. The key distinction between her capabilities approach and social contract theory is that social contract theory is procedural, while Nussbaum's approach is content-focused. In social contract theory, fairness is achieved through the procedures that set the conditions for selecting evaluative standards, which are meant to ensure fairness and impartiality[16]. All the principles that emerge from this method are considered just. There is no independent criterion for determining the correctness of a finding. Nussbaum's capabilities approach begins with an intuitive understanding of what is inseparably connected to a dignified human life (Nussbaum, 2006: 82). Procedures are deemed valid if they lead to the correct content (Nussbaum, 2006: 82). In other words, the underlying premise of the capabilities approach is that what matters for justice is the quality of human existence. A procedure that contradicts our intuitions about dignity and justice is considered invalid. Nussbaum's main critique of social contract theory is that reason alone cannot fulfill the essential task of developing a credible theory of justice without also incorporating some understanding of the good.

The basic premise of Nussbaum's approach is that justice is defined by considering the capabilities with which each individual should be provided. In her interpretation of capability theory, these capabilities are determined with regard to the idea of dignity, which is rooted in species membership—in mind. Dignity, and by extension justice and rights, are based on the unique characteristics of each species. Nussbaum rejects the idea that justice should be based solely on a characteristic such as rationality, which is inherent in human beings. Instead, she anchors dignity in a broader range of capabilities that are essential to human flourishing (Nussbaum, 2006a: 326). In this way, she extends the scope of justice to individuals who may not have full rationality, such as people with severe cognitive disabilities. Her approach centres on a shared criterion: the capabilities that enable individuals to act and become who they can be, tailored to their specific circumstances.

[16] Here I follow Nussbaum, who emphasizes the term "impartiality." However, for the further elaboration of the dissertation, it is important to highlight the difference between the concepts of impartiality and reciprocity. Impartiality could be defined as a state in which both my own and other people's interests are equally important to me, while reciprocity is a state in which the interests of others are important to me, but under the condition that my interests are also important to them (Hartley 2014).

To sum up, Nussbaum's capability approach offers a distinct perspective on justice, focusing on the individual's capabilities and placing human dignity at the heart of its framework. By emphasising the inherent potential of every person, it aims to provide a more inclusive basis for determining justice and the distribution of rights, intending to create a society that respects and nurtures the diverse capabilities of its members. However, while Nussbaum's theory brings valuable insights, it has also faced significant criticism. In the following chapter, I will critically examine objections to Nussbaum's approach.

1.4. Section Four: Why Martha Nussbaum's Alternative Doesn't Work

Nussbaum's capabilities approach, which grounds the dignity and rights of individuals in a specific set of essential capabilities, has attracted substantial criticism, particularly regarding its application to the diversity of human and non-human experiences. While Nussbaum's list of capabilities is designed to be flexible, allowing for the "multiple realisability" of different conceptions of the good life, critics (Sen 2004; Robeyns 2017; Claassen 2014; Begon 2023) argue that the very existence of a universal list still imposes a narrow view of human flourishing. This occurs despite the flexibility she intends, as it may unintentionally limit the scope of justice by presenting a vision of human well-being that does not fully reflect the wide variety of human and non-human experiences. This paragraph explores key objections to Nussbaum's theory, focusing on its normative limitations, insensitivity to pluralism, and the challenges of grounding rights in species membership. The critique of Nussbaum's list can be divided into two main strands: one critiques the content of the list itself, arguing that it is exclusionary, and the other addresses the philosophical justification of the list, particularly in terms of universalism and species-based norms. Both strands highlight the limitations of imposing a singular, universal approach to justice that does not adequately account for diversity.

As mentioned, one significant critique of Nussbaum's approach comes from the content of her list of essential capabilities. A major criticism is that Nussbaum's framework runs the risk of being too rigid to meet the diverse and changing needs and values of individuals, particularly individuals with disabilities (Begon 2023; Claassen 2014). By relying on a fixed list of capabilities, her framework may overlook the fact that individual needs and values can vary greatly and change over time, so a system that is more sensitive to pluralism is needed to ensure that everyone is included and treated fairly. Nussbaum's account of human flourishing appears overly detailed and perfectionist because it propagates a particular vision of the good life (Schuppert 2014: 71)

In contrast to Nussbaum, Amartya Sen (2004) rejects the idea of a fixed list of capabilities. His capability approach emphasizes procedural flexibility, focusing on individuals' freedom to make choices and pursue their own goals. Sen prioritizes the process of achieving well-being over prescribing specific outcomes, allowing for

adaptation to diverse needs and preferences. Namely, he (2004; 2005; 2009) argues that capabilities should be defined through public deliberation and democratic consultation, rather than through a fixed list. Sen stresses that in pluralistic societies, conceptions of well-being are shaped by diverse cultural, social, and political influences. Insisting on a universal set of capabilities can lead to dogmatism, as it fails to accommodate evolving social values and the complexities of individual lives (Sen, 2004). By advocating for a process that reflects societal values and allows for ongoing adjustments, Sen's approach offers greater flexibility, focusing on preventing severe deprivations like poverty and highlighting the need for context-sensitive assessments of capabilities.

However, even Sen's emphasis on democratic deliberation faces its own challenges. Public decision-making processes are inevitably shaped by societal constraints and biases, which can undermine fairness and effectiveness. As Baccarini and Lekić Barunčić (2023) argue, public standards of justification and evaluation are influenced by various perspectives and contextual factors, making it difficult to establish truly impartial standards of justice. These challenges demonstrate that even within a democratic framework, achieving fairness in the assessment of capabilities is complex and subject to real-world limitations.

A further objection to Nussbaum's approach is based on her reliance on species membership as the foundation of human dignity. Nussbaum argues that human dignity is tied to the possession of species-typical capabilities, which she considers essential for a flourishing life. Shane Glackin (2016) sharpens this critique, pointing out that the concept of species is fluid and arbitrary and thus we cannot speak about species-typical capabilities. Evolutionary theory shows that species are not fixed entities but categories of organisms with shared traits (Rachels: 1987). Therefore, grounding rights in species membership is problematic because it ignores the diversity of individual experiences and capacities within a species. Glackin (2016) stresses that rights should be based on individual characteristics rather than species membership, as each person possesses unique qualities that cannot be captured by a general, species-based standard. Moreover, Glackin argues that there is no inherent reason to prefer certain capacities over others in every situation. For instance, we cannot assume that hearing is inherently better than being deaf, as in the case of individuals from the Deaf community (2016: 7).[17] Glackin uses an example from science fiction to illustrate his point, referring to the concept of "remaking" in literature. "Remaking" is an involuntary type of body modification used to punish criminals. It involved various grotesque disfigurements, such as twisting the neck 180 degrees so the criminal would always be seen behind him, or surgically attaching body modifications ranging from

[17] I use the term "capacities" in later sections to differentiate my viewpoints from the specific terminology of the capability approach. In this section, however, "capacities" and "capabilities" are used interchangeably and mean the same thing. Here, Glackin is referring to what Nussbaum calls "capabilities."

extra limbs to steam-powered iron wheels in place of their legs. One of the criminals who experienced the punishment of mutilation of the body was sent to the colony for life-long slavery. However, the ship that transported him to this enslavement was captured by a floating city, whose inhabitants were also taken prisoner, and the said criminal and the other criminals on board were forced to remain in the service of this city. Namely, this makes a difference for prisoners. Instead of being sent to prison, they, like all the inhabitants of the floating city, are captured, but while there, they are accepted as free and equal citizens. By finding maintenance work under the city, the criminal was able to make productive use of his disfigured body and was treated with dignity. Glackin (2016) states that, over time, the criminal has become a different kind of human being. Namely, although the majority will experience his process as permanent humiliation and torture, which is the point of such punishment, the quality and dignity of his life have been greatly improved precisely by the changes that deprive him of his alleged specific dignity. However unintentional (and unjust) his initial transformation was, the said criminal has ceased to judge his biological health by the 'form of life typical of his species'. The only "form of life" now relevant to any such judgment is his own; the fact that human lives usually go best in a certain way is of no particular importance to him (Glackin 2016: 7). This example highlights how individuals who deviate from species norms may still live dignified and fulfilling lives, showing that dignity should not be tied to the possession of species-typical characteristics.

In line with this, Glackin suggests that deviations from species norms should not diminish an individual's dignity or worth. This perspective challenges Nussbaum's assumption that there is a universal standard of human dignity based on species-typical characteristics. He argues that people who deviate from these norms—whether due to disability, personal choice, or cultural factors—should not be seen as less dignified or deserving of fewer rights. This critique points to an implicit assumption in Nussbaum's framework that may lead to disabled individuals being viewed as "diminished" versions of able-bodied individuals, rather than recognising their differences as legitimate and valuable (2016: 10-12).

Furthermore, Glackin criticises Nussbaum for not sufficiently considering the diversity of views held by people with disabilities, many of whom do not perceive their disability as a deficit. A truly liberal approach, according to Glackin (2016), should support an open, negotiated understanding of human capabilities that respects individual differences and avoids imposing a rigid, species-based norm. This critique aligns with the broader challenge to Nussbaum's approach, which is seen as failing to respect the plurality of human experiences, particularly with regard to disability.

Glackin's criticisms are partly replied by Jessica Begon (2023). In particular, it opposes the assumption that disabilities are inherently associated with impairments, which are meant as atypical characteristics in her definition. Begon (2023) argues that whether an impairment leads to a disability depends on broader societal contexts and

evaluations of individuals. In societies where impairments can be compensated for by changes in the environment or by the support of helpers, they may not lead to disability at all. Moreover, some people may not positively assess certain impaired capabilities, thus not identifying them as disabilities. This view, Begon contends, contrasts with species-determined perspectives like Nussbaum's, which rely on a set of universal, valuable human functionings. By emphasizing the variety of valid human functionings, Begon critiques the imposition of a singular list of capabilities, as it risks excluding individuals who do not value or cannot perform these functionings, while also undermining their authority over their own conception of the good. For example, deaf and blind individuals may reject neurotypical social interaction, and asexual individuals may repudiate sexual satisfaction. In focusing on a specific list of valuable functionings, Nussbaum's approach risks violating individual autonomy and may lead to the imposition of what is considered "best" for all, rather than supporting more content-neutral freedom. Begon's position encourages a reflective process to critically revise the current scope of disability, questioning whether it accurately captures the diverse and self-determined nature of human functionings (2023: 217).

Views such as those of Glackin's and Begon's, which attribute a fundamental role to the variety of personal perspectives, are challenged through the concept of adaptive preferences. This concept suggests that individuals who have adapted to deprivation or disability may not recognise what is really good for them. Contrary to this assumption, Elizabeth Barnes (2016) challenges Nussbaum's reliance on the concept of "adaptive preferences." Barnes warns that an over-reliance on this idea can lead to paternalism, which undermines the autonomy of individuals by assuming that they need guidance to better understand their wellbeing. She argues that this perspective ignores the lived reality of people with disabilities who can find meaning and value in their own experiences, even if those experiences do not conform to a predetermined standard of wellbeing. Her critique challenges the assumption that everyone can or should aspire to the same capabilities and emphasises the importance of recognising the diverse forms of human flourishing. Barnes argues that people with disabilities may have their own conceptions of the good life that should be respected, rather than imposed by a universal standard of capabilities.

In line with the above criticism of Nussbaum's theory being non-pluralistic and possibly paternalistic, Rutger Claassen (2014) criticises Nussbaum's capabilities approach, arguing that it imposes an overly standardised vision of human flourishing, potentially leading to excessive paternalism. Claassen's concern is that Nussbaum's framework risks prescribing too rigid set of functionings, thereby limiting individual autonomy. In order to refute Nussbaum's theory, he pictures illustrations of possible persons who are impaired in some of the components of Nussbaum's capabilities list, but their life still goes well according to their perspectives. Claassen gives an example of a person who lacks the capability for play but leads a fulfilling life through engaging in analytical and intellectual activities. He also considers an individual who

does not form deep emotional relationships yet finds satisfaction in solitude, dedicating themselves to personal projects and meditation. Another example is a person without sensory abilities such as hearing or sight, who nonetheless manages to achieve their own vision of a good life by adapting their interests and activities to their abilities. Claassen's critique highlights the danger of Nussbaum's theory becoming too rigid in its definition of human flourishing, leaving insufficient room for individual differences and personal preferences (Claassen 2014: 60-62).

In conclusion, while Nussbaum's capabilities approach offers a compelling framework for addressing human dignity and rights, it faces significant challenges. The reliance on a fixed list of essential capabilities risks imposing a narrow, culturally specific view of human flourishing, which may not reflect the diversity of values and experiences across societies. Furthermore, grounding dignity in species membership and focusing on species-typical characteristics is problematic, as it overlooks individual variations within a species and the social construction of categories like disability.

To address these limitations, further discussion is needed to refine and develop a more inclusive framework that respects diversity and better meets the complex needs of all individuals in society. In the following sections, I will explore potential solutions and responses to Nussbaum's approach, aiming to propose a more inclusive model of justice that embraces pluralism.

1.5. Section Five: Solution and replies to Nussbaum's alternative

In this section, I will analyse several approaches that address the limitations of Nussbaum's framework in the context of justice and the inclusion of individuals with severe cognitive disabilities. I will examine the theories proposed by Gabriele Badano (2014), Henry Richardson (2006), Samuel Freeman (2018) and Cynthia Stark (2007).

The reason I engage with these theories is to provide a broader context for understanding approaches to justice that include individuals with severe cognitive disabilities and to highlight different attempts to integrate these individuals into theoretical frameworks of justice. I will critically analyse their proposals, point out their strengths and weaknesses, and then offer my own solution based on the principles of political liberal inclusion. This analysis will contribute to a deeper understanding of the ways in which different theories address the issue of justice for individuals who are unable to actively participate in social and political processes, while also serving as a foundation for the further development of inclusive approaches.

I will proceed as follows: I will first analyse the theory of Gabriele Badano (2014) and then move on to the theories of Henry Richardson (2006), Samuel Freeman (2018) and Cynthia Stark (2007). Finally, I will present my own solution, which is based on the principles of political liberal inclusion.

A. Gabriele Badano: Minimal reasonableness and minimal rationality as constituents of personhood

Gabriele Badano (2014) copes with a critical challenge within political philosophy: the need to ensure that individuals who may lack the capacity for reasonableness and rationality are not unjustly excluded from the scope of justice. Similarly, to Nussbaum's focus, Badano's central contention is that John Rawls' political liberalism falls short in recognizing the status and demands of many individuals with disabilities, and he argues that mere extensions of Rawls' theory won't repair this issue. Instead, Badano asserts that the demands of individuals with disabilities must be considered as a matter of political justice, necessitating a fundamental revision of political liberalism. To address this pressing concern, Badano proposes a significant departure from Rawlsian principles by redefining the structuring principle of rights. He suggests that by fixing the concept of person in their minimum capacity for moral powers, namely reasonableness and rationality, we can encompass a broader range of individuals within the framework of justice. Badano argues that the key lies in re-conceptualizing the very nature of moral powers (2014).

In contrast to the Rawlsian concept of reasonableness, which refers to the capacities required for justice and public reasoning, Badano (2014) considers this approach to be too limited. He argues that the idea that only reasonable and rational people should debate policy acceptability should be discarded. Many individuals, especially those with cognitive disabilities, may not fulfil Rawls' strict criteria for reasonableness in both senses. To address this gap and foster inclusivity, Badano introduces the concept of "minimally reasonable" persons. These are individuals who, although they may not meet the stringent criteria of full reasonableness, can recognize at least some blatant instances of wrongdoing (Badano 2014: 416). In this way, Badano proposes a more accommodating and inclusive understanding of reasonableness, enabling the participation of a broader spectrum of individuals, including those with disabilities, in the realm of justice. In analysing Badano's perspective, it becomes clear that he seeks to strike a balance between the core principles of political liberalism and the imperative of accommodating individuals with disabilities. By redefining the requirements for reasonableness, Badano endeavours to ensure that justice is not an exclusive privilege but a right extended to all members of society, regardless of their cognitive capacities. This approach reflects a commitment to addressing the challenges of reasonable pluralism while upholding the principles of fairness and justice.

Moreover, Badano's alternative to Rawls' concept of rationality significantly differs from Rawls' thesis. He suggests that some key assumptions in Rawls' idea of rationality should be reconsidered. One such assumption is the idea that individuals modify their goals considering their evolving understanding of what constitutes the good. Additionally, Badano suggests that the notion of individuals choosing the most effective means to achieve their goals should be replaced with a focus on the most

likely choice (2014). As a result of these revisions, Badano (2014) introduces the concept of "minimally rational" person. These individuals possess their own set of goals or interests and demonstrate a desire to see at least some of these aims realized. In other words, minimally rational individuals exhibit a willingness to engage in activities aimed at fulfilling some of their goals, or they express a desire for these goals to be achieved through actions that may be carried out by someone else. The significance of Badano's redefinition of rationality becomes clear in the context of the debate on the scope of justice described above, i.e. within the framework of political liberalism. Political liberalism should include individuals with severe cognitive disabilities and be more orientated towards their experiences and capacities. By shifting the criteria for rationality away from complex goal modification and the selection of the most effective means, Badano (2014) opens the door for many people with disabilities to meet the standards of minimal rationality. For instance, in basic areas such as nourishment, emotional connection, and pain management, individuals with cognitive disabilities can demonstrate a clear desire for the realization of their goals. His redefinition of rationality as "minimally rational" thus aims to ensure that persons with limited cognitive capacities are not unfairly excluded from the principles of justice. This reinterpretation also emphasises the adaptability and responsiveness of political theories such as Rawlsian liberalism in the face of the imperative of justice for all, regardless of cognitive capacities. It also emphasises the importance of recognising the diverse ways in which individuals, including those with disabilities, pursue their goals and well-being, which may differ from traditional notions of rationality.

However, this revision is not a rejection of Rawlsian political liberal perspective but rather a refinement that aligns it more closely with the principles of fairness, equality, and justice for all, regardless of their cognitive capacities. In line with this, Badano contends that all reasonable doctrines would accept his inclusive view of personhood:

> *"The religious doctrine would regard those with the minimal moral powers as persons because they may have a role to play in God's plan, because they bear the image of God, or based on analogous considerations. Other doctrines would stress that those with the minimal moral powers are sentient beings (classic utilitarianism), or beings who add to the variety of walks of life individuals can draw from (Mill's liberalism)"* (Badano, 2014: 417).

Badano asserts that his revised approach to political liberalism addresses a critical shortcoming in Rawls' framework. Specifically, Badano contends that his revision is more effective than Rawls' original proposal in ensuring that rights, opportunities, and distributive shares are extended to individuals who do not fit the framework of fully cooperative members within society.

Badano's (2014) argument has several advantages over Nussbaum's framework. It prioritizes inclusivity and social justice, asserting that a just society should not exclude individuals based on their cognitive abilities. His revision seeks to reconcile the ideals of political liberalism with the diverse realities of human cognitive differences. However, a key challenge remains: the potential ambiguity or exclusivity in defining who is included within the scope of justice. In other words, Badano's concept of minimal reasonableness, which requires a basic consideration of others, may be too restrictive or unclear when it comes to defining who should be included in the scope of justice. This lack of clarity in identifying minimally reasonable individuals could lead to arbitrary exclusions and inconsistencies in how justice is applied. The lack of clarity in determining who should be considered minimally reasonable thus represents a potential disadvantage of his approach.

To illustrate this, consider the first category in which Badano's criterion appears to be problematic, and that is the case of psychopathy. Some scholars argue that psychopaths may fail to meet the minimal reasonableness criteria set by Badano (2014) because they show a reduced level of moral concern for issues of harm and fairness (McGuire et al. 2014: 495-508). This exclusion is problematic because it unfairly denies justice to individuals who, despite their moral deficiencies, still require protection and consideration within the framework of justice.[18] Nadelhoffer and Sinnott-Armstrong (2013), further highlight this issue by emphasizing the challenges in attributing moral responsibility to those suffering from psychopathy. They point out that these individuals often have a diminished capacity for empathy and moral sensitivity, suggesting that their level of moral responsibility may not be comparable to that of individuals without psychopathy. This perspective reinforces the concern that excluding psychopathic individuals from justice based on minimal reasonableness criteria may be unjust, as it fails to account for their unique psychological condition. However, there are less straightforward, additional scenarios that can illustrate why Badano's (2014) theory is not fully satisfactory. Consider individuals with schizophrenia. A study suggests that individuals with schizophrenia exhibit subtle impairments in their behaviours related to fairness, but not necessarily in their willingness to engage in altruistic punishment (McGuire et al. 2014: 495–508). Certain authors, such as Glannon (264-265: 1997), are of the opinion that a schizophrenic who commits an act of violence during a psychosis is not responsible for this act because of the psychosis. Some other scholars are of the opinion that

[18] Including individuals with psychopathy in the scope of justice is essential, as their exclusion challenges the core principles of fairness and equality. Focusing on inclusivity ensures that all individuals, regardless of mental health conditions, are treated with dignity, and that the justice system remains responsive to diverse needs. The approach I support emphasizes justifying actions based on a clear understanding of their condition and its societal impact, rather than excluding individuals due to their differences. This argument will be further developed in a paper with Elvio Baccarini and Luca Malatesti.

individuals with schizophrenia can still possess at least a minimal level of reasonableness (Saks 2007, Cooper 2002). In any case, the situation of individuals with schizophrenia reveals an inherent ambiguity in Badano's (2014) definition of personhood. In the context of political liberalism, it is unlikely to be accepted that both those with psychopathy and those with schizophrenia should be excluded from the scope of justice if they cannot demonstrate adequate minimal reasonableness according to Badano's (2014) criteria. Moreover, there are additional circumstances in which individuals might be perceived as lacking minimal reasonableness, such as individuals suffering from severe dementia (Mendez et al 2005). Again, the claim that they have no claim to justice is problematic, at least when considered within the framework of political liberalism.

To summarise, the application of the Badano criterion in cases where individuals do not clearly fit into the category of minimal reasonableness underlines its inadequacy. The complexity of psychopathy, schizophrenia and severe dementia shows how difficult it is to draw clear boundaries for inclusion within the scope of justice. This complexity requires a more pluralistic and context-specific approach to ensure the main goal: that justice is accessible to all individuals, regardless of their cognitive conditions.

In addition to the problem of applying Badano's "minimal reasonableness" mentioned above, analogous challenges arise when we consider his concept of "minimal rationality". As a reminder, Badano (2014) defines minimally rational individuals as individuals who have their own purposes or interests, suggesting that they wish to see at least some of these purposes fulfilled and are willing to take actions themselves or wish others to take actions that fulfil these interests. To illustrate the complexity of this criterion, let us examine a scenario involving children with disabilities who display uncontrollable hostility towards dental procedures intended for their benefit. In such cases, the opposition to these procedures may be so intense and rigid that administering general anaesthesia becomes necessary (Schulz-Weidner, Nelly, et al. 2022). Now, we can say that these children have their own goals or interests (to be saved from pain) and want these to be satisfied. However, what does it mean to say that they want some individuals to perform actions to achieve this? The criterion is, to say the least, indeterminate or even unfulfilled by these individuals, in a sense that is not immediately obvious. This example highlights the challenges posed by the criterion of minimal rationality, especially in situations where individuals, due to their specific conditions or circumstances, may not conform neatly to the criteria set forth by Badano (2014). The inherent ambiguity in defining whose interests and actions qualify for minimal rationality underscores the need for a more nuanced and adaptable approach when addressing the diverse ways in which individuals pursue their interests and well-being. In view of the counterexamples, a different strategy is required, which must depart from the traditional models that presuppose certain characteristics as a precondition for inclusion in the scope of justice. These traditional models function

as a membership card for the “club of justice”, which admits individuals who fulfil certain criteria, such as being minimally reasonable or minimally rational, or simply being a member of the human species. Those who do not fulfil these criteria are excluded from the scope of justice by default.

In the next two sections, I will analyse two approaches that appeal to the normative strength of humanity. The first is the theory of Henry Richardson (2006) and the second is by Samuel Freeman (2018). I will argue that this is not a fully satisfactory solution because it does not respect the pluralism of reasonable conceptions of justice.

B. Henry Richardson’s Approach: Adaptation of Original Position for Individuals with Severe Cognitive Disabilities

Richardson (2006) suggests that we can include individuals with severe cognitive disabilities in the framework of justice by modifying Rawls’s concept of the *original position*. To remind, the original position is a thought experiment where representatives, acting as impartial legislators, choose the principles of justice behind a "veil of ignorance." This veil ensures fairness by preventing the legislators from knowing any personal details about themselves or the individuals they represent (such as their socioeconomic status, talents, or beliefs). This method is intended to help create fair principles of justice, as the legislators would have to create rules that would be just for everyone, regardless of their particular circumstances. Richardson’s innovation is that he explicitly includes individuals with severe cognitive disabilities among the individuals represented in the original position. This is a response to the discomfort of their exclusion from justice discussions and aims to ensure that their interests are represented when principles of justice are created.

He (2006: 443-444) provides three ways to adapt Rawls’s principles to be more inclusive of individuals with disabilities, including severe cognitive disabilities:

1. *A simple extension of Rawls’s principles to accommodate those with severe disabilities.*
2. *A set of principles inspired by Nussbaum's capabilities approach, which emphasizes ensuring that all individuals have the capabilities to live a life of dignity.*
3. *A hybrid approach that integrates elements from both Rawls’s and Nussbaum’s frameworks* (Richardson 2006: 443-444).

Richardson (2006: 439) also argues that issues of disability should be addressed at the constitutional stage, rather than just the legislative stage, to ensure a more comprehensive approach to justice. In this context, he introduces two new principles:

1. NPG (Needs-based Primary Goods): This principle guarantees that every citizen has a decent minimum of opportunity, income, wealth, and self-respect.

2. NC (Needs-based Capabilities): This principle replaces Rawls's primary goods[19] with Nussbaum's ten central capabilities, ensuring a minimum threshold of well-being for everyone, regardless of their capacities.

In addition, Richardson incorporates the concept of "species-typical functioning". This concept integrates health and disability into a broader justice framework by focusing on the basic capabilities that people should have in order to live healthy, fulfilling lives.

Although Richardson does not propose a final solution, he wants to show how the combination of Rawls' and Nussbaum's approaches can lead to more comprehensive theories of justice. He uses the Initial Choice Situation (ICS) to explore how different sets of principles might address justice for people with disabilities, weighing the trade-offs between Rawls', Nussbaum's and hybrid approaches.

Finally, Richardson (2006) challenges the ideal of reciprocity, which was central to Rawls's early work but is less emphasized in his later work. Reciprocity suggests that justice relies on mutual cooperation among individuals. Richardson argues that this principle is not essential for a just society and should not exclude individuals who cannot participate in cooperation, such as those with severe cognitive disabilities. Instead, Richardson proposes that justice should be based on the inherent worth and dignity of all individuals, not on their capacity to cooperate. This shift allows for a more inclusive understanding of justice, ensuring that individuals with disabilities are not left out of the moral and political community.

Richardson's proposal to include persons with severe cognitive disabilities in theory of justice encounters significant objections. In particular, it encounters similar challenges to Freeman's (2018) that I will present later. I argue that Richardson's view, which bases justice on the idea of shared humanity, does not adequately account for pluralism. The problem is that certain comprehensive doctrines that include typical features of reasonableness reject the idea that membership in the human species is a normatively relevant basis for justice. Critics of speciesism, such as Singer (2009), for example, claim that moral considerations should not be constrained by species boundaries.

[19] Rawls's (1999) primary goods refer to the basic goods and resources that individuals need to pursue their conception of the good life and live a flourishing life in a just society. These goods are fundamental to well-being and freedom, and Rawls argues that they should be distributed fairly according to the principles of justice. They include: (1) basic rights and liberties, such as freedom of speech and the right to vote; (2) opportunity, which ensures fair access to education and opportunities for self-development; (3) income and wealth, providing the material resources necessary for meeting needs and desires; and (4) the social basis of self-respect, which ensures recognition of one's dignity. Rawls's difference principle allows for inequalities in these goods only if they benefit the least advantaged members of society.

This objection is also supported by the findings of evolutionary theory, as I argued in section four in my critique of Nussbaum's theory. Evolutionary theory shows that species are not fixed or essential categories, but rather fluid groupings of organisms with overlapping characteristics (Rachels, 1987). This undermines the justification for basing rights or claims to justice on species membership alone.

To sum up, Richardson extends Rawls's Original Position by including representatives for individuals with severe cognitive disabilities, where the inclusion is sustained through a principle of humanity that requires all human beings to be protected by justice. While this approach is appealing, such inclusion should not be assumed *a priori*. The scope of justice is a contested issue, and presupposing an answer undermines respect for pluralism. If species membership is not a valid basis for justice[20], this assumption becomes problematic. Instead of automatically including certain groups in a way that is contested, such inclusion should emerge from deliberation among representatives of diverse reasonable perspectives. Otherwise, Richardson's approach risks imposing a specific moral viewpoint without engaging with alternative conceptions of justice. A deliberative process, considering what could be justified to an idealised constituency with different reasonable views, would ensure that inclusivity results from thoughtful discussion rather than being assumed from the outset when questions are contested. This kind of approach would better respect pluralism by allowing different perspectives on justice to shape decisions about inclusion.

C. Samuel Freeman: Contractualist Approach to Individuals with Severe Cognitive Disabilities

Samuel Freeman (2018) defends the inclusion of individuals with severe cognitive disabilities in the scope of justice by appealing to contractualist principles. His approach builds upon Rawls' political liberalism while addressing the common critique that contractualism, which focuses on rational agreement among free and equal persons, struggles to recognize the rights of those unable to participate in social cooperation. Freeman (2018) argues that justice must extend to individuals with severe cognitive disabilities by recognising their fundamental interests and moral worth through representation by trustees or guardians. He also challenges alternative approaches, such as the capabilities framework, and proposes a contractualist justification that respects the dignity of all persons, regardless of cognitive capacity.

In this section I will analyse Freeman's main argument, examining how he modifies Rawlsian contractualism to include individuals with severe cognitive disabilities in the scope of justice. I will also consider the implications of his argument for the design of just institutions and policies that address the exceptional needs of these individuals.

[20] As I have shown in the section in which I criticise Nussbaum's approach.

Freeman (2018) recognizes a key challenge for contractualist justice: If justice is based on agreement among free and equal people, how can it apply to those who cannot reason or recognize rights and duties? This excludes individuals with severe cognitive disabilities and weakens their moral status. However, Freeman (2018) argues that this criticism misinterprets the nature of contractualist justification. He emphasises that the test of legitimacy and justice does not depend on the actual ability of individuals to participate in rational agreement but rather on whether the principles governing society can be justified to them or their representatives in a way that respects their fundamental interests and moral worth. He argues that a political conception of justice, such as Rawls' liberalism, is not based on contingent individual capacities but on the idea that all persons deserve to be treated with fairness and equal concern.

In this context, Freeman (2018) proposes that individuals with severe cognitive disabilities can be represented in contractualist reasoning by trustees or guardians who act on their behalf. This approach ensures that their fundamental interests are taken into account in the justification of moral and political principles. He argues that just as representatives in Rawls' original position act on behalf of free and equal persons by considering their rational interests, trustees can play a similar role in ensuring that the rights and needs of individuals with cognitive disabilities are addressed.

This argument is based on two key claims. First, he asserts that individuals with severe cognitive disabilities have moral worth and share basic needs with other citizens, such as protection, care, and dignity. Justice requires that these needs be considered when shaping fair institutions and policies. Second, he argues that even though these individuals may not understand or endorse public reasons themselves, their interests can still be represented through trustees. Just as children or elderly individuals with dementia remain part of a just society despite cognitive limitations, so do those with severe disabilities. By relying on trustees, Freeman (2018) ensures that justice includes everyone, recognizing different needs and capacities rather than excluding certain groups.

Freeman (2018) also critiques the capabilities approach, such as that of Nussbaum, which focuses on ensuring people have real freedoms to achieve a decent life by considering what individuals can actually do and be. While this approach has been influential in discussions on disability, Freeman argues that it may not fully apply to those with severe cognitive impairments. He points out that the capabilities framework assumes a level of agency and autonomy that may not be relevant for individuals with profound disabilities. Since some people require lifelong care rather than expanded choices, focusing on capabilities might overlook their basic needs. Instead, Freeman suggests that justice should prioritise a theory of basic needs. He maintains that contractualist justice, when extended through trusteeship, effectively secures the rights of individuals with severe cognitive disabilities without requiring an entirely different framework.

Freeman's argument has important implications for how just societies should be structured. If justice requires addressing the fundamental interests of individuals with severe cognitive disabilities, then society must ensure their basic needs—such as healthcare, education, and social support—are met while respecting their dignity and moral worth. First, just institutions must provide comprehensive care and assistance, recognising the exceptional needs of individuals with severe disabilities and ensuring they have the necessary resources for a decent life. Second, these individuals must have legal protections that secure their rights, including oversight to ensure their trustees act in their best interests. Finally, justice requires not only material support but also social and cultural inclusion, ensuring that individuals with disabilities are treated with respect and given opportunities to engage in their communities. Freeman's argument reinforces the idea that justice is not only for those who can actively participate in social cooperation but must extend to all individuals, regardless of their cognitive abilities.

To sum up, Freeman's (2018) contractualist approach offers a compelling defence of the inclusion of individuals with severe cognitive disabilities within the scope of justice. First, Freeman remarks that lacking capacities for being reasonable and rational does not exclude such individuals from justice, neither as those that are protected through the principles of justice, nor as those who are included in the construction (justification) of principles of justice. The important idea is that there are no real life individuals included in the process of justification of principles of justice. This process of justification is based on a thought experiment where we imagine representatives of real-life individuals instantiating reasonableness and rationality. It is not needed that those individuals that they represent are reasonable and rational. Thus, there is no obstacle to including individuals with severe cognitive disabilities in justice. They are included through their representatives. But this is true for all others, as well. Freeman's contribution highlights the flexibility and inclusivity of contractualist justice. Rather than being inherently exclusionary, contractualism—when properly developed—can provide a strong foundation for protecting the rights and dignity of individuals with severe cognitive disabilities, ensuring that justice truly applies to all members of society. The central idea is that all human beings must be included in justice. This is referred to as the "principle of humanity".

Although Freeman's contractualist approach offers a strong argument for including individuals with severe cognitive disabilities in justice, it faces significant objections. Indeed, it faces similar challenges to Richardson's proposal. I argue that Freeman's (2018) view of the role of humanity in reasoning about justice does not respect pluralism. The concern is that some reasonable doctrines reject the idea that membership in a species is a normatively relevant basis for justice. For example, critics of speciesism such as Singer (2009) argue that moral considerations should not be constrained by species boundaries. The argument is also supported by evolutionary theory, as I mentioned in section four when I criticised Nussbaum's theory - which

shows that species are not fixed entities, but categories of organisms with shared characteristics (Rachels, 1987). This undermines the idea of basing rights on membership of a species.

Another significant critique of Freeman's approach concerns the indeterminacy of his justification for inclusion. While he argues that justice should account for the fundamental interests of individuals with severe cognitive disabilities, he does not provide a clear standard for determining what these interests entail or how they should be balanced against the claims of other citizens. This lack of clarity risks making the scope of justice overly vague or contingent on external moral considerations rather than being firmly grounded within contractualism itself.

To conclude this part, a more defensible approach would involve addressing the inclusion of individuals with severe cognitive disabilities through a deliberative process that respects reasonable pluralism, rather than assuming their status within justice from the outset. This would allow for a more transparent and inclusive justification that is responsive to the diversity of perspectives within liberal political thought. In the next section, I will examine the final alternative to Nussbaum's approach, one by Cynthia Stark, and then offer my own solution for including individuals with severe cognitive disabilities within the scope of justice.

D. Cynthia Stark's Approach: Beyond Productivity in Social Contracts

In her work, Cynthia Stark (2007) explores how justice operates within a political liberal social contract, centred on the concept of society as a fair system of cooperation among free and equal individuals. She (2007: 131) distinguishes between *fully cooperating individuals* and *non-cooperators*, which include human nonpersons (lacking moral powers) and individuals with severe impairments (who have moral powers but are "unable to cooperate given a society's particular level of technological advancement, even if that society is committed to making accommodations"). Stark argues that a just society should ideally focus on those capable of contributing productively to its social and economic systems. Accordingly, the principles of justice should be justifiable to representatives who act on behalf of these contributors. She emphasizes that justice must acknowledge and fairly reward the efforts of those productive members who help generate the goods and resources on which society depends.

However, Stark also acknowledges that justice should not neglect those individuals who are unable to contribute productively to society. Their interests are only taken into account at the second stage of reasoning about justice, i.e. at the constitutional stage. At this stage, justice should fulfil the needs of all members of society, i.e. not only those who are economically or socially productive. In doing so, Stark argues that the social contract must balance the needs of both productive and non-productive

members (*fully cooperating individuals* and *non-cooperators)*, ensuring fairness for everyone within the society:

> *"My proposal is to retain the fully cooperating assumption in the original position but to drop it at the constitutional stage of the theory. Ideal constitutional conventioneers should imagine that they might be disabled in a way that prevents them from participating in a scheme of cooperation and should fashion the constitutional provision for the social minimum with this possibility in mind. The minimum would, in this case, presumably be as high and as comprehensive as possible, within the constraints imposed by the difference principle, and would cover all the claims of need had by the non-cooperating in the areas of shelter, food, clothing, transportation, utilities and the like* (Stark 2007: 138*).* (…) *In short, my proposal meets the liberal principle of legitimacy by allowing that the terms of cooperation for participating citizens are determined by representatives of those citizens (in the original position) and that the policies for meeting the needs of the non-cooperating, who are dependent upon the goods produced by participants, are determined by representatives of those citizens. Some representatives in the constitutional stage will represent (though they will not know they represent) individuals who do not qualify as persons in Rawls's sense. They will represent individuals who lack the two moral powers. In this case, the ideal deliberators serve as trustees for such individuals. So long as the policies for addressing the needs of the non-cooperating are adopted by representatives of the non-cooperating, whether those representatives turn out to have represented persons or non-persons, those policies are justifiable to those non-cooperating individuals who are owed a justification. So, those policies are legitimate by the lights of hypothetical consent theory* (Stark 2007: 140*)."*

Stark's proposal bears similarities to the one I will endorse, particularly in recognising the importance of principles that address the needs of individuals independently of their contribution to society. However, there are key differences between **the** approach I will propose and Stark's. First, I argue that the justification of principles addressing needs should be placed at the fundamental level of justice, as needs are a significant concern for productive members of society as well, contrary to the idealisation in Rawls's framework. This idealisation, I believe, distances justice from real-life issues, even for productive members.

Second, I contend that if needs are not addressed at the fundamental level, their inclusion at a lower level lacks a clear foundation (Hartley 2009). In my view, an appropriate sequence of justification of principles and their application must begin

with the more abstract levels and gradually move to the less abstract levels, culminating in practical application. For this reason, needs should be included in the most fundamental level of justice. Although the initial focus may be on the needs of productive members (because their representatives legislate), this does not exclude the needs of others. I argue that defining public justification as the justifiability of principles to rational and reasonable members of society does not entail the permanent exclusion of individuals with severe cognitive disabilities from justice.

2. CHAPTER TWO: SECTION ONE: NEW SOLUTION: IDEAL REASONABLE AGENTS (IRAS) MODEL FOR JUSTICE[21]

The main aim of the first chapter was to address the shortcomings of Rawls's theory, which emphasises reasonableness and rationality as central attributes of justice and thereby fails to include individuals — such as people with severe cognitive disabilities—within its scope. To fill this gap, I propose a revised model of public justification centred on the concept of ideal reasonable agents (IRAs). According to this model, IRAs — individuals capable of reasonableness and rationality — serve as impartial legislators of principles of justice. Their role is to evaluate and justify principles of justice not only for themselves but also on behalf of individuals unable to participate directly in the justification process. This ensures that the rights and interests of all individuals, including those who lack reasonableness and rationality, are recognized and protected through a process of universalization. By universalising principles of justice, IRAs extend the rights they establish for themselves to those they represent (Martinić and Baccarini 2023).

Importantly, this model maintains the inclusion of individuals incapable of reasonableness and rationality as "beneficiaries of justice." While these individuals do not partake in the legislative process, their needs and interests are safeguarded within the framework of justice. Drawing on Thomas Schramme (2021), the model emphasizes the necessity of interpreting non-verbal signals as meaningful communication, ensuring that these individuals are genuinely included in the justice system. The findings of Stacy Clifford (2012) emphasise the limits of the assumption that language is a universally clear medium. It is thus a matter of choosing to recognise non-verbal speech acts in the interactions between IRAs and beneficiaries of justice.

This nuanced approach bridges the gap between the rights of reasonable and rational individuals and those who cannot fulfil these criteria, promoting a more inclusive and just justice system. It reflects the ideal of society as a fair system of cooperation in which the principles of justice emerge through a reflective equilibrium — a method of reconciling principles and judgements by taking into account the different interests and needs of all individuals. It is considered an equilibrium because our principles and judgements end up aligned. It is also a balance because we are aware of the principles that guide our judgements and the foundations from which they are derived. In an ideal scenario, everything is in harmony, but this equilibrium may not necessarily remain stable (Rawls 1999: 18). I will now elaborate on how this model operates, illustrating its capacity to harmonise substantive judgments about fairness with the principles of justice.

[21] The analyses and theses for this approach were developed in collaboration with my supervisor Elvio Baccarini and the JOPS research project. Specifically, the idea was formed in the joint article, *"Capabilities and Justice for People Who Lack the Capacity for Reason and Rationality,"* published in *Filozofska istraživanja* 43.3 (2023): 495

First it is crucial to elaborate more on the role and structure of *Ideal Reasonable Agents* (IRAs) and their importance for justice as a framework that goes beyond the traditionally rational and reasonable. Ideal Reasonable Agents (IRAs) represent a theoretical construct used to imagine a version of people who make the best possible use of their capacities for reasonableness and rationality. In the real world, human beings often encounter limitations such as personal biases, emotional influences, cognitive errors, and social pressures that can prevent them from consistently acting or thinking according to principles of fairness and logic. In contrast, IRAs are an idealisation designed to remove these limitations and enable them to act as perfect agents of impartiality and rational decision-making. This idealisation plays an important role in philosophical and ethical concepts, especially those related to justice. By imagining how such agents would reason and decide, the concept provides a basis for determining the principles of justice that could be universally justified. The consistency of IRAs ensures that their reasoning is not influenced by self-interest or situational factors. This allows them to objectively judge what is fair and just for all, including those who cannot participate in the process themselves. IRAs are not intended to describe actual individuals but provide a model for ideal reasoning and decision-making that can help with real-world ethical and political considerations. They set a standard for the rational and impartial application of justice, free from the inconsistencies and errors that typically influence human behaviour (Martinić and Baccarini 2023). In this theoretical model for the construction of concepts of justice, the IRAs thus serve as an ideal basis for establishing principles of justice and the specification of their application. Principles and specifications are considered justifiable if they are acceptable to all idealised persons. Since IRAs, as reasonable and rational actors, must specify the principles of justice, they do so impartially and avoid a selfish perspective or a bias towards their own interests. Consequently, IRAs extend the principles of justice to those who lack the capacities for reasonableness and rationality (i.e., persons who cannot be legislators of principles of justice themselves), who then become *the beneficiaries of justice* (Martinić and Baccarini 2023).

An objection can be raised against this kind of extension of Rawls's scope of justice. Specifically, one might ask: Why should reasonable people act fairly and impartially towards those who are unable to participate equally in the system of social co-operation or in the process of justifying principles and applications of justice? I offer two answers to this question (Martinić and Baccarini 2023). First, I argue that such an extension is necessary to maintain the consistency of the concept of reasonableness. The distinct feature of reasonableness is the rejection of public decisions based on personal interest, instead favouring impartial considerations of fairness. This distinguishes reasonable agents from those who are merely rational—focused solely on their conception of the good. If reasonable agents limited their impartiality to those who can cooperate on equal terms in society or the justification process, they would reduce fairness to a consideration of their own interests in

cooperating with equals. This would be a more sophisticated form of self-interest than that demonstrated by those incapable of impartiality, even toward their equals, but it would still reflect rationality (as the capacity to pursue one's conception of the good) rather than reasonableness (as the capacity for fairness). Thus, extending the scope of justice to include those who lack reasonableness and rationality aligns with Rawls's definition while expanding upon it. This extension clarifies rather than revises Rawls's framework (Martinić and Baccarini 2023). In other words: At the heart of the argument is the idea that ideal reasonable agents, characterised by their commitment to fairness and impartiality, should extend justice to those who cannot participate equally in social cooperation or in the justification of principles of justice. This extension is considered necessary to maintain the integrity of the concept of reasonableness. It must be clear that there is a difference between reasonableness and rationality. Rationality refers to the individual pursuing their personal idea of what is good or beneficial for them. It is goal-orientated and often self-serving. Reasonableness, on the other hand, means prioritising fairness and making decisions that are impartial, even if these decisions are not in line with one's own interests. It requires an ethical commitment to prioritise fairness over personal gain. The IRA upholds the consistency of reasonableness. If reasonable agents were only fair and impartial to those who can reciprocate (i.e., equal participants in society or in the justification process), they would be acting in subtly self-interested ways. For example, they might choose fairness not for its own sake, but because it benefits them in a system of mutual co-operation. This would reduce reasonableness to a sophisticated form of rationality in which fairness is merely a strategy to serve one's own interests among equals. Such behaviour would contradict the nature of reasonableness, which presupposes impartiality and fairness as universal principles that do not depend on reciprocity or self-interest. If reasonable people only acted fairly towards their equals, they would contradict the very principles that define their reasonableness. By including those who cannot reciprocate, they maintain the integrity and consistency of their commitment to impartial fairness and ensure that justice is applied universally rather than being a tool for strategic self-interest (Martinić and Baccarini 2023).

The second answer to the question of why reasonable people should act fairly and impartially towards those unable to participate equally in the system of social co-operation or in the process of justifying principles and applications of justice lies in Rawls' own definition of reasonableness. Rawls defines reasonableness as a characteristic of individuals who are not biassed towards their own needs. He also describes reasonable individuals as those who do not seek to dominate others in a disadvantaged position. This includes individuals who cannot reason, rationalise or have a say in the content of justice (Martinić and Baccarini 2023). Rawls emphasizes that reasonable individuals "are willing to propose certain principles [...] and abide by them even at the expense of their own interests if circumstances require it" (Rawls 2001: 191). Those who impose principles of justice "motivated, for example, by their

greater power or superior bargaining position" (Rawls 2001: 191) cannot be considered reasonable. However, when Rawls describes reasonable individuals as ideal legislators who establish principles of justice not from their own interests but impartially for all, he limits this description to relationships among reasonable individuals. A key condition for such impartial and fair behavior, he notes, is that "others are encouraged to act appropriately" (Rawls 2001: 191). Nonetheless, consistent with Samuel Freeman's (2018) interpretation of Rawls's work, I argue that this condition reflects the boundaries of Rawls's specific aims in the context of particular discussions, rather than a definitive limitation on who can be included within the scope of fairness as recognized by reasonable individuals. Freeman (2018) contends that in differentiating his position from utilitarianism, Rawls explicitly emphasizes that moral contractarian conceptions presuppose that social cooperation should work for the benefit of every individual in society; otherwise, it constitutes unfair exploitation. This does not imply, however, that every social relationship must involve mutually advantageous reciprocity (Freeman 2018: 182–183). More precisely, Freeman's argument demonstrates that including individuals who lack reasonableness and rationality as beneficiaries of justice is not excluded by Rawls's conception (Martinić and Baccarini 2023).

The intent of my argument is to show why such inclusion is necessary. This is why I stress that the motivation of reasonable individuals is fairness, and correspondingly, impartiality—an approach centered on fairness rather than the pursuit of particular self-interests. Consequently, it is consistent to define IRAs as ideal reasonable agents who also take into account the interests of the most vulnerable, those who lack the capacities required to serve as ideal reasonable agents themselves. It is therefore necessary to define principles of justice that, as justifications of universal principles of justice, also include principles of application covering those who do not participate in this process. When reasonable individuals establish principles of justice and their applications, they determine rights and protections that extend to those lacking the capacities required for reasonableness (Martinić and Baccarini 2023).

Further, the extension of the scope of justice to include as beneficiaries those who lack the capacities for reasonableness and rationality can also be supported through the method of "reflective equilibrium." As mentioned above, reflective equilibrium is a method aimed at achieving "coherence between all beliefs relevant to moral reasoning," beginning with "moral judgments of varying levels that the reasoner holds" (Baccarini 2007: 9). This definition is provided for further clarity. The argument being made is that, using reflective equilibrium, it becomes difficult to claim that a person possesses the capacity for fairness or reasonableness (i.e., the ability to make just decisions) while being indifferent to the suffering of others. This holds true regardless of whether the individuals experiencing the suffering are capable of reason or rationality, or whether they can engage in reciprocal social cooperation. The method of reflective equilibrium thus requires us to adjust and revise our beliefs

until they harmonise in a well-rounded and consistent way. In this case, it encourages us to recognise that fairness is not just about reciprocity or cooperation, but also about minimising suffering and providing basic protection to all individuals regardless of their capacity for reason or rationality. For example, people with severe cognitive impairments may not be able to participate in social co-operation in the traditional sense, but their suffering and well-being are still morally relevant. From the perspective of reflective equilibrium, we adapt our moral principles to include them as beneficiaries of justice. Thus, those who are unable to co—operate reciprocally - such as people with severe cognitive disabilities who cannot reason or participate in discussions about justice — should also be considered under the umbrella of justice. Society in this case would extend justice to ensure that they receive the care and protection they need, such as access to medical care, social services or physical assistance, even if they are unable to contribute equally to the social contract.

What rights and corresponding capabilities should be protected through the consistent application of the concept of reasonableness and the universalization of protection for capacities and rights? At this point, as in my earlier argument, I extend Rawls's considerations. Rawls himself focused on identifying fundamental rights and freedoms that reasonable and rational individuals would establish, concentrating exclusively on fundamental interests strictly tied to their capacities for reasonableness and rationality (such as the protection of political liberties and the fair value of these liberties). However, this does not exhaust all their fundamental interests. It is also necessary to protect other essential interests and needs. Following Kimberley Brownlee, I refer to "non-contingent needs that are necessary conditions for the non-contingent goals that individuals necessarily have" (Brownlee 2012: 188). These needs go beyond mere survival. Survival alone does not capture the richness of human life. Instead, survival must be understood as the preservation of the individual as a being with certain essential characteristics. For example, non-conditional needs include protection from suffering, which severely limits a person's ability to engage in meaningful activities or pursue goals that constitute a fulfilling life. A person suffering from chronic pain or mental health problems is unable to lead a fulfilling life, and justice would require that this need for protection from unnecessary suffering be met. Another example of non-contingent needs is the need for freedom of movement in an environment where that movement is meaningful and fulfilling. This refers to the individual's ability to move freely in their social or physical environment in a way that contributes to their well-being. For example, a person with mobility problems should have the right to enter public spaces and participate in society without unnecessary barriers. These needs contribute to a person's overall fulfilment. Access to a meaningful environment can be crucial to a person's wellbeing, especially for those who are otherwise unable to contribute in typical social or economic ways. Thus, in the process of justifying the content of justice, reasonable and rational individuals will include the protection from suffering and the protection of freedom of movement as rights. Such capacities and rights should also be protected for all

individuals, based on the consistency of the concept of reasonableness and its universalization (Martinić and Baccarini 2023).

The remaining challenge is extending the scope of justice to the specific capacities and corresponding rights that pertain to individuals with severe cognitive disabilities. So far, my argument has emphasized the need to universalize rights to include such individuals because reasonable individuals, by virtue of their reasonableness, do not limit the reach of established rights to only certain persons. However, as the earlier argument demonstrates, reasonable individuals universalize the protection of capacities and corresponding rights that are relevant to themselves. The critique from Nussbaum theory, however, points out that for individuals with severe cognitive disabilities, it is necessary to emphasize their unique capacities and corresponding rights. This is why I supplement my first response with a second. I achieve this through an argument related to Rawls's well-known model of the "original position." This model envisions Ideal Reasonable Agents (IRAs) as hypothetical architects of fundamental principles of justice. To implement ideas of reasonableness and ensure the freedom and equality of all, these IRAs must not favour their own interests. This is achieved by situating them behind a "veil of ignorance," where they know nothing about their particular status or beliefs. Accordingly, they will design principles of justice that, among other things, protect them from possible adverse outcomes. Such outcomes include conditions of severe cognitive disabilities due to accidents or aging processes. Therefore, individuals behind the veil of ignorance, in constructing a comprehensive conception of justice, must design principles that guarantee rights even in such circumstances. For example, these rights might include the right to care for those unable to care for themselves, the right to environmental adaptations (including socialization conditions), and more. This alone, however, is still insufficient to guarantee such rights for individuals who have always been in states of severe cognitive disability. Behind the veil of ignorance, individuals still possess general information about themselves—namely, that they have the capacities for rationality and reasonableness, as it is these capacities that grant them a place in determining principles of justice from behind the veil of ignorance. The question arises again: why would they extend rights to those who lack the capacities for rationality and reasonableness? The answer is similar to the one I have already offered. They universalize rights precisely because of their reasonableness. As noted, because reasonable individuals establish principles of justice impartially and not solely for themselves, they establish rights universally. If rights are universal, then rights pertaining to individuals who lack the capacities for rationality and reasonableness, such as those with severe cognitive disabilities, apply to all such individuals (Martinić and Baccarini 2023).

How, then, can reasonable individuals determine the relevant capacities to support for individuals with severe cognitive disabilities—capacities they themselves might need if they were to find themselves in such states—and the corresponding rights? This is

not a straightforward task and requires attentiveness, the ability to interpret the conditions of others, and openness to evolving understanding. It is important to listen to the voices of individuals who lack the capacities for rationality and reasonableness through the various modes of expression they possess. For example, Thomas Schramme highlights the significance of recognizing non-intellectual forms of expression, which allow for broader communication with individuals lacking rational and reasonable capacities. One of Schramme's examples includes individuals with dementia, who express discomfort with certain practices in care homes non-verbally, enabling caregivers to provide better care (Schramme 2021). This can be seen as a form of engagement in striving for thoughtful and effective care practices (Martinić and Baccarini 2023).

In conclusion, the extension of justice to include individuals who lack the capacities for reasonableness and rationality represents a crucial development in our understanding of fairness and moral responsibility. By utilizing the concept of Ideal Reasonable Agents (IRAs), we can establish a framework of justice that transcends the limitations of human biases, personal interests, and situational constraints. This idealized model encourages us to recognize that justice should not be restricted to those who can actively participate in social cooperation or the process of justifying principles of justice. Instead, it requires the inclusion of all individuals, regardless of their ability to reason or engage in reciprocal cooperation. The consistency of the concept of reasonableness demands that fairness be universally applied, even to those who cannot reciprocate. This inclusion is necessary to maintain the integrity of reasonableness itself, as limiting justice to those capable of cooperation would reduce fairness to a self-interested form of rationality. Furthermore, Rawls's own definition of reasonableness supports this universal application, emphasizing impartiality and the rejection of bias in favor of one's own interests. By including those who lack the capacity for reason or rationality, we ensure that justice is applied universally, without discrimination. Moreover, the method of reflective equilibrium strengthens the argument by demonstrating that true fairness must consider the well-being and protection of all individuals, even those who are unable to participate in the social contract. Through this method, we harmonize our moral principles to ensure that they are inclusive and just for everyone, particularly those who are most vulnerable (Martinić and Baccarini 2023).

The extension of justice to those with severe cognitive disabilities and other vulnerable groups requires a thoughtful and evolving approach. It involves recognizing the unique capacities and needs of these individuals, while ensuring that the rights they are entitled to are universal and non-contingent. Through the lens of the original position and the veil of ignorance, reasonable individuals would design principles of justice that protect even those who are not capable of contributing to the formation of those principles, including those with severe cognitive impairments. Ultimately, the commitment to justice, reasonableness, and fairness demands that we

extend protections to all individuals, not just those capable of rational or reasonable thought. By universalizing rights and considering the needs of the most vulnerable, we create a more inclusive and just society, one that recognizes the inherent dignity of all people, regardless of their cognitive capacities. This vision of justice requires us to look beyond traditional boundaries and redefine fairness in a way that benefits everyone, ensuring that no one is left behind.

2.1. Section Two: Extending the principle of justice: the case of non-human animals

This section explores the extension of justice principles to non-human animals, addressing both theoretical ideals and practical applications. In previous discussions, I have addressed the challenge of recognising and protecting the rights of individuals who do not conform to traditional notions of rationality and reasonableness, particularly those with cognitive disabilities. My argument centred on the pursuit of universal principles of justice, applicable to all individuals regardless of their cognitive capacities. To achieve this, I proposed a model of public justification based on Ideal Reasonable Agents (IRAs). By serving as impartial legislators, IRAs establish principles of justice that respect the needs and rights of all, including those who cannot directly participate in the reasoning process.

On this basis, questions of justice arise in relation to beings that share with humans the characteristics relevant to justice. This leads me to the issue of justice for non-human animals. Some scholars, such as Eva Feder Kittay (1999, 2001, 2005a, 2005b), have outright rejected, on a principled level, the idea that non-human animals possess moral status and, consequently, that they should be included within the scope of justice. I will argue that Kittay's argument is unsatisfactory.

Some Rawlsian theorising, developed by scholars such as Abbey Ruth (2007), Mark Rowlands (1997), Alasdair Cochrane, Robert Garner, Siobhan O'Sullivan (2018) and Brian Berkey (2017), has attempted to provide a framework for the inclusion of non-human animals in justice. However, I will also argue that their theories do not provide a fully adequate justification for the inclusion of non-human animals at a principled level. Instead, I will defend the view that the IRA model offers a superior approach.

Nevertheless, challenges arise in real-world applications. The broader the scope of those entitled to rights, the greater the difficulty in ensuring that all these rights are effectively protected in practice. This raises the issue of moral conflict. Bernard Williams (1981), and following him, Baccarini (1994) [22], have explored moral conflict on a principled level. Within this framework, ideal principles are established, and in a world without contingent, real-life conflicts between them, it would be optimal to uphold them all. However, when such conflicts do occur, we are faced with moral

[22] For more see: Williams, Bernard. *Moral luck: philosophical papers 1973-1980.* Cambridge University Press, 1981. and Baccarini, Elvio. *Moralni sudovi.* 1994.

dilemmas that require us to weigh competing principles and determine which holds greater normative weight—a discussion that remains at the level of abstract principles.

That, however, is not my focus. Instead, I address what can realistically be achieved in the real world, taking into account not only the scarcity of resources but also existing social and institutional relationships. In other words, I examine what aspects of justice for non-human animals can feasibly be translated into real-world policies, considering both the practical constraints and the perspectives of those who currently hold decision-making power. In constitutional democracies, this means engaging with the views and priorities of participants in democratic processes.

I proceed as follows: I present Kittay's approach, then Rawlsian theories, and at the end of the chapter my own solution for the inclusion of non-human animals in the scope of justice.

A. Eva Feder Kittay Approach[23]

In this section, I focus on Eva Feder Kittay's response to the exclusion of individuals with cognitive disabilities from the realm of moral personhood. Kittay argues that species membership alone determines moral status, setting her apart from scholars such as McMahan (1996, 2002), Donaldson and Kymlicka (2011, 2014), Oliver (2020), and Taylor (2017). She grounds this determination in "social relations," using the family analogy to highlight our shared humanity as the basis for moral inclusion. Although I am very much in favour of including individuals with cognitive disabilities in the framework of moral personhood, I have objections to Kittay's methodology. I contend that her approach unjustly excludes non-human animals, revealing a problematic form of speciesism. By commenting on her reliance on species membership and social relations as determinants of moral status, I aim to highlight the limitations of her framework and advocate for a more inclusive understanding of moral community.

I will begin with the definition of moral personhood that Kittay (2005) herself provides and criticises. Namely, the definition is that moral personhood is the designation of beings who are entitled to moral protection and consideration in the realm of what can be called "moral."[24] This categorisation usually depends on the presence of certain characteristics, such as the capacity to recognise morally right or wrong actions and the possession of psychological traits such as rationality and autonomy (McMahan 1996; 2002).

These criteria often lead to the exclusion of many beings from the moral sphere, especially those who lack these capacities, such as people with severe cognitive

[23] The ideas explored and analyzed in this paragraph have already been proposed in my article "Evaluation of Eva Feder Kittay's Framework on Cognitive Disabilities and Moral Status of Non-Human Animals," Etica & Politica (2024).

[24] https://medicine.missouri.edu/centers-institutes-labs/health-ethics/faq/personhood 03.08.2023.

disabilities. For this reason, Kittay's theory, as outlined in her 2005 paper, is based on the claim that moral considerations should be based on an individual's membership in a species. In particular, she argues that the function of "social relations" should be to shape an individual's moral identity (Kittay 2005a). In other words, the fact of being human should be enough to justify moral considerations. This view shifts the focus from cognitive capacities to the fact that someone is part of the human species.

By 'social relations' she means:

> *"... a place in matrix of relationships embedded in social practices through which the relations acquire meanings. It is by virtue of meanings that the relationships acquire in social practices that duties are delineated, ways we enter and exit relationships are determined, emotional responses are deemed appropriate, and so forth. A social relation in this sense need not to be dependent on ongoing interpersonal relationship between conscious individuals. (...) Identities that we acquire are ones in which social relations play a constitutive role, conferring moral status and moral duties. These identities are part and parcel of social matrix of practices, roles, and understandings, which are themselves enmeshed in a moral world"* (Kittay 2005a: 111).

As evident from the quote, social relations hold a distinct significance for individuals, primarily due to the deep emotional bonds they entail. In other words, the relationships we have with others and the way we are embedded in a social context are fundamental to who we are as moral beings. With this, Kittay wants to emphasise the importance of care, relationships and interdependence in understanding moral personhood, rather than focusing only on individual cognitive traits.

In this context, Kittay emphasises the uniqueness of human social relations. She argues that the uniqueness of these social relations between people is accompanied by distinct and exceptionally strong moral connections, obligations and claims between them. In other words, the obligations and entitlements refer to the duties and rights that arise from these human relationships. The relationship between a parent and a child, for example, involves significant moral obligations (such as care, protection and nurture) and rights (such as the right to be cared for and loved).

Kittay compares human relationships with those between humans and non-human animals. While the relationship between a pet owner and their pet is significant and often emotionally meaningful, it does not carry the same moral weight as human relationships such as parenthood. This is because, according to Kittay, the moral obligations and demands in a parent-child relationship are much stronger and more profound than in a human-animal relationship. Kittay's view reinforces a human-centred moral framework in which human relationships are prioritised and the moral significance of relationships with non-human animals is seen as secondary.

Parenthood has a special significance for Kittay, as she argues in her earlier work "Love's Labour" (1999), where she explains this significance in terms of the inherent dignity of being "a mother's child" By saying "We are all a mother's child'," Kittay (1999; 2005b) emphasises that all human beings, regardless of their circumstances, share the common experience of being cared for by a maternal figure. This experience is universal and fundamental to human life. Moreover, it indicates that the value of this care is profound and forms the basis for many of our moral and ethical claims. The care of a mother (or maternal figure) is fundamental to our development and well-being. Kittay's claim implies that everyone is equally entitled to the benefits and rights that come from being "a mother's child" This means that the care, love and protection typically associated with motherhood are things to which all humans have a moral claim. This forms the basis of her broader ethical perspective, which emphasises the importance of care and relational duties in defining a just and moral society.

Kittay (2005b) explains this inherent dignity in the following way:

> *We utter these locutions when we want to remind our interlocutor (or ourselves) of the humanity of someone who seems to have been vanquished from our moral domain — the enemy we fight, the evildoer we want to punish, the homeless person living a life that is hardly recognisable as human, the inhabitant of a body noticeably twisted and a brain that only slowly takes in its world. We may say it even of ourselves when we have exerted ourselves on another's behalf and need to remind someone (perhaps ourselves) of our own need for care. It is herein that I hear a claim to equal dignity, one that is an alternative to conceptions dominating philosophical discourse. It is a claim with both moral and political consequences. Unlike most claims to equality where we invoke some common property, we each possess as individuals and from which we make claims to equal treatment, welfare, opportunity, resources, social goods, capabilities, rights, or dignity, when I assert that 'I too am some mother's child' I invoke a property that I have only in virtue of a property another person has. One is the child of a mother only because another person is someone who mothered one* (Kittay 2005b; 113).

It is important to note that Kittay's work emphasises that the uniqueness of parenthood is not tied to gender or biology, but that it is defined by caring for a dependent and vulnerable other. In other words, what makes parenting unique, according to Kittay, is the commitment to someone who is dependent and vulnerable, such as an infant or a person with disabilities. These activities are central to the parent-child relationship and give parenting its moral and ethical significance. Kittay (1999; 2005b) argues that the value of caregiving is significant because it is not just a practical necessity, but a moral activity that emphasises the intrinsic worth of the individuals involved. The act

of caring demonstrates that the person being cared for has an intrinsic value— a value that does not depend on their capacities or status, but simply on their existence as a vulnerable human being. At the same time, the role of the carer is also recognised as valuable, reflecting the importance of the relationship itself. Kittay points out that infants' survival and well-being depend on the care they receive. The fact that infants survive and thrive because of the care they receive emphasises the crucial role that care plays in sustaining life. Without care, the most vulnerable individuals would not be able to survive, which emphasises the essential role of care in human life. Thus, the survival of people who are dependent— on care, such as infants, is proof of the importance of care. This survival confirms that care is a fundamental aspect of human life, necessary for both physical survival and the development of moral and social beings. In highlighting the value of caregiving, Kittay also argues for the recognition of the value of the carer. Caring for others is not only necessary for the survival of the person in need of care, but also enriches the carer and gives moral significance to their role. In light of this, Kittay emphasises the unique relationship between a mother or carer and the child, in which the carer prioritises the child's needs over their own interests. This caring relationship is based on a specific love from which arises the duty to care for the child when necessary (Kittay 2005b; 116-118).

Kittay argues that this dignity, which arises from the caring relationship in which a person cares for a dependent person, has a moral value of its own. She believes that this kind of dignity is just as important as any other and should not be considered less valuable simply because it is associated with dependency. In doing so, she challenges the common notion that dependence on others somehow diminishes a person's dignity. Instead, she argues that these caring relationships— in which someone provides care, and another receives it — are essential to what it means to be human and to our understanding of dignity. She emphasises that caring for others and caring about others are fundamental aspects of our humanity and should be respected as such.

To support the above assertions, Kittay points to the widespread use of the term "child of a mother" in various cultures, which once again emphasises the deep-rooted nature of caregiving as a fundamental aspect of human experience (Kittay 2005b; 116–117). It shows that the bond between carer and cared-for is not just a specific or isolated idea, but a core aspect of humanity that transcends cultural and societal boundaries.

To further illustrate the importance of dignity in care, Kittay refers to her daughter Sesha, who has severe cognitive disabilities.

Kittay argues that Sesha's worth and dignity do not stem from her capacity to think or reason logically, but from the love and care she receives from those around her. Recognising Sesha's value through her relationships prevents dehumanisation and affirms the value of the carers. In other words, when we value Sesha because of her relationships and the care she receives, we avoid treating her as less than human and recognise the importance of those who care for her. Kittay argues that caring

relationships are essential for individuals to have intrinsic worth, regardless of their physical and mental fluctuations. The dignity that comes from caregiving is rooted in our common humanity — our need for care and our vulnerability — and it is evident through the commitment of caregivers. Kittay emphasises that, especially in difficult times, we should remember that everyone deserves dignity because we are all vulnerable in some way. She argues that we can build a more empathetic and compassionate community (Kittay 2005b: 118). In this sense, in her earlier work, Kittay (2001) contrasts the joyful moments she shares with Sesha with the struggles of those affected by neglect and institutions. The meaningful, joyful moments they share challenge the notion that Sesha's life is defined only by her limitations. By caring for Sesha at home, Kittay helps others to recognise her humanity, which promotes a broader understanding of what it means to be human (Kittay 2001: 567).

In sum, Kittay believes that each person's identity is shaped by their relationships and the activities they participate in. This means our moral status and obligations come from these connections. She uses her own experience as a mother to her daughter Sesha, who has severe cognitive disabilities, as an example. Both Kittay and Sesha have identities formed by their relationship with each other (Mercer 2017: 15). Furthermore, Kittay argues that caregiving relationships are uniquely human and crucial for understanding each person's worth. These relationships involve a one-sided commitment to the other's well-being, where one person takes on the responsibility of meeting the needs of the other without expecting the same in return. To care effectively, one must be fully aware and responsive to the other's needs, making oneself transparent to those needs. Moreover, Kittay argues that these relationships, where one side may never be able to reciprocate the concern, are ethically significant because they are non-instrumental. Even if some people are not moral agents, Kittay maintains that their dignity nevertheless lies in their place within the moral community (Mercer 2017: 15).

In line with her account of care, Kittay (2005a) acknowledges that capacities like rationality and understanding what is good for oneself are important in a moral society. However, she argues that there are other important capacities that are often undervalued but are vital for our moral lives. For example, "caring and responding appropriately to caring, empathy and compassion, a sense of what is harmonious and loving, and the capacity for kindness and appreciation for those who are kind" (Kittay 2005a: 122).

Kittay argues that this is precisely where unjust marginalisation begins. For people with severe cognitive disabilities, like her daughter Sesha, are often not ascribed the psychological capacities required for full moral recognition. Kittay, on the other hand, emphasises that although Sesha is not recognised for her cognitive capacities, she "enriches the lives of others by her warmth, her serene and harmonious spirit and her infectious zest for life, and who has never acted maliciously or tried to harm anyone" (Kittay 2005a: 123). Kittay's point is that rationality and autonomy are often

mistakenly seen as necessary for moral worth, leading to the neglect of other important capacities. She proves her point by emphasising the positive qualities of Sesha and arguing that these traditional capacities are not the only ones that should determine moral status.

The main reason why these important capacities are neglected is the frequent mischaracterisation of people with severe cognitive disabilities. Kittay points out that they are not unresponsive beings who have no awareness of their surroundings, as some authors claim (McMahan 1996; 2002). She backs this up with the example of the behaviour of her daughter Sesha who is:

> *... enormously responsive, forming deep personal relationships with her family and her long-standing caregivers and friendly relations with her therapist and teachers, more distant relatives, and our friends. Although she will tend to be shy with strangers, certain strangers are quite able to engage her. (She has a special fondness for good-looking men!)* (Kittay 2005a: 126).

In describing whether Sesha can connect her past and future selves from "within", from her own life experiences, or has a narrative of her own, Kittay (2005a) argues that Sesha has a strong and distinct sense of self, even though her connections may be less strong than ours. In other words, even though Sesha's capacity to connect her past and future experiences is not as strong as others, she has a clear and individualised sense of who she is. Kittay claims that, despite her different cognitive capacities, Sesha's sense of personal identity and continuity is as coherent as her own. Given Sesha's situation, Kittay is deeply concerned about what lies ahead for her daughter. As Sesha may not be able to advocate for her own future, Kittay takes on the responsibility of representing and protecting Sesha’s interests as a third party. This mediating role allows Kittay to think not only about her own future, but also about Sesha's connection to it (Kittay 2005a: 128).

All this leads to the conclusion that when we refer only to certain capacities, we define the conceptual criteria for certain cognitive disabilities that exclude one from humanity (Kittay 2005a: 129). Furthermore, we contribute to the mentioned unfair mischaracterisation of people with severe cognitive disabilities.

When it comes to this false characterisation of people with severe cognitive disabilities, they are often compared to non-human animals. In other words, people with severe cognitive disabilities are sometimes according to Kittay - wrongly compared to non-human animals, implying they are less human in some way. In this context, she emphasises that this kind of comparison is unacceptable and nonsensical. Kittay emphasises that Sesha behaves like a human, not a dog. Sesha does everything she can, as a human would, often imperfectly, but it is "humanly imperfect, not canine perfect" (Kittay 2005a: 128). For example, despite all that Sesha cannot or seems unable to comprehend, according to Kittay, Sesha's receptivity to music and her

sensitivity to others have remained remarkably intact. Sesha's musical empathy is impressively sustained by the strange mix of gifts and drawbacks she possesses. This unevenness is common in many people with severe cognitive disabilities. According to Kittay (2005a), this is not a characteristic of the non-human animals with which they are associated (Kittay 2005a: 128).

Although Kittay (2005a) agrees that the non-human-animal/human-animal comparison is easiest to understand with primates because they are so similar to us, and certainly gorillas and some clever chimpanzees can do many things that her daughter cannot, Kittay cannot understand the comparison between humans with severe cognitive disabilities and dogs. That's because people simply do not know enough about what it's like to be a dog, to think like a dog, to perceive the world like a dog, or to compare a human's intelligence to an intelligence of a dog. On the other hand, Kittay acknowledges that no gorilla or dog, no matter how devoted she is to them, can be her daughter — with all the emotional, social, and moral implications that entails (Kittay 2005a: 130).

Because of the special emphasis on human relations, Kittay is exposed to the criticism of speciesism. Authors such as McMahan (1996; 2002) and Singer (2009) claim that speciesism as discrimination on the basis of species is comparable to nationalism and racism, which are also based on "group membership". That is, Kittay's focus on human relationships unfairly favours humans over non-human animals, much like nationalism or racism favours certain groups over others based on their group membership.

Kittay counters that nationalism and racism are not only about group membership, but also about a reliance on certain traits – "property types" – that are considered superior or exclusive to a group. For example, racism involves the belief that one ethnic group has desirable traits that another does not have or opposes (Kittay 2005a: 119). Kittay argues that the real problem with racism and nationalism is that they involve the belief that only one group has certain traits that entitle it to special privileges and power, leading to harm and division. She adds that focusing on intrinsic traits to define who is "us" and who is '"them" can be more problematic and discriminatory than focussing on membership of a particular species. Thus, Kittay believes that focussing on particular traits to establish moral worth is more reminiscent of racism or nationalism than simply recognising membership of a species (2005a: 121).

Accordingly, she claims that belonging to a family, rather than racism or nationalism, is the proper moral analogue for belonging to a community of moral equals based on belonging to a species (Kittay 2005a: 124). In other words, belonging to a family (or species) should be a fundamental basis for moral consideration, rather than focussing on traits that separate groups from one another.

To conclude the part about Kittay's approach, I would like to briefly emphasise what has been done. Namely, I have examined Kittay's response to the exclusion of

individuals with cognitive disabilities from moral personhood and explored her argument that species membership alone should determine moral status. In doing so, Kittay challenges the traditional view that cognitive capacities such as rationality and autonomy are prerequisites for moral consideration. Instead, she emphasises the importance of "social relations" and our common humanity, arguing for a moral framework in which being human, rather than possessing certain cognitive traits, is the basis for moral status. Kittay's focus on duties of caregiving and relational duties highlights how our relationships with others define our moral obligations and identity. She argues that relationships, such as that between a parent and a child, are fundamental to our understanding of moral worth and dignity. Kittay contrasts this with the often problematic comparisons between people with severe cognitive disabilities and non-human animals, arguing that such comparisons are both unjust and misleading.

Despite the strength of Kittay's arguments, there are, as I announced at the beginning of this section, unresolved problems and objections to her theory. One important issue is the fact that she relies on relationships of care and empathy to determine moral status, which some critics argue can lead to inconsistency and subjectivity (McMahan, 2002; Nussbaum, 2006). There is also a danger that Kittay's framework may unintentionally reinforce paternalistic attitudes or perpetuate stereotypes of dependency and vulnerability. Although Kittay defends her approach against accusations of speciesism by emphasising the unique moral significance of human relationships, I believe that her response does not fully address the broader ethical concerns about speciesism. This is because speciesism not only involves an unjustified prioritisation of human interests over those of non-human animals but also calls into question the moral hierarchy that places the interests and welfare of humans above those of non-human animals without sufficient justification (Martinić 2020). In short, the defence of Kittay's position requires further justification of speciesism (Mercer 2017: 27).

In the following discussion, I will examine these objections to Kittay's theory in more detail. Specifically, I will critically analyse her arguments regarding speciesism and the moral status of humans with cognitive disabilities in comparison to non-human animals. As noted earlier, Kittay asserts that species membership plays a central role in forming moral bonds. I will challenge the validity of this claim and propose a broader, more inclusive approach to justice and rights. I will also address the ambiguity in Kittay's concept of "doing something in a human way" and evaluate evidence of cognitive and emotional capacities in non-human animals that undermine her assertions. In contrast to Kittay's assumptions, I will argue that the bonds between humans and their pets demonstrate deep emotional connections that transcend species boundaries, and that these relationships emphasise the need for a broader view of

moral status inclusion[25]. Furthermore, I will raise concerns about the paternalism implicit in Kittay's framework and question whether it unintentionally reinforces harmful hierarchies and stereotypes. In addressing these counter-arguments, I would like to argue for a re-evaluation of our just treatment of both humans with cognitive disabilities and non-human animals.

Let me begin my rebuttal by first examining Kittay's concept of doing something in a "human way". To critically engage with her concept of "doing something in a human way," it is necessary to unpack several layers of her argument and address the ambiguities and challenges it presents. As a reminder, Kittay uses the term "in a human way" to distinguish human interactions and experiences from those of non-human animals. In particular, she (2005b) refers to this concept when discussing how her daughter Sesha engages with music, suggesting that Sesha's interaction with music is uniquely human in comparison to a dog's interaction. The term "in a human way" is central to Kittay's argument, but it is inherently ambiguous. This ambiguity arises from the following points: diverse human experiences, cultural and personal differences and the lack of a clear definition.

Firstly, it is an undeniable fact that human interaction with music spans an incredibly broad spectrum – it is multi-layered. In other words, people engage with music in different ways and music is not only an auditory phenomenon, but also a cultural, emotional and intellectual one. A conductor's approach to a symphony, for example, may involve an intricate interpretation of the composer's intent, incorporating historical context, music theory and personal expression to guide an orchestra's performance. In contrast, a layperson may connect with the same piece on a personal, emotional level, using it as a source of comfort, inspiration or entertainment. These different ways of engaging with music reflect the complex ways in which people interact with music. Furthermore, the diversity of human responses to music also extends to the cultural and social dimension. Different cultures have different musical traditions, and people within these cultures may attach different meanings and importance to the same piece of music. The communal experience of music, whether in religious rituals, celebratory occasions or shared moments of listening, emphasises the importance of music in shaping our identity and fostering social relationships. This diversity further complicates the notion of a single "human way" Kittay's concept of what constitutes "the human way" lacks a clear, detailed definition. Without a precise explanation of what characteristics or criteria define this "human way", the concept remains vague. This ambiguity makes it difficult to understand how and why certain behaviours, such as Sesha's engagement with music, should be considered uniquely human in comparison to behaviours of non-human animals.

[25] I retain the notion of *moral status inclusion* in line with Kittay's methodology, but also use it here to refer to the status of a being that should be included in the domain of justice.

Given this ambiguity, Kittay's argument runs into several problems: unclear foundation, potential overgeneralisation, and need for nuanced explanation.

For Kittay's thesis to hold, it must establish a solid framework that defines what "in a human way" means. This framework should encompass the various ways in which humans engage in activities such as music, while also accounting for the various ways in which non-human animals engage in similar activities. Furthermore, the concept of "in a human way" could lead to overgeneralisations. If the term is too broad or poorly defined, it risks being used to exclude some beings from certain moral considerations simply because their engagement in activities might differ from the experiences of others. Kittay could argue that variations in human activities still fall within the realm of the "human way" In this case, she could argue that Sesha's actions, even if they exhibit some variation, still fall within the realm of these human variations. However, she would need to explain in more detail how these variations fit into her framework and how they relate to Sesha's behaviour. A nuanced explanation is essential to avoid the accusation that the concept is too vague to support her argument. Thus, for Kittay's argument to be valid, she would need to clarify how the concept of doing something "in a human way" accounts for this variation. To do this, she would need to identify the characteristics or criteria that define what is truly "human" in these activities and how they relate to Sesha's behaviour. Kittay's task would be to provide a more robust framework for understanding and identifying what constitutes "the human way" in different activities. This framework would ideally encompass the various ways in which humans engage in these activities, while including beings such as Sesha. Until this is clarified, the concerns and objections regarding the ambiguity of this concept will remain as valid criticisms of their argument.

The problem with Kittay's definition of "doing something 'in a human way' is further deepened by the evidence that certain non-human animals exhibit behaviours and cognitive capacities that Kittay characterises as uniquely human. This problem poses several significant challenges to her argument. First, there is an overlap in cognitive capacities. Indeed, research (Morell: 2008; de Waal and van Roosmalen: 1979; Melis, Hare, and Tomasello: 2006a; 2006b) has shown that many non-human animals have cognitive capacities and behaviours that were previously thought to be unique to humans.

Dogs, for example, are a convincing example of this phenomenon. In the book "Minds of Their Own" (2008), researchers present convincing evidence that dogs like Rico have "uncanny linguistic capacities". Rico had the extraordinary capacity to learn and recall words as quickly as a human child. This capacity is considered a fundamental building block of language acquisition, and Rico's approach in this regard was very similar to that of humans. Remarkably, the researchers discovered similar linguistic capacities in other dogs, such as Betsy, who had an extensive vocabulary of almost three hundred words. Most remarkably, even our closest relatives, the great apes,

could not match Betsy's remarkable capacity to hear a word just once or twice and recognise its representation based on the audio pattern.

I argue that the discoveries outlined above pose a direct challenge to Kittay's claim that certain capacities are exclusive to humans and are not present in non-human animals. I question whether the capacities Kittay refers to, such as caring, appropriate responses to caring, empathy, compassion, a sense of harmony and love, and the capacity for kindness and appreciation of those who are kind (cited in Kittay 2005a: 122), are not present in non-human animals.

Indeed, if non-human animals exhibit similar behaviours or cognitive capacities this undermines the claim that these traits define what it means to do something "in a human way." This evidence calls into question the validity of distinguishing human behaviours as fundamentally different from those of other animals. Moreover, it leads to a redefinition of human uniqueness. The presence of these human-like traits in non-human animals suggests that the criteria used to define what “human” way is may need to be re-evaluated. If behaviours and cognitive capacities are shared by all species, then the concept of “doing something in a human way” is less about exclusive traits and more about variations within a spectrum of cognitive and emotional experiences. When non-human animals exhibit traits that Kittay associates with human interactions, it challenges the moral hierarchy that prioritises human experiences over those of other animals. Kittay’s framework may unintentionally reinforce speciesist attitudes by implying that the cognitive and emotional capacities of nonhuman animals are less significant, despite evidence to the contrary.

Therefore, I would like to conclude my first counter-argument by stating that the evidence that nonhuman animals share cognitive and emotional qualities with humans emphasises the need for an ethical framework that recognises these shared qualities. Kittay’s definition, which relies on the uniqueness of human behaviours, may not do justice to the moral significance of non-human animals that share similar traits.

My second counter-argument deals with the possible extension of Kittay's concept of "social relations". Namely, I believe that it can be challenged by examining the strong emotional bonds between humans and non-human animals, especially pets. This counter-argument suggests that Kittay's framework can be extended to these non-human relationships, showing that such bonds are not only meaningful but also consistent with her emphasis on deep emotional connections.

Pet ownership[26] is a compelling example of how deep emotional bonds can develop between humans and non-human animals. Contrary to Kittay’s view that such bonds are limited to human relationships or familial bonds, the relationship between a human

[26] I have left the accepted term pet "owner" even though it is a problematic term, but only so as not to divert focus from my main point, which is the possibility of the special relationship Kittay argues with other non-human animals.

and a pet often reflects the depth and importance of familial bonds. People who own pets often refer to their companions as beloved family members, emphasising the deep emotional connection they experience. For example, their primary concern is for the safety and well-being of their animal companions. This level of commitment reflects a strong bond that is not limited to the human species or familial ties and challenges the notion that humans can build or value such meaningful relationships only with other humans.

Furthermore, the increasing recognition of pets as valued members of households illustrates a shift in the way society recognises non-human animals. Terms such as "pet'," "companion" and "friend" signify more than mere ownership — they indicate a moral and emotional esteem that reflects people's familial relationships. Furthermore, this recognition is also reflected in legal protections and societal attitudes that increasingly focus on the welfare of pets and recognise their status as individuals with significant emotional value. This distinction plays a crucial role in how these animals are treated (Alvaro 2017: 769). For example, pet owners often go to great lengths to ensure the well-being and satisfaction of their animals, from veterinary care to emotional support. These actions demonstrate a deep sense of responsibility and affection comparable to the care and concern typically reserved for human family members. Indeed, what distinguishes a pet from other animals is essentially the attribution of certain human-like characteristics, including a distinct personality (Sunstein and Nussbaum 2004: 97). This attribution goes beyond simply recognising the existence of an animal; it recognises that pets have unique, individual characteristics that make them special in the eyes of their human companions.

The naming of pets is another aspect that emphasises their importance. The act of naming a pet emphasises its individuality and signifies a personal relationship. This practise is in contrast to the way livestock or laboratory animals are often treated, where they are usually seen as resources rather than individuals. The naming of pets not only emphasises their unique identity but also reflects the deep emotional bonds that their owners form with them (Sanders 2003: 411).

The bond between humans and their pets is particularly evident in difficult situations. People experiencing homelessness, for example, often prioritise the care of their pets despite their own limited resources. This phenomenon therefore extends not only to situations in which the conditions for keeping pets are optimal, but also to situations in which basic needs are scarce. Homeless people who keep a pet emphasise how important this companionship is for their psychological well-being. A study conducted in Sydney, Australia (2021) looks at the life experiences of homeless people who have kept a pet despite the difficulties they face. It becomes clear that pets serve as a protective shield against social isolation, alienation, loneliness and mental health problems. At the same time, the bond between humans and non-human animals provides pet owners with unwavering affection, emotional stability and a heightened sense of security (Cleary et al., 2021).

When examining the deep emotional bonds between humans and pets, it becomes clear that these relationships can and should be understood within Kittay's concept of "social relations" The emotional depth and moral significance of pet ownership refutes the notion that such bonds are limited to human or familial contexts. Instead, these relationships show that strong emotional bonds can transcend species boundaries. They call for a broader understanding of the value and recognition of non-human companions in discussions of social bonds and moral worth.

The third counter-argument against Kittay's theory centres on the concern that it may inadvertently support paternalistic views and reinforce stereotypes of vulnerability and dependency, especially in relation to people with disabilities. While Kittay emphasises the importance of compassion and empathy in recognising the moral worth of people, particularly people with disabilities, this emphasis may inadvertently encourage prejudice and paternalistic attitudes.

A major problem with Kittay's theory is that it defines the moral worth of people with disabilities primarily in terms of their dependence on the care and support of others. This focus on dependency can reinforce social norms that view people with disabilities as passive recipients of care rather than active agents with autonomy and agency. By emphasising the need for care in the lives of people with disabilities, Kittay's approach risks perpetuating stereotypes that view these individuals as inherently vulnerable and unable to contribute meaningfully to society beyond their role as care recipients.

Lennard Davis (2016), for example, discusses how cultural representations and historical contexts have shaped societal perceptions of disability, often leading to stereotypes that marginalise and stigmatise people with disabilities. Thus, Kittay's theory, by emphasising dependency, may inadvertently contribute to these stereotypes by portraying people with disabilities as people in need of constant care rather than as individuals capable of autonomy and self-determination.

Another concern is that Kittay's emphasis on caring and empathy as primary determinants of moral status may undermine efforts to promote the independence and self-determination of people with disabilities. By prioritising caring relationships over factors such as autonomy, Kittay's framework may marginalise the voices and experiences of people with disabilities who assert their right to make decisions about their own lives.

Batavia (2001) argues that portraying people with disabilities primarily as an oppressed minority may inadvertently support paternalistic views that undermine their autonomy and dignity. Kittay's focus on care could therefore be seen as an unintended contribution to a representation that emphasises the dependency of people with disabilities, potentially sidelining their ability to advocate for their needs and preferences. This could lead to a form of well-intentioned paternalism that portrays

people with disabilities as weak and dependent rather than as capable, self-determined individuals.

Furthermore, Kittay's theory risks essentialising the experiences of people with disabilities by focusing on care as the primary basis for moral inclusion. This perspective overlooks the diversity of life experiences of people with disabilities. Not all people with disabilities rely on caring relationships to find moral worth or fulfilment. Some find meaning in their independence, work, creativity or other aspects of their lives that do not involve dependence on the care of others.

By emphasising care, Kittay's theory may inadvertently exclude those who do not fit into this framework of care, thus limiting the recognition of the full range of experiences and contributions of people with disabilities. This could lead to an oversimplified understanding of disability that fails to recognise the multiple ways in which people with disabilities experience and express their moral worth.

To avoid these pitfalls, for the reasons outlined above, it is crucial to recognise the agency and autonomy of people with disabilities alongside their need for care, and to do so in a way that respects their full humanity and diverse experiences.

The final counterargument against Kittay's view centres on her intuition regarding the moral status of nonhuman animals in comparison to humans, especially persons with severe cognitive disabilities. I argue that Kittay's view, which ascribes a significantly weaker moral status to nonhuman animals, is based on a mischaracterisation of these animals. This mischaracterisation has significant ethical implications and leads to a problematic attitude towards nonhuman animals that reflects exactly the same kind of marginalisation that Kittay criticises when it comes to people with severe disabilities.

Namely, Kittay's argument that non-human animals have a much weaker moral status than humans with severe cognitive disabilities is based on a false or incomplete understanding of non-human animals. A false or incomplete understanding of non-human animals refers to misconceptions or limited perspectives about the nature, capacities, and experiences of nonhuman animals. This has already become clear, for example, in the description of the underestimation of the cognitive capacities of non-human animals. The assertion that non-human animals do not have complex cognitive capacities, such as problem solving, tool use or the ability to form social bonds is false; in reality, many species have these capacities. As we have seen, there is also a common misconception that non-human animals do not feel emotions like humans. However, research shows that many animals exhibit behaviours that indicate that they feel emotions such as joy, grief, empathy and fear. Another common but flawed belief is that humans have a unique moral status that places them above non-human animals. This view often ignores the sentience and capacity for suffering of non-human animals and thus justifies their exploitation. All this leads to the social complexity being overlooked when it is assumed that non-human animals have simple, instinct-

driven social structures. In contrast, many species, especially primates, elephants and whales, have complex social hierarchies, communication methods and even cultures.

This mischaracterisation is problematic because it supports an attitude that rejects the moral worth of animals, which is similar to the dismissive attitude that some people have towards people with severe cognitive disabilities. The central criticism is that Kittay's perspective assumes a clear moral distinction between humans and animals based on the asserted superiority of human characteristics. However, this distinction not only unjustifiably misrepresents the capacities and experiences of non-human animals, but also perpetuates a hierarchical view of moral status that exalts humans and diminishes the intrinsic value of other sentient beings. The concept of *species narcissism*, introduced by philosophers Kymlicka and Donaldson (2014), describes this belief that humans possess a fundamentally superior moral status purely by virtue of their humanity. This belief sustains a rigid moral hierarchy that places humans above all other life forms, enabling the marginalisation and exploitation of non-human animals. Moreover, species narcissism parallels the marginalisation of human groups based on perceived deficits in traits such as cognitive ability or autonomy. By asserting that certain characteristics—like rationality—determine moral worth, this anthropocentric worldview not only harms non-human animals but also perpetuates other forms of unjust discrimination (Kymlicka and Donaldson, 2014). Kymlicka and Donaldson (2014) present a compelling counter-argument to species narcissism, advocating for the recognition of the shared vulnerabilities and capacities of humans and non-human animals. They argue that if we acknowledge that all sentient beings are capable of suffering harm and experiencing well-being, it becomes evident that the welfare of humans and non-human animals is interconnected rather than fundamentally distinct. By adopting their inclusive view, moral worth is no longer determined by the degree to which a being resembles certain human traits. Instead, it is grounded in the inherent capacity of sentient beings to experience life, suffering, and well-being.

The discomfort that Kittay and others may experience when comparing humans to non-human animals often stems from societal attitudes towards non-human animals. As Oliver (2020: 117-118) observes, objections to such comparisons reflect not an inherent moral distinction but the negative connotations associated with how non-human animals are treated. Widespread exploitation and mistreatment of non-human animals create a societal context in which analogies between humans and non-human animals are seen as offensive. However, this discomfort exposes deeper ethical contradictions in how society values different forms of life. If non-human animals were treated with the same respect and dignity as humans, such comparisons would likely evoke understanding and compassion rather than unease. Oliver's (2020) reasoning suggests that the resistance to these analogies is rooted in societal speciesism rather than any substantive moral difference.

To conclude this section, I would like to briefly summarise what I have done so far. I have presented several counter-arguments to Kittay's claims and proposed a broader and more inclusive framework that challenges the limitations of her approach. Firstly, I critiqued the ambiguity of Kittay's concept of "doing something in a human way." By highlighting the diverse and complex ways in which humans engage with activities such as music, alongside the lack of a clear definition in Kittay's argument, I demonstrated that this concept is too vague to support her assertion that certain behaviours are uniquely human. Furthermore, evidence of similar cognitive and emotional capacities in non-human animals undermines the exclusivity of these traits to humans, challenging the moral hierarchy Kittay seeks to establish. Secondly, I argued that the deep emotional bonds between humans and their pets demonstrate that Kittay's concept of "social relations" can and should extend to non-human animals. These relationships, which often mirror familial bonds, reveal that meaningful and morally significant connections are not confined to humans. This broader understanding of moral worth transcends species boundaries and invites a more inclusive ethical perspective. Thirdly, I addressed concerns that Kittay's framework may inadvertently reinforce paternalistic views and stereotypes of vulnerability and dependency, particularly with regard to people with disabilities. By prioritising care and dependency as primary criteria for moral inclusion, Kittay risks perpetuating harmful stereotypes and marginalising the autonomy and agency of individuals with disabilities. A more balanced ethical framework should integrate the recognition of agency and diverse experiences of people with disabilities alongside the need for care, avoiding these pitfalls. Finally, I criticised Kittay's claim that non-human animals have a significantly weaker moral status than people with severe cognitive disabilities do. I argued that this perspective mischaracterises non-human animals and perpetuates a hierarchical view of moral status rooted in speciesism. In addition, I argue that by confronting and dismantling speciesist assumptions, we create space for a more equitable relationship with all living beings, addressing the moral contradictions inherent in traditional anthropocentric worldviews. As Kymlicka and Donaldson (2014) argue, this shift is essential for building a society that respects the dignity of all sentient beings and promotes their capacity to thrive.

In conclusion, while Kittay's arguments provide valuable insights into the moral inclusion of people with cognitive disabilities, they fall short in addressing the broader implications of speciesism and the moral status of non-human animals. A more expansive perspective, one that acknowledges the shared capacities and moral worth of all sentient beings, is necessary to foster a just and compassionate framework.

In the next section of the chapter, I will examine responses and theories that have attempted to step forward by including non-human animals and gain insight into the development of Rawlsian thought and its implications for animal welfare.

B. Expanding Justice to Non-Human Animals: Rawlsian Theoretical Approaches and Limitations

There are already proposed Rawlsian approaches that attempt to include non-human animals within the domain of justice, but I will not delve into all of them here. I present some of the most representative examples of the debate. By comparing the proposals and, mainly, the criticisms they address to each other I will justify the need for a new proposal, as the one I offer. In this section, I am focusing on the contributions of Mark Rowlands (1997), Abbey Ruth (2007), Brian Berkey (2017), and the political turn discussed by Cochran, Garner, and O'Sullivan (2018). Each of these scholars offers a distinct perspective on how nonhuman animals can be incorporated into justice theory, particularly Rawlsian contractarianism, and how moral deliberation can transcend rational actors.

While these approaches are valuable, they also encounter significant limitations. I will argue that Rowlands' reliance on contractarianism, Ruth's non-rights-based perspective, Berkey's critique of basic assumptions, and Cochran, Garner, and O'Sullivan's policy turn focus on institutional reform provide important insights but do not fully resolve the challenges of integrating animals into justice. These theories either fail to provide a comprehensive theoretical basis for the integration of animals into justice, ignore practical applications, or attempt to reconcile moral considerations with systemic change.

A problem that needs to be addressed when extending justice to include non-human animals is the increasing problem of competition for resources and protection. This is not a problem at the general level of idealized justice that I discussed so far. But it is a problem in the real world. A coherent framework that can bridge the gap between theoretical principles and real-world applications is crucial for addressing the inclusion of non-human animals in scope of justice. In the next section, I present an alternative approach that distinguishes between ideal and real-world justice and offers a more structured model that not only addresses the theoretical challenges but also suggests practical avenues for legal and institutional reform. The intention is to offer a model that better supports the inclusion of nonhuman animals in the moral community and help to ensure that justice transcends human interests. In this section I will continue with the analysis of the above theories, starting with Mark Rowlands' approach.

Rowlands (1997), in *Contractarianism and Animal Rights*, presents an alternative perspective on incorporating animal welfare within Rawlsian justice. He argues that contractarianism, especially Rawls's version, provides a theoretical basis for including non-human animals in moral frameworks. This challenges the idea that Rawlsian justice is solely about human agents, expanding the concept of justice to include animals.

I consider Rowlands' argument important because it aligns with the idea I aim to demonstrate, which I emphasized through the IRA model. Namely, Rowlands criticizes the orthodox contractarian argument that restricts moral status to rational agents, excluding non-human animals from direct moral consideration. Rowlands questions the assumption that rationality is a morally relevant property that justifies this exclusion. The argument from intuitive equality states that moral status should not depend on properties that individuals possess purely by chance or without merit. Since rationality is considered an undeserved property, it is not a valid basis for excluding non-rational beings from the moral community. Thus, if justice is based on such morally arbitrary properties, then the scope of moral consideration should be extended to all beings who may experience suffering or well-being. In other words, Rowlands argues that rationality is an arbitrary property that does not deserve special moral status and therefore should not serve as a basis for limiting the scope of the social contract. Furthermore, he emphasises Rawls' definition of moral persons as beings capable of having a conception of the good and a sense of justice, rather than as purely rational agents. This broader definition suggests that equal justice should not be limited to humans but should extend to all beings capable of moral personhood. This idea is very similar to those that I expose when I present the IRA model. Namely, it reveals that agents engaged in the construction of justice are not only concerned with their own advantages (rational), but also with justice as such (reasonable). In other words, Rowlands argues that the principles of justice derived from the social contract should apply not only to those who formulate the contract (rational agents), but also to those who are affected by it. Non-human animals fall into this moral community as sentient beings that can experience suffering and well-being. Rowlands' approach therefore extends the principles of contractarianism to all sentient beings, regardless of their rational capacities (Rowlands 1997).

This emphasis on sentience is key to Rowlands' argument. He claims that the capacity for pleasure and pain — sentience — is a morally relevant property that grounds moral status. According to Rowlands, sentient beings have interests that must be considered in any moral framework, and as such, non-human animals should also be included within the scope of justice. By linking moral consideration to sentience, Rowlands argues that contractarianism can be an effective tool for extending rights and moral protection to non-human animals (Rowlands 1997).

Although I agree with Rowland's strategy that widely corresponds to my own proposal, he leaves unresolved the problem that I remark above, i.e., the increasing conflicts and competition for resources and protection of rights when we extend beneficiaries of rights. This is a problem that I resolve by distinguishing between ideal and real-world justice. Ideal justice extends moral consideration to all sentient beings, recognizing their inherent value and right to protection. However, real-world justice must also account for practical considerations, such as limited resources, clashes among rights, legal reforms, societal norms, and institutional frameworks that can

ensure these moral principles are actually realised. By distinguishing between these two levels of justice, I aim to address both the theoretical and practical challenges of extending justice to non-human animals, which Rowlands' framework does not fully resolve. While Rowlands provides a useful starting point, my approach goes further by rethinking the very principles of justice and proposing mechanisms for change that can make such inclusivity a reality in practice. Before fully presenting my own solution, I will examine further Rawlsian approaches that criticise Rowlands' attempt to include animal welfare within justice while maintaining Rawls's original ideas. I will begin by exploring Abbey Ruth's approach (2007), then move on to Brian Berkey's perspective (2017), and finally consider the political turn highlighted by Cochran, Garner, and O'Sullivan (2018) in the discussion of justice for non-human animals.

Abbey Ruth (2007) highlights a crucial gap in Rawls' theory of justice: while Rawls did not include non-human animals as participants in his framework, he acknowledged that humans have moral obligations towards them (Rawls 1995:20). This observation is significant because it opens the door for extending principles of justice beyond human society, which is the main aim of this section. Therefore, discussing Ruth's approach is essential to my argument, as it provides a foundation for challenging the anthropocentric limits of Rawlsian justice and advocating for the inclusion of non-human animals within a moral and political framework. This exclusion, however, has been widely criticised, with scholars such as Robert Garner (2003; 2012) and Tom Regan (1981; 1983) arguing that it is both arbitrary and inconsistent with the egalitarian ideals Rawls promotes.

Ruth's argument can be divided into three main premises. The first is that she takes Rawls's remarks about the moral status of animals at face value and attempts to draw out their significance in a manner that seems more consistent with Rawls's original design. Here, we see her departure from Rowlands (1997), who Ruth believes strayed too far from Rawls's original intention. Ruth explains this first premise by highlighting that Rawls views human and non-human animal relationships within the domain of morality, rather than justice. She elaborates this by suggesting that duties toward animals arise from their capacity for pleasure and pain, possibly drawing on a utilitarian perspective or Aristotelian ideas about sociability. A key question then arises: if duties toward animals are not grounded in the social contract, where do they come from? Rawls does not provide a detailed answer but refers to these obligations as "considered beliefs".[27] This aligns with his broader method of reflective

[27] Here I follow Ruth's interpretation, but Rawls uses the concept of "considered judgements", which refers to moral intuitions that reflect our deepest sense of fairness and arise when individuals reflect on justice impartially free from bias. These judgements, based on a consistent and thoughtful application of principles of justice, serve as the basis for the construction of principles of justice. If there is a discrepancy between these judgements and the principles, the principles may need to be revised to be consistent with them (Rawls, 1999).

equilibrium, where principles of justice must be balanced with widely held moral convictions. Ruth, relying on this method, emphasises that in many societies, there is a shared belief that animals matter morally, reflected in laws against cruelty and institutions for animal welfare.

The second premise is that Ruth maintains, rather than elides, the distinction between justice and morality. She proposes that Rawls's notion of "justice as fairness" could be extended to include moral consideration for non-human animals. Namely, she argues that while Rawls's theory primarily addresses rights and justice between rational agents, this does not prevent it from covering moral duties towards animals, such as preventing cruelty or unnecessary harm. Her key argument here is that she does not wish to frame her theory in terms of rights, as rights-based theories tend to focus on legal entities, duties, and obligations. Instead, she proposes a non-rights-based perspective. This approach, she argues, is more flexible than rights-based theories because it accommodates the needs of animals and diverse cultural understandings of how to treat animals.

Third premise is focused on Ruth comparing the relevance of *A Theory of Justice* and *Political Liberalism* for animal ethics, highlighting that both Garner and Rowlands focused solely on the former, despite *Political Liberalism* being available at the time of their writings. Ruth critiques John Rawls' *Political Liberalism* for its failure to extend justice-based moral consideration to animals. She argues that Rawls' commitment to pluralism, while central to his political philosophy, ultimately weakens the normative foundation for animal ethics. In *A Theory of Justice*, Rawls provides a framework for human rights and justice but excludes non-human animals from this sphere. In *Political Liberalism*, he reinforces the idea that justice applies only to reasonable comprehensive doctrines, further marginalizing concerns about animal welfare. Ruth suggests that this exclusion undermines efforts to protect animals from cruelty and exploitation. Her argument suggests that pluralism, as Rawls defines it, is both a strength and a limitation. On one hand, it allows for a diversity of moral and cultural perspectives and their equality in the political domain. On the other, it risks justifying morally questionable practices—such as animal cruelty—if they are embedded within "reasonable" doctrines. This paradox raises the question of whether a more inclusive version of justice should extend beyond human interests. This is why she implies that instead of trying to fit animal ethics into Rawlsian liberalism, it may be more productive to explore ethical traditions that are already more attuned to non-human concerns.

In conclusion, while Ruth's approach offers valuable insights into how Rawls's theory can be extended to include non-human animals, it does not come without its limitations. I will now outline what I believe is wrong with her approach, which will then set the stage for my own proposal.

Ruth's non-rights-based perspective is flawed because, contrary to Ruth, I argue that rights are central to discussions about justice. Rights provide a clear and structured means of addressing obligations that are significant for the beings to whom they apply. While Ruth highlights that rights-based approaches focus on legal duties, I believe their strength lies precisely in the fact that they offer a precise and legally enforceable framework. This is why I have argued that IRAs (Ideal Reasonable Agents) universalise rights to ensure that obligations of justice are both clear and enforceable, while still maintaining flexibility to accommodate different perspectives.

Furthermore, I believe Ruth's proposal lacks a clear foundation for justice-based inclusion. Although it is valuable for initiating the discussion, an alternative justification is necessary to provide a more robust theoretical framework. A further problem is that Ruth's approach is not sustained through the kind of justification needed in the space of controversial issues and it seems to rely solely on moral intuitions, which raises concerns. It is unclear who decides on moral considerations when they are controversial.

In addition, Ruth's theory does not explain what kind of protection non-human animals should be granted – should it be based on sentience, cognitive abilities, human interests, or something else? For these reasons, I will argue for a stronger theoretical foundation with clearer policy applications.

Next, I will explore Brian Berkey's (2017) critique of traditional justice theories, particularly in relation to their exclusion of non-human animals. Berkey challenges the foundational assumptions that prevent the inclusion of non-human animals in moral and political frameworks. His analysis identifies three key obstacles: the "contribution/capacity basis of entitlement," political liberalism, and institutionalism. These critiques are central to understanding the limitations of existing theories and provide an important perspective on why a more inclusive approach to justice is necessary.

The first challenge Berkey identifies is the "contribution/capacity basis of entitlement." Traditional justice theories often tie entitlement to an individual's ability to contribute to a cooperative social system. Since in such theories non-human animals are assumed to lack the ability to contribute in the same way humans can, they are excluded from direct entitlements. This exclusion reflects an inherent bias, as it assumes that the capacity to participate in social systems is a justifiable basis for granting rights. However, Berkey argues that this assumption is arbitrary and unjust, as it disregards the intrinsic value of non-human animals, who, despite their inability to contribute in human terms, have interests and well-being that deserve protection (Berkey: 2017).

The second challenge arises from political liberalism, particularly in the Rawlsian tradition. Rawls's theory of justice as fairness is predicated on the idea of social cooperation among free and equal citizens. This model excludes non-human animals

because they are not considered autonomous or rational beings capable of engaging in such cooperation. By focusing on human citizens as the primary subjects of justice, political liberalism inherently marginalises animals, making it difficult to extend justice to them without significantly revising its core principles (Berkey: 2017).

The third challenge is institutionalism, which applies the principles of justice exclusively to social institutions and not individual behaviour. The problem is that, in a political liberal view as Rawls's, applying principles of justice to institutions and requiring their enforcement demands justifiability to all reasonable persons, or, we can say, coherence with all reasonable doctrines. The problem is to show that this can be done in the case of non-human animals. This is why it is important to have the possibility to engage an alternative solution for the protection of non-human animals that is not burdened by such strong demands of justification. This could be appealing to direct entitlements to which it must be responded by individual responsibility. Thus, the Rawlsian view must be complemented by this additional normative frame (Berkey: 2017).

These challenges illustrate the limitations of adapting existing theories to include non-human animals. Berkey contends that such attempts are insufficient because the foundational assumptions of these theories—such as the emphasis on human capacities, social cooperation, and institutional frameworks—are inherently exclusionary. Rather than attempting to stretch these frameworks to accommodate animals, Berkey advocates for a radical rethinking of justice itself. This would involve moving away from anthropocentric principles and developing a more inclusive theory that recognises the moral relevance of sentience and the interests of all beings, human and non-human alike.

I believe my approach is not subject to Berkey's criticisms, by offering a more structured framework for rethinking justice—one that extends consideration of justice beyond individuals that exercise capacities for social cooperation and recognises the interests and welfare of all sentient beings. The advantage of my proposal is to simplify the justification of non-human animals' rights. As I show in the explanation of the IRA model, non-human animals' rights follow from general reasoning about justice as a coherent part of justice, without the need of additional sources of norms.

Last approach I will analyse is by Cochrane, Garner, and O'Sullivan (2018), that represents exactly the inverse view of Berkey's criticism of institutionalism. I believe their approach is interesting exactly because of this, i.e., because they examine the shifting discourse in animal welfare, particularly regarding the "political turn." This refers to the increasing incorporation of political concepts and language into discussions of animal welfare, moving beyond traditional moral philosophy to a framework that addresses systemic and institutional justice for animals. They (2018) identify key aspects of this political turn: relationships and positive duties, pragmatism, and the avoidance of first principles. They argue that this shift signifies

a movement towards recognising animals' interests within political institutions and structures. Instead of merely advocating for individual moral obligations, the political turn emphasises the importance of institutional action to enforce justice for animals.

Cochrane, Garner and O'Sullivan (2018) point out that this shift emphasises the need for political action and structural change to ensure justice for animals. Animal welfare is no longer just a matter of personal morality, or individualized responsibility, but a political issue that requires systemic reform. On this point, they differ from Ruth's (2007) and Berkey's (2017) approaches and emphasise the increasing focus on justice in the field of animal welfare. They note that achieving justice for animals requires more than moral conviction and individual responsibility — it requires political action and institutional change. While the political turn has brought new dimensions to the discussion, the authors (2018) believe that it has not yet fully crystallised into a coherent and distinct movement. They suggest that future research should examine the role of political institutions, the political economy that supports animal interests, and how non-human animals can be meaningfully represented within these structures. Emphasising justice in these discussions is crucial as it reshapes political structures to benefit both humans and non-human animals, going beyond lifestyle reforms and individual moral obligations.

While I agree with the authors on the need for systemic change, their approach is not entirely satisfactory because it overlooks the necessity for a clear framework of justice that could guide political action. Without a well-defined justification for the inclusion of animals, political action can become fragmented and less effective. Again, the IRA model offers a more coherent solution, providing a structured, rights-based approach that can both guide political reforms and ensure the inclusion of animals within the scope of justice.

In conclusion, the exploration of these various approaches reveals both the potential and the difficulties of extending Rawlsian justice and other traditional theories to include non-human animals. While Rowlands, Ruth, Berkey, and Cochran et al. have significantly advanced the conversation, their approaches do not fully address the theoretical and practical complexities of integrating non-human animals into justice. They either fail to challenge the foundational assumptions of justice, or lack a clear justificatory framework.

With this I have completed my attempt to show the implication of the IRA model requiring the inclusion of non-human animals in the scope of justice. Besides, I have shown why the IRA model is more successful than some prominent Rawlsian proposals in achieving this inclusion. The problem that I address, now, is related to real-life implementation. Precisely, this problem appears in two forms. First, the question whether this implementation could be realistic, considering actual relations in society and its organization. Second, the issue of conflicts regarding rights among beneficiaries of justice in an extended view of inclusivity.

C. Ideal Justice and Real-World Justice: The Case of Non-Human Animals in Society

In this section, I explore the distinction between ideal justice and real-world justice, particularly in relation to non-human animals. This distinction is vital because, while ideal justice provides a normative framework for treating all beings fairly, real-world justice considers what is practically achievable within existing social, economic, and political constraints. As mentioned earlier, my focus here is on practical strategies rather than normative balancing. To this end, I adopt the approach of "doing the best we can, starting from where we are" (Wolff: 2011).

Ideal justice represents a world in which the principles of justice can be applied universally and without exceptions. In this ideal framework, all living beings in need of protection — whether human or not - are treated fairly, regardless of their species belonging, their cognitive abilities or their usefulness in social cooperation. This is ensured by the hypothetical construction of Ideal Reasonable Agents (IRAs), impartial and rational decision-makers who determine the fairest course of action. From this perspective, for example, unnecessary cruelty such as factory farming, medical animal testing or the use of animals for aesthetic purposes would be prohibited, as all beings that can feel pain have a right to protection from suffering.[28]

However, since I am not only concerned with a theoretical description of the ideal justice system, but would like to offer a proposal with practical benefits, I must also consider the real-world level of justice at a certain stage of the deliberations, i.e. a pragmatic method of feasibility within the existing social relations. Thus, I must engage with real-world justice and step into the real world. Real-world justice does not consider which principle or value is stronger in principle, or which principle we should "save" only for principled reasons. Instead, it concerns what is actually achievable right now. Therefore, I am concerned with what can be realistically implemented from those ideal principles—having in mind a sense of justice but also considering what is feasible in the present time. This includes the reality that certain things—such as people's strong resistance—will require us to, to some extent, compromise on justice because people firmly reject certain concepts of justice. In this context, the focus is on practical strategies, specifically how feasible a given policy is

[28] The argument presented here challenges the notion of universalizing the right to life, asserting that it is not always applicable in a blanket manner, particularly when the quality of life and the potential for well-being are taken into account. As Singer (1981; 1998; 2003; 2009) suggests, the ability to experience pleasure or pain, rather than mere species membership, should determine moral consideration. However, this does not mean that the right to life should be generalized indiscriminately to all beings. I am not discussing a matter of granting the right to life to every human or every non-human animal; rather, it is about protecting beings from suffering, particularly when their quality of life is minimal or nonexistent.

in light of the resistance it may encounter. I will now give some examples where the distinction between ideal and real-world justice is evident[29].

The first example is factory farming. As I stated before, on an ideal justice level, factory farming would be abolished entirely. All sentient animals would have the right to live free from cruelty, exploitation, and unnecessary suffering. This would align with a universal application of the principle that all beings capable of experiencing pain should be treated with dignity and respect.

In the real world, however, the complete abolition of factory farming may not be feasible in the short term due to economic interests, entrenched practises and cultural preferences. A pragmatic approach to real-world justice could be to push for incremental reforms, such as improving living conditions for animals, introducing better regulations for the treatment of animals in the food industry, or supporting alternatives such as plant-based diets. Justice in the real world could also focus on raising awareness and bringing about societal change towards a more humane treatment of animals, even if a complete abolition of factory farming is not immediately achievable. It is important to point out that no matter how unjust the current situation is for non-human animals, systemic change cannot happen overnight. However, we should be less ignorant when it comes to the current conditions in which many farm animals, including cows, pigs and chickens, are still kept in conditions that cause immeasurable suffering. While some countries have passed legislation to improve animal welfare, these measures are often limited and inhumane conditions persist on many farms. Stress is a major problem for farm animals, often caused by unfavourable environmental factors, and can lead to serious consequences ranging from discomfort to death. Research has shown that both acute and chronic stress affect the hormonal and behavioural responses of animals, with chronic stress —which is common in intensive farming — receiving less attention (Dantzer and Mormède, 1983). Furthermore, practises such as deliberately inducing anaemia in calves to produce paler, more expensive meat raise significant ethical concerns (Singer, 1998).

I argue that in light of current practises, several steps can be taken to improve it and promote justice for farm animals in the real world. For example, by seeking to understand the mechanisms of stress, including tolerance and sensitisation, we can support the development of more humane husbandry practises, improve veterinary care and reduce unnecessary suffering in the field. In this context, at the real-world level of justice, this means supporting husbandry practises that prioritise animal

[29]The author Federica Liveriero in *Relational Liberalism: Democratic Co-authorship in a Pluralistic World* also develops questions of adapting principles of justice to real-life through practical compromises (Vol. 24, Springer Nature, 2023). She uses the example of same sex marriage. These questions are also explored further by Jonathan Wolff in *Ethics and Public Policy: A Philosophical Inquiry* (New York: Routledge Press, 2011), whose strategies I will discuss in more detail in this section.

welfare, such as free-range or pasture-based husbandry, which allows non-human animals to express their natural behaviours and live in an enriching environment, improving their health and reducing stress-related problems.

Laws and regulations prescribing humane treatment are also crucial, such as the ban on gestation cages for pigs or the establishment of minimum space requirements for laying hens (Watnick: 2016). Research into alternative methods of food production, including plant-based proteins, cultured meat and insect-based foods, can reduce reliance on traditional livestock farming and minimise animal suffering while ensuring food safety (Lisboa et al. 2024). In addition, animal husbandry conditions can be improved through research into animal welfare, the development of better breeding methods and more humane tools. Consumers can make more ethical choices through awareness campaigns and labelling schemes such as "Certified Humane", which encourage companies to adopt better practises. Increased demand for ethical products can change the market and industry standards[30]. Community engagement through discussions and programmes such as Community Supported Agriculture (CSA) can lead to policy changes (Brown and Miller: 2008). Advocating for government regulations and international agreements also ensures long-term improvements and aligns animal welfare with productivity. The co-operation of all – individuals, communities, industry and governments – is essential for sustainable change in agriculture (Fernandes et al. 2023[31]).

The next example is animal experiments in medicine. From the point of view of ideal justice, no non-human animal should suffer for human health, just as no non-human animal should suffer for food production, as already mentioned. From the perspective of ideal reasonable agents (IRAs), any form of animal experimentation would be prohibited, as it inevitably involves harm and suffering. If we adhere to the principles of universal compassion and justice, the use of non-human animals in medical testing would be deemed morally unacceptable.

However, in the real world, the issue is more complex. The medical and pharmaceutical industries, particularly those involved in vaccine development and life-saving treatments, provide evidence that animal testing is sometimes necessary to ensure the safety of new treatments before they are used on humans. This creates a conflict between advancing human health and minimising animal suffering.

In the context of real-world justice, a more nuanced approach is required. As Wolff (2011) acknowledges, scientific experimentation on animals often involves a cost-benefit analysis where the potential benefits to human health may justify some level of harm to non-human animals. However, this justification depends on the ethical

[30] https://ascot-meats.com/ethical-meat-consumption-a-guide-to-conscious-eating-habits/ Accessed on 07.03.2025.

[31] https://www.mdpi.com/2077-0472/9/6/132 Accessed on 07.03.2025.

value placed on animal lives versus human benefits. Wolff argues that while experiments causing harm to animals may be justified if they lead to significant medical advancements, we must also be cautious of speciesism—the assumption that human lives are inherently more valuable than non-human animal lives simply due to species membership. Real-world justice, therefore, demands strategies that minimise harm to animals while still allowing for potential medical benefits. Wolff's classification of experiments—mild, moderate, severe, and unclassified—provides a framework for assessing the degree of suffering inflicted on animals. This can inform regulations aimed at limiting harm. For example, mild experiments, which cause minimal discomfort or pain, may be more justifiable than severe experiments, where animals experience significant suffering.

A pragmatic approach to justice would advocate for the 3Rs principle—Replacement, Reduction, and Refinement—as a guiding framework for medical research.

- Replacement involves using alternatives such as in vitro models or computer simulations.
- Reduction seeks to minimise the number of animals used in research.
- Refinement focuses on improving experimental techniques to reduce suffering, such as employing humane endpoints and minimising discomfort (Clark: 2018).

Public policy and ethics committees are also crucial to ensure that animals are treated fairly in research. This means that ethics committees in the real world should carefully consider whether the harm caused to animals is justified by the potential benefits of the experiment. Animal welfare laws lay down rules for humane treatment in research. It is also important that researchers are transparent – they should publicise their methods, share their findings and report any adverse events to ensure accountability (Ormandy et al. 2019[32]).

The development of alternatives to animal testing is leading to significant progress in science and medicine. Modern methods such as computer modelling, in vitro testing and "organs on chips[33]" offer more precise, ethical and efficient approaches to disease research and drug development. These innovations in real-world could reduce reliance on animal testing and improve the relevance of results to human health. Computational modelling and simulation enable high-tech programmes to predict the body's responses to chemicals and replace traditional testing with in silico methods that rapidly analyse the toxicity and pharmacokinetics of drugs. In vitro tests performed in the laboratory on human cells provide more accurate predictions of the effects of drugs and have already begun to replace animal testing in evaluating the safety of vaccines. The "Organs on Chips" developed at Harvard use microfluidic

[32] Available at: https://www.mdpi.com/2076-2615/9/9/622 Accessed on 07.03.2025.

[33] https://wyss.harvard.edu/technology/human-organs-on-chips/ Accessed on 07.03.2025.

assemblies with human cells to mimic the functions of organs such as the lungs, intestines and kidneys, enabling drug and chemical testing under conditions that closely mimic human biology. However, despite these advances, animal testing remains widespread due to various systemic factors. Pharmaceutical and chemical companies often rely on animal testing to protect themselves legally in the event of adverse reactions in humans. Financial incentives and long-standing scientific traditions also contribute to the persistence of animal testing, as researchers working with animal models receive more recognition and funding, while alternative methods are slower to be accepted. In education, many institutions continue to use animals for training, even though there are high-quality alternatives such as interactive simulations, videos and virtual reality that avoid unnecessary animal suffering (Martinić: 2020)[34]. As mentioned earlier, we need to realise that this will not change overnight, but we should be less ignorant about it. As Wolff (2011) suggests, achieving justice in the real world requires both short-term improvements and long-term commitments to systemic change. In this sense, true moral progress cannot rely solely on individuals seeking alternatives or avoiding ethical dilemmas. As Rosalind Hursthouse (2011) argues, institutional efforts through policy reform, research funding and public awareness are required to drive meaningful and lasting social change.

The aesthetic use of non-human animals, such as in cosmetic testing or the use of fur in fashion, presents another example of the tension between ideal and real-world justice. Ideal justice would demand that no non-human animal is harmed or exploited for the sake of human beauty or fashion needs. However, in practice, the beauty and fashion industries continue to rely on animal testing and animal-derived materials, often placing the pursuit of human desires above the ethical treatment of animals. Besides, there are relevant economic interests related to the fashion industry that we need to take into consideration (these are not only the corporate interests of capitalists and the privileged, but also, for example, the jobs of members of the working class). Real-world justice requires acknowledging this complexity and finding solutions that can reduce harm while balancing the needs and desires of human industries with the welfare of animals.

The use of animal-based materials in the fashion industry, such as leather, fur, wool and feathers, has a long tradition and symbolises luxury and status. However, with increasing awareness of animal rights and sustainability, criticism has been levelled at the inhumane conditions on farms and the serious environmental problems associated with the production of these materials. As in the case of farm animals, fur

[34] Martinić (2020), following: Cheluvappa, R., Scowen, P., & Eri, R. (2017). "Ethics of animal research in human disease remediation, its institutional teaching; and alternatives to animal experimentation." *Pharmacology Research & Perspectives, 5(4),* e00332, p. 10; Anderegg, C. et al. (2012). *A Critical Review of Animal Experiments.* In: Dvostruka Duga, Čakovec and Prijatelji životinja, pp. 18–20; and Singer, P. (1998). *Animal Liberation.* Zagreb: IBIS Grafika, p. 73.

and leather often come from industries where animals are forced to live and suffer in poor conditions. Leather is often seen as a by-product of the meat industry, but many farms breed animals specifically for leather production, leading to more ethical dilemmas. Leather production also has a significant impact on the environment due to the use of tanning agents, chemicals and pollution. While alternative materials such as faux fur and vegan leather are considered more environmentally friendly, they also have their own ecological footprint as they are based on petrochemical resources and require intensive production. However, new solutions such as Piñatex (pineapple leather) and Myla (mushroom leather) offer sustainable alternatives that do not exploit animals (Plannthin 2016: 69-117.)

Brands such as Stella McCartney[35] have already shown that haute couture can thrive without animals, fuelling the demand for ethical fashion. Consumers are increasingly demanding transparency and details about production conditions, prompting brands to be accountable and provide ethical information.

Wolff's (2011) argument that moral considerations should focus on the morally relevant characteristics of animals, such as their capacity for suffering, supports this shift, as it prioritises the avoidance of unnecessary pain or harm over the fulfilment of aesthetic desires. Again, change cannot happen overnight but, as Plannthin (2016) writes, in the future:

> *In the future producers, manufactures, fashion brands, and designers will undoubtedly encounter problems of consequence and reliability in their decisions when working with animals whether it involves fur, skin, wool, or feathers. The question then will be this: Is it more fruitful to take action regarding these issues by becoming more knowledgeable and insightful or by taking traditional defensive role. Knowing and showing will be the way to transparency. To demonstrate and explain choices toward an ethical approach and to be honest and transparent: These are the next steps. Labeling will must be improved in order to tell the true and full story of what kind of animals are used, where were they bred, and what kind of slaughter methods were used. The potential for collaboration across disciplinary boundaries should not be ignored in this industry. Ethical thinkers, NGO, farmers, manufacturers, and designers should all collaborate together for future production and possibilities* (Plannthin 2016:117).

The next example is environmental destruction and habitat loss. In an ideal world, humanity would act swiftly to prevent environmental degradation and protect the habitats of non-human animals from destruction. This would include rigorous policies

[35] https://www.stellamccartney.com/gb/en/sustainability/fur-free fur.html?srsltid=AfmBOoqtc9UYULQjegQbhi_Er2BgoR4qJhjZn35X3Ag_Et1yN9jyuIvO Accessed on 07.03.2025.

to mitigate climate change, prevent deforestation, and reduce pollution — actions that ensure the survival of diverse species in their natural environments.

In the real world, addressing environmental destruction often involves balancing economic development, political interests, and immediate human needs with long-term ecological goals. Real-world justice in this case might focus on finding practical solutions, such as implementing sustainable practices in agriculture, investing in green technologies, and enacting policies that promote biodiversity protection while accounting for the realities of industries that depend on land and natural resources. While total preservation of ecosystems might not always be feasible, a focus on creating more sustainable and ethical practices would still be a step toward realizing some of the ideal justice goals.

The last example I will present to show how the distinction between ideal world justice and real-world justice works is pet ownership and breeding. From the ideal level of justice, pet ownership should involve a deep respect for animal rights, including ensuring that pets are kept in environments that meet their physical and psychological needs. The ideal world would prohibit the breeding of animals for profit, as this can often lead to overpopulation, abandonment, and exploitation. Instead, animal companions would be adopted from shelters or sanctuaries, where their welfare is prioritized.

Real-world justice recognizes that pet ownership is widespread, and changing people's attitudes and behaviors around pet breeding and adoption will take time. A more practical approach could involve tightening regulations on breeding practices, encouraging adoption over buying, and implementing education campaigns about the responsibilities of pet ownership. Though ideal justice may call for a complete overhaul of pet ownership practices, real-world justice can work toward minimizing harm by fostering more ethical, compassionate pet care practices. For example, practices causing unnecessary suffering, such as constant chaining or using shock collars, are incompatible with humane treatment principles (Martinić, 2021). Proper care should consider the specific needs of each species. Dogs and cats require regular exercise and mental stimulation, while birds thrive in environments that allow natural behaviours, such as aviaries instead of small cages (Korsgaard, 2018). Addressing broader issues, such as pet overpopulation, is critical. Strategies to prevent overpopulation and ensure animal welfare in the real world should include promoting sterilisation to prevent unwanted litters and ensuring the placement of pets in suitable homes as key. Policies that enforce anti-cruelty laws and minimum standards of care must be consistently enforced, with mechanisms in place to address neglect or abuse. Education plays a key role in encouraging responsible pet ownership. For example, training pets, especially dogs, helps them adapt to human-centred environments and prevents risks to themselves and others. Training should foster mutual understanding rather than control (Bok, 2011). Prospective and current pet owners must understand the responsibilities involved in pet ownership. Pets should not be kept solely for

entertainment or image but must be cared for properly. Those unable to meet these responsibilities should seek alternatives to ensure the animals' welfare (Martinić, 2021). In addition, supporting animal shelters and rescue organizations with funding, resources, and volunteers enables them to provide medical care, rehabilitation, and adoption services. Making veterinary care available and affordable through subsidized or low-cost programs is also essential, helping pet owners secure necessary treatments. Fournier and Geller (2004) argue that the main issue is that current environmental factors often encourage negative behaviors and prevent change. To address this, we should use behavior analysis to adjust these factors and promote positive behaviors. They suggest that animal welfare agencies, traditional research groups not involved in animal welfare (e.g., behavior analysts, community psychologists, and epidemiologists), and officials who research other community problems (e.g., county health departments, city councils, and urban planners) should work together to create solutions (Fournier and Geller 64: 2004).

The examples presented above illustrate that ideal justice offers a vision of a world where all living beings, human and non-human, are treated fairly and with respect. However, real-world justice acknowledges the complexities, constraints, and resistance in the world as it currently stands. Still, developing a conception of ideal justice is strongly relevant. Namely, it defines the limits of legitimate behaviours, the goals to be achieved, the compromises that we need to accept in the long-term path to the complete achievement of justice. While the application of ideal principles may not be immediately achievable, real-world justice involves crafting pragmatic strategies that respect these ideals while also considering the socio-political and economic realities at hand. The goal is to make incremental progress, always working toward the ideals, but within a framework that accounts for what is possible in the present moment.

CONCLUSION OF PART ONE

In part one of this dissertation, I have critically examined Nussbaum's alternative (2006) to Rawls's framework of justice (1971; 1999; 2001; 2005), focussing on her critique of the social contract tradition and the development of her capabilities approach. This analysis was set against the backdrop of Rawlsian principles, whose fundamental emphasis on rationality and mutual advantage reveals significant gaps in addressing the needs of individuals who are unable to be rational and reasonable. Rawls' theory assumes that all individuals have the capacity to engage in rational decision-making and are capable of participating in the social contract. However, this overlooks the needs of those who, due to cognitive or developmental impairments, are unable to meet these criteria. The danger here is that Rawls' framework does not fully include individuals with severe cognitive disabilities in its considerations of justice. The need for a more inclusive approach to justice is crucial to ensure that all individuals, regardless of their cognitive capacities, are recognised and supported within the framework of justice.

Nussbaum's capabilities approach provides a compelling alternative, grounding justice in a set of fundamental capabilities necessary for a life of dignity. However, as I have shown, Nussbaum's framework faces challenges, particularly its reliance on a fixed list of essential capabilities and its emphasis on grounding rights on species membership. This fails to account for the diversity of human and non-human experiences, especially regarding disability and cultural differences. The universalism and species-based norms within the approach can exclude those who do not fit a single vision of flourishing. Despite its flexibility, Nussbaum's approach risks imposing a restrictive view of well-being and does not fully address pluralistic values. I have argued that to overcome these limitations, a more flexible and inclusive model of justice is needed—one that respects diversity and accommodates different perspectives on well-being and dignity. Further refinement of the capabilities framework is required to better meet the complex needs of all individuals in society.

I have explored various approaches to address the limitations of Nussbaum's capabilities framework in relation to justice and the inclusion of individuals with severe cognitive disabilities. The theories of Badano, Richardson, Stark and Freeman offer valuable insights into how justice can be extended to better include those who cannot fully participate in conventional frameworks of reasoning and deliberation. Building on these ideas, I proposed a new model of public justification centred on ideal reasonable agents (IRAs).

IRA's model of justice builds on the principle that duties and rights are established through the reasoning of individuals capable of impartiality and universalisation. By acting as impartial legislators, IRAs ensure that justice principles are justified not only for themselves but also for individuals who cannot directly engage in the process, such as those with cognitive disabilities. This approach aims to transcend the traditional "membership ticket" models of justice, offering a more inclusive and adaptable framework that respects the dignity and rights of all individuals. Through critical analysis of existing theories and the proposal of a revised model, I have argued for a pluralistic and context-sensitive approach that better accommodates the diverse realities of human experience. This revised framework, centred on the concept of ideal reasonable agents, seeks to uphold the principles of inclusivity and fairness, addressing the complexities of justice in a way that respects the dignity of all individuals, regardless of their capacities.

In this chapter, I have also explored the extension of the principles of justice to non-human animals, building on the IRA's framework established in previous discussions of the inclusion of individuals with cognitive disabilities. In this context, I have examined important contributions from scholars such as Kittay, who has extended the concept of moral personhood to those with cognitive disabilities, and Rawlsian theorists, who have explored the possibility of extending justice to non-human animals. Through this lens, the chapter has argued that while ideal justice demands the equal treatment of all living beings, real-world justice requires pragmatic

strategies to move progressively towards these ideals, recognising the complexities of human and non-human animal relationships and the limitations of existing systems.

Ultimately, in this first part of the dissertation, I hope to have provided a comprehensive analysis of the philosophical foundations and practical challenges of extending the Rawlsian scope of justice to non-reasonable and non-rational beings. In particular, I hope to have offered a more inclusive and practicable approach that incorporates both individuals with severe cognitive disabilities and animal welfare, grounded in both the principles of justice and the realities of the modern world.

3. PART TWO: OBJECTIVITY OF EVALUATIVE STANDARDS IN PSYCHIATRIC CLASSIFICATION OF MENTAL DISORDERS

3.1. Chapter Three: Section One: Introduction to the problem

The first part of the dissertation dealt with the justification for the inclusion of individuals with severe disabilities in the scope of justice and thus laid the foundation for a critical examination of the second challenge of this dissertation — the critique of psychiatry.[36] In particular, my inspiration for dealing with this issue concerns the question of the protection of freedom and autonomy in psychiatry. This issue has long been the subject of debate, initially brought to the fore by the anti-psychiatry movement and other thinkers such as Thomas Szasz (1960; 1994; 2000) and Michel Foucault (1989). Building on these discussions, this chapter examines the extent to which psychiatry can achieve various goals that could be in tension: (i) establishing objective evaluative standards[37] that (ii) protect individual autonomy while recognising freedom and equality. This is particularly important given the historical context in which psychiatric patients have often been subjected to inappropriate and dehumanising treatment.

Two influential figures in anti-psychiatry discourse are Thomas Szasz (1960; 1994; 2000) and Michel Foucault (1989). Szasz, a prominent critic of conventional psychiatric practises, emphasises the tension between psychiatric classifications and the principles of individual autonomy and equality. His work challenges us to question whether psychiatry can overcome its historical stigmatisation and truly recognise the value of individuals with mental health problems. Similarly, Foucault's seminal contributions illuminate the complex power dynamics inherent in the field of psychiatry and mental health more broadly. His research challenges us to examine how psychiatric practises have historically functioned as mechanisms of social control and normalisation, often overshadowing the imperative to treat the individual as a free and equal moral agent.

Thus, the main aim of this chapter is to answer the criticism by searching for objective evaluative standards. This will be achieved by developing a specific method inspired

[36] In the first part of the dissertation, the primary focus was on individuals with *severe* cognitive disabilities. In the second part, disabilities will be examined in a broader context, which will include not only severe cognitive disabilities but various forms of mental disorders.

[37] As I mentioned in the introduction to this work, by evaluative standards I mean criteria that are used to evaluate and judge something. In the context of psychiatry, these standards are used to assess mental health and determine whether someone has a mental disorder. They help professionals decide whether certain symptoms or behaviours meet the criteria for a particular diagnosis. These standards aim to provide a uniform method for the diagnosis and treatment of mental disorders while recognising individual differences and needs.

by Gaus' (2011) concept of weak external epistemology and a specific form of public justification based on convergence. In this way, this chapter attempts to develop a framework that ensures objectivity in psychiatric evaluations while respecting individual autonomy and pluralism.

To this end, I will first present Szasz's challenge to psychiatric objectivity and then turn to Foucault's critique to examine its contributions and implications for the development of fair and objective standards in the classification of mental disorders.

A. Thomas Szasz: Values, objectivity and the dynamics of power in the diagnosis of mental disorders

An important aspect of examining the criteria for categorising objective and fair evaluative standards and the intersection of this categorisation with the principles of justice, both within psychiatry and beyond, is to examine Szasz's perspective on the role of values and moral norms in defining mental disorders.[38] As mentioned, his perspective offers valuable insights into the ways in which societal values and moral judgements shape our understanding of mental health, which will be essential for a nuanced discussion of the intersection of disability and justice. Szasz thus argues that the categorisation of a condition as a mental disorder acts as a mechanism for enforcing value-laden decisions, enabling psychiatrists and the wider psychiatric establishment to exercise a form of oppressive power over individuals. In his view, this transforms psychiatry from a science-based discipline into an instrument of social control, where the pathologisation of deviance, dissent or non-conformity legitimises coercive measures under the guise of medical necessity.

The central argument of Szasz (1960; 1994; 2000) questions the legitimacy of the concept of "mental disorder." According to Szasz, the concept of mental disorder is fundamentally flawed and should be rejected. He argues that unlike physical disorders, which are defined based on observable and measurable abnormalities in the body, mental disorders are determined by value judgements and are therefore inherently subjective. Szasz argues that the term "disorder" should be reserved exclusively for medical use, where it refers to physical abnormalities or somatic pathologies—, i.e. undesirable changes in the structure or function of the body (Szasz: 1994). From this perspective, Szasz strongly asserts that even when brain lesions are observed in individuals with mental health conditions, these findings should not automatically be classified as evidence of a mental disorder. Instead, he believes such conditions should be recognised as brain disorders, which he considers diseases of the central nervous system rather than conditions of the mind (Szasz 1994: 35-39).

[38] This discussion does not delve into the specific terminology of the philosophy of psychiatry. Instead, the focus is on establishing a method for defining objective evaluative standards. Therefore, in this context, the terms "disability," "illness," and "disorder" are used interchangeably to mean the same thing.

This distinction between brain disorders and mental disorders, Szasz argues, is critical for maintaining scientific objectivity. He asserts that objectivity can only be achieved by relying on naturalistic categories grounded in observable, biological phenomena. However, Szasz contends that this objectivity is undermined in the realm of psychiatric discourse. He highlights how, historically, psychiatric diagnoses in the nineteenth century primarily focused on identifying physical lesions or abnormalities within the body. During this period, psychiatry maintained closer alignment with the natural sciences. However, in the twentieth century, this focus shifted significantly. Diagnoses in psychiatry increasingly began to serve purposes beyond medical classification, including justifying treatment modalities, obtaining government funding, and fulfilling institutional objectives (Szasz: 1994). Szasz (1994: 37) argues that this shift has led to psychiatric diagnoses becoming entangled with economic, personal, political, and social considerations. He claims that these external influences compromise the scientific validity of mental disorder classifications. Unlike somatic disorders, which Szasz argues retain objectivity due to their basis in natural scientific principles and empirical evidence, mental disorders are highly susceptible to subjective interpretation. This lack of objectivity, according to Szasz, is due to the central role of value judgements in defining mental disorders. These judgements often reflect societal norms, cultural expectations, and professional biases, rather than objective scientific criteria.

Consequently, Szasz suggests that the classification of a condition as a mental disorder becomes a mechanism for imposing value-laden decisions. This process, he argues, enables psychiatrists and the broader mental health establishment to exert a form of oppressive power over individuals diagnosed with mental disorders. Szasz views this as particularly problematic, as it shifts psychiatry away from its potential grounding in scientific principles and turns it into an instrument of social control. In this sense, he contends that the authority granted to psychiatrists to diagnose and treat mental disorders has profound ethical implications. By categorising behaviours or experiences as disorders, psychiatry may inadvertently or intentionally pathologise deviance, dissent, or nonconformity, thereby legitimising coercive interventions in the lives of individuals who may not meet strict criteria for disease in a naturalistic sense.

Szasz's critique (1960; 1961; 1994; 2000) is rooted in his broader philosophical stance on the distinction between disease and deviance. He argues that whereas somatic diseases can be objectively identified through biological markers and physiological evidence, mental disorders are defined by subjective interpretations of behaviour, thought, and emotion. For instance, behaviours deemed undesirable or socially unacceptable may be labelled as symptoms of a mental disorder, even when no underlying biological pathology is evident. In this view, psychiatry operates in a fundamentally different manner from other medical disciplines, as it is often tasked

with addressing moral, social, and existential dilemmas rather than physiological abnormalities.

To further illustrate his argument, Szasz critiques the evolution of psychiatric practice and its implications for society. He observes that, over time, psychiatry has expanded its scope to include a wide range of behaviours and experiences, many of which may not correspond to identifiable neurological or biological conditions. For example, conditions such as depression, anxiety, or schizophrenia are frequently diagnosed based on clinical interviews and subjective reports, rather than objective tests or biomarkers. While these conditions undoubtedly cause significant distress and impairment for individuals, Szasz argues that their classification as disorders is influenced by cultural and historical contexts. As such, he warns against conflating personal or societal discomfort with medical pathology.

In summary, Szasz's critique (1960; 1961; 1994; 2000) of the concept of mental disorder challenges the foundation of contemporary psychiatric practice. He calls for a re-evaluation of the criteria used to define and diagnose mental disorders, urging greater caution in distinguishing between genuine medical conditions and socially constructed categories. By highlighting the influence of value judgements, economic incentives, and social norms on psychiatric diagnoses, Szasz seeks to provoke a critical examination of the ethical and scientific underpinnings of mental health care. His work continues to generate debate within the fields of psychiatry, psychology, and philosophy, encouraging ongoing reflection on the nature of mental illness and the role of psychiatry in modern society.

In this context, it is important to examine the implications for patient autonomy and informed consent. Here it may be useful to refer to "The Duty to be Well Informed: The Case of Depression" by Charlotte Blease (2014), a thought-provoking theory and an appropriate exploration of the ethical duty of physicians to provide patients with accurate and complete information about their conditions. Blease highlights the complexity of patient education and informed consent and emphasises that physicians should stay updated on the evolving understanding of mental illness. She points out the gap between the medical community's nuanced understanding of depression and the public's simplistic perceptions. She emphasises how pharmaceutical marketing and rising antidepressant prescription rates are shaping patients' perceptions of depression, often leading to misconceptions. Blease (2014) argues that the term "antidepressant" can give the false impression that depression can be treated with a single "miracle cure", like antibiotics for infections. She criticises the view that depression is solely a "biological disease" caused by a "chemical imbalance" in the brain. This simplistic view can have a negative impact on patients' understanding of their illness, their treatment decisions and their long-term prognosis. She uses empirical evidence to show that this view oversimplifies the complex interplay of biological, psychological and social factors that contribute to depression. Blease argues in favour of a more comprehensive understanding of depression that includes

a range of treatment options such as psychological therapies and environmental changes in addition to medication. She emphasises the ethical responsibility of healthcare providers to provide patients with accurate and holistic information about their illness. Furthermore, Blease (2014) argues that informed consent and patient education are crucial for promoting autonomy and well-being. She raises ethical concerns about inadequate patient education, stating that withholding or misrepresenting information about depression compromises patient autonomy and undermines trust in the medical profession. Blease examines factors such as misunderstanding, expediency and patient pressures that lead physicians to provide inadequate information. She emphasises the need for greater awareness and accountability within the medical community.

Blease's criticism of oversimplified explanations for depression aligns with Szasz's arguments about the subjective nature of psychiatric diagnoses. Both Blease (2014) and Szasz (1994, 2000) emphasize how social, economic, and political factors influence the understanding and classification of mental illness. Blease specifically targets the "chemical imbalance" theory of depression that relies on a reduction of mental disorders to mere natural phenomena, neglecting wider components, such as those social and political. Szasz is partly associated with these critiques as he says that psychiatry is fully immersed in social and political factors. They are, thus, associated by the thesis that affirms the primary relevance of these factors in psychiatry. Szasz argues that psychiatric diagnoses often lack empirical verification compared to somatic medicine and are susceptible to non-medical influences. Both Blease and Szasz address power dynamics in psychiatry. Szasz criticizes the authoritarian power that psychiatrists can exert over patients based on subjective diagnoses, which can be oppressive when influenced by external factors. Blease echoes this concern by highlighting how oversimplified biological explanations limit patient understanding and choice, perpetuating a paternalistic approach to mental health care. Ultimately, both Blease and Szasz argue against overly simplistic or reductionist explanations that obscure the complexity of the human experience and perpetuate harmful power dynamics within the healthcare system. However, despite these similarities, a key difference exists between their perspectives. Szasz advocates for a strict adherence to naturalistic facts, arguing that psychiatry should avoid subjective or socially constructed classifications. In other words, he affirms that psychiatry, to be successful, needs to realize coherently the reduction of its classifications and diagnoses on natural facts, while the critique is that it cannot be successful. In contrast, Blease calls for greater awareness of the broader context, emphasizing the need to account for the complex social and political dimensions of mental health. Thus, her project is, in a sense, opposite: realize the awareness of the flaws of reductionism and the need to properly deal with social and political components.

The power dynamics in psychiatry that Szasz and Blease criticise are not merely theoretical — they have real, historical consequences. As mentioned above, Szasz argues that psychiatry functions as an instrument of social control, while Blease emphasises the importance of social and political factors in mental health. These debates are not merely abstract. Psychiatry's use of power has always sparked controversy, from the use of asylums to modern debates about the medicalisation of daily life. A clear example is the earlier categorisation of homosexuality as a mental disorder (Szasz 1994: 36). The Diagnostic and Statistical Manual of Mental Disorders (DSM) of the American Psychiatric Association (APA) categorised homosexuality as a mental disorder until 1973. This categorisation was based on current social conventions and prejudices and not on scientific facts. As knowledgeable mental health professionals, psychiatrists contributed significantly to the pathologisation of homosexuality. This diagnosis had serious consequences, as it was used to justify a variety of discriminatory practises such as forced institutionalization and so-called "conversion therapies", which often caused great harm to those affected.

The term "drapetomania" is another cruel historical illustration. The desire of enslaved African Americans to escape slavery was labelled a mental disorder by some American psychiatrists in the 19th century. This diagnosis served to delegitimise the slaves' legitimate desires for freedom and human rights, as it was based on the beliefs of white supremacy and the desire to preserve the institution of slavery (Rajapakse 2024).

In various authoritarian regimes, including the Soviet Union and some totalitarian states, political dissent was pathologised as a mental disorder. Psychiatrists in these contexts were complicit in diagnosing dissidents with "sluggish schizophrenia" (Wilkinson 1986) or similar conditions, thus silencing political opposition by labelling it as a form of mental disorder.

While modern psychiatry has made progress in recognising gender dysphoria as a legitimate condition that deserves medical and psychological support, there have been cases in the past where transgender people were pathologised and subjected to coercive and harmful treatments, often reflecting societal prejudices against gender diversity In general, and usually, this means the imposition of gender identity, which is assumed as corresponding to the given neurophysiological characteristics of individuals. The murder of Brandon Teena is one of the best-documented cases of anti-transgender violence and serves as a stark example of the dangers faced by transgender people. Brandon was a young transgender man living in Falls City, Nebraska, who had moved there in search of a fresh start. He befriended a group of people, including Lana Tisdel, with whom he became romantically involved. However, when two of his acquaintances, John Lotter and Tom Nissen, found out that Brandon was transgender, they reacted with hostility and violence. In December 1993, after learning of his transgender identity, Lotter and Nissen lured Brandon to a secluded location where they sexually abused and brutally beat him. Fearing for his

life, Brandon reported the assault to the local police. However, instead of taking immediate action to protect him, he was treated with hostility and indifference by the police. The local sheriff even subjected him to invasive and humiliating questioning about his gender identity instead of focusing on the crime committed against him. Days later, on 31 December 1993, Lotter and Nissen sought out Brandon at a friend's house and murdered him along with two other people present — Lisa Lambert and Phillip DeVine. Brandon was shot and stabbed to death in what was clearly an act of transphobic violence. Brandon's murder sparked nationwide outrage and became a landmark case in the discussion of hate crimes against transgender people. His story was widely publicised through media coverage and later inspired the 1999 film Boys Don't Cry, which further raised awareness of the challenges faced by transgender people. The murder of Brandon Teena is a painful reminder of the discrimination, violence and lack of legal protection that transgender people have historically faced. It also emphasises the importance of continuing to fight for transgender rights, legal protection and social acceptance[39].

However, there are rare opposite cases on the other side. A notable example is the case of the treatment of David Reimer, originally born Bruce Reimer, by Dr John Money in the 1960s and 70s. Following a damaged circumcision, Dr Money recommended that Bruce be raised as a girl, Brenda, as part of an experiment to explore the influence of gender socialisation. Money's approach was based on the belief that gender identity can be shaped entirely through upbringing and is not an innate aspect of a person's identity. Despite Money's initial success, Reimer struggled with his assigned gender and was informed of his past at the age of 14. He decided to live as a man again and took the name David. Although he underwent corrective surgeries and hormone treatments, Reimer faced significant emotional and psychological trauma throughout his life. In 1997, he shared his story, exposing the unethical practices he suffered, including forced "sexual rehearsal play" in therapy. His experience highlights the harm caused by treating gender diversity as a problem and the importance of respecting a person's self-identified gender (Slayton et all)[40]. A key issue in cases like David Reimer's is the imposition of a gender identity that does not align with the individual's own sense of self. Typically, such impositions come from those who insist on strict adherence to morphological and biological identity, rejecting gender diversity. However, Reimer's case demonstrates that the opposite can also occur—where an assigned gender identity is imposed despite the person's biological characteristics.

---

[39] https://lambdalegal.org/case/brandon-v-richardson-county/ and https://en.wikipedia.org/wiki/Brandon_Teena Accesed on 3.3.2025.

[40] Slayton, Kelly, Alexander Grigorievskiy, and Live Statistics. "DAVID REIMER AND THE GENDER EXPERIMENT." https://wiki2.org/en/David_Reimer Accesed on 15.08.2024.

The fundamental problem in both scenarios is the failure to respect an individual's own understanding of their gender. Reimer's experience underscores the harm caused when external authorities, whether medical, social, or ideological, override a person's autonomy in defining their identity. This highlights the broader issue of psychiatric and medical interventions that, rather than supporting the individual's freedom, enforce socially constructed norms, often leading to lifelong psychological distress.

These examples show how psychiatric diagnoses reflected social and cultural norms, as Szasz described, and sometimes led to the stigma and oppression of people based on their sexuality, ethnicity, politics, or gender identity. Psychiatrists, as leaders and decision-makers in their field, have considerable authority when it comes to asserting their statements as scientific fact and recommending rational treatments (Szasz 1994: 37). Their categorisations can even lead to legal decisions. Consequently, according to Szasz, they act more as legislators than as scientists. For this reason, he argues that psychiatry differs from traditional medicine because of this different social role and authority.

While the conventional role of medicine is to cure the sick, psychiatry often finds itself in the role of confining and controlling "deviants" for treatment or, in other cases, acting as an arbiter of legal and political decisions. Szasz claims that this positioning of psychiatry transforms it into a branch of oppressive politics that threatens the respect due to individuals as moral agents (Szasz 1994: 39). The danger is that the categorisation of disorders is used as an additional means for the powerful to exert control over the vulnerable or, more generally, to impose values held by a few on society as a whole. In light of all this, Szasz believes that psychiatry has become a moral foundation in contemporary Western society. Its institutions and interventions legitimise hierarchical power dynamics and exert a considerable influence on our daily lives (Szasz 2000: 15).

In conclusion, Szasz's critique of psychiatry challenges the foundational principles of mental disorder classification, asserting that these are shaped more by subjective value judgments and societal norms than by objective scientific criteria. His arguments reveal the ethical and political complexities inherent in psychiatry's role, particularly its capacity to pathologize deviance and exert social control. This critique is further supported by historical examples where psychiatric diagnoses have been misused to justify discrimination and oppression, emphasizing the need for caution and critical reflection in mental health care. The parallels between Szasz's and Blease's perspectives underscore the enduring tension between science, ethics, and the sociopolitical dimensions of mental health practice.

In the next section, I will examine Foucault's critique of psychiatry, which offers an incisive analysis of how power and knowledge intersect to shape psychiatric practices and the construction of mental disorder.

B. Michael Foucault: The ideological role of mental disorder and population regulation

Foucault holds similar views as Szasz (1994; 2000) in his work "Madness and Civilisation" (1989), in which he examines the evolving definition and social connotations of "madness" in relation to the authority of psychiatrists. Foucault's insights run parallel to Szasz's concerns and illuminate how the concept of mental illness has evolved with the growing influence of psychiatric authority.

When analysing the origin of the asylum, Foucault says:

> *As positivism imposes itself on medicine and psychiatry, this practice becomes more and more obscure, the psychiatrist's power becomes more and more miraculous, and the doctor-patient pair sinks deeper and deeper into an unusual world. In the eyes of the patient, the doctor becomes a thaumaturge; the authority he had borrowed from order, morality, and family now seemed to come from himself; he is believed to possess these powers because he is a doctor* (Foucault 1989: 261).

For Foucault (1989), mental disorder, if we look at it through the concept of illness, had above all an ideological role due to the extension of political control over the population. Foucault's concept of power is closely linked to knowledge. Namely, he argued that power functions through the dissemination and control of knowledge. Rationality in this context is not just a question of reason but is deeply interwoven with power relations and the knowledge systems on which they are based. He emphasised the historical and context-dependent nature of rationality. In other words, rationality is not a universal and ahistorical concept, but depends, according to Foucault, on specific historical and cultural contexts (Fraser 1981). Different historical periods produce different forms of rationality. In the era of classicism, madness stood in opposition to reason and has since been reduced to a social disease. The twist that Foucault adds is that others' ideas about what it means to be "bad" are largely the product of regimes of power that hollow out historically arbitrary standards of judgement and use them to present themselves as timeless, universal moral standards and essential truths. So, if we visualise and question the reasons why these problems are seen as "real" or "the way things have to be", it becomes easier to address or resist the problems people face when dealing with this confusing way of understanding their situation (Lock et al. 2005: 5 of 23).

The change in the way people viewed mental illness had less to do with compassion for those affected and more to do with social and economic demands. Foucault (1989) noted that when Europe first studied mental illness through medicine, it began to classify and define it based on notions of "rationality". In this early phase, known as psychiatry, the therapist took on a role similar to that of a priest. The person with a mental disorder was like a child who was considered irrational and confessed their problems, while the therapist, who was considered rational, held a position of

authority and was expected to manage the situation without the use of physical force. When Foucault talks about homosexuality, he says that it is not something regulated, but that he sees it as a product of culture. He bases his arguments on the fact that society, through scientific and medical institutions, creates norms and criteria to categorise conditions such as mental disorders. It defines certain conditions as normal or abnormal, and so they are labelled and stigmatised by the individual. He talks about the role of psychiatric institutions as well as prisons, which manifests itself in the control of minors who do not behave according to societal norms. By highlighting the importance of power in the process of pathologizing mental illness, Foucault argues that categories are not discovered independently in the world but are shaped and imposed by society. Foucault argues that sexuality is not repressed but is a central point of power and social control (1989).

He introduced the concept of "biopower", which refers to the regulation and control of populations by various institutions, including those related to sexuality. Rather than viewing sexuality as a private and individual matter, Foucault examined how it became an object of public discourse and regulation. Since it embodies what Foucault calls the "history of the present" — which is also always a concept of the future — biopower is therefore contemporary. Biopower reveals the processes, relations and practises that shape and utilise political subjects, as well as the forces that have shaped and continue to shape modernity. It is not contemporary, however, because its meaning must be obscured and concealed, for the forces of power always find a way to hide their traces (Cisney and Morar: 2020).

Jasper Friedrich's (2021) reflections on Foucault's theory of mental health offer a comprehensive analysis of the strengths and limitations of Foucault's approach while presenting an argument for a critical theory of mental health that does justice to the complexity of contemporary understandings of mental disorders. He begins by recognising the significant contributions of Foucault and other thinkers associated with the anti-psychiatry movement in questioning the authority of orthodox psychiatry and questioning established categories of health and illness. He highlights how these scholars, including Foucault, have critiqued the power dynamics inherent in psychiatric labelling and the medicalisation of suffering, which serve to depoliticise the experiences of people who do not conform to social norms of rationality and normality. However, Friedrich also points out the limitations of this approach, in particular the tendency to either romanticise madness or dismiss mental illness as a myth. He emphasises the importance of addressing the real experiences of suffering associated with mental disorders, rather than focusing solely on criticising psychiatric authority and the construction of mental health categories.

Friedrich (2021) then addresses the conceptualisation of mental health and distinguishes between the negative concept, which denotes the absence of mental disorders, and the positive concept, which emphasises psychological well-being and resilience. He argues for looking at mental health issues through the lens of the

positive concept and recognising mental disorders as a continuum of experiences ranging from mild suffering to clinically diagnosable illnesses. Furthermore, Friedrich (2021) acknowledges the critique of normalising tendencies within mental health discourse, particularly in relation to the intersections with biopower and the production of a healthy and productive population. Ultimately, however, he argues in favour of retaining the term "mental health" within the discourse of critical theory, as it resonates with the everyday experiences of people struggling with mental health problems.

To summarize, Foucault argued that mental disorder served an ideological function. The shift in the view of mental disorders is not due to compassion for those who suffer, but to social and economic imperatives. Against this background, he drew parallels between psychiatry and religious denomination. In this context, the analyst or therapist took on the role of a priest, and the person in psychiatric condition confessed their "sins." This dynamic positioned the therapist as an adult figure versus the supposedly childlike irrationality of the patient. His most famous case is that of homosexuality, which was pathologized because of the role of scientific and medical institutions in creating norms and stigmatising certain conditions as normal or abnormal. The control of minors who deviated from social norms, whether in psychiatric institutions or in prisons, illustrated the influence of power in this process. In this context, Foucault introduced the concept of "biopower" to describe the regulation and control of populations by institutions, including those related to sexuality. He explored how sexuality became an object of public discourse and regulation, challenging traditional notions of oppression.

C. Towards Objectivity and Pluralism in Psychiatric Classification

The previous sections highlighted the challenges and questions regarding the classification and recognition of mental disorders, aiming to underscore the oppressive nature of society rather than its freedom. This oppression is particularly evident in psychiatry, where the anti-psychiatry movement has raised significant concerns about power dynamics within psychiatric practice. As mentioned above, Szasz and Foucault argue in their critique that psychiatric classifications often reinforce social norms and power structures rather than addressing mental health issues in an unbiased way. Szasz’s assertion that mental disorders are inherently value-laden and influenced by external factors emphasises the criticism of psychiatry's tendency to present personal or cultural values as if they were objective medical facts. Similarly, Foucault’s analysis demonstrates how psychiatric practices historically evolved as tools of societal control, embedding broader power dynamics within their structure, rather than focusing solely on medical concerns.

This critique is particularly relevant in the context of Rawls' (2005) assertion that open and democratic societies should cultivate a rich tapestry of diverse beliefs, values and perspectives. Pluralism, a fundamental virtue of such societies, demonstrates their

commitment to a wide range of viewpoints and ideas. However, for pluralism to truly thrive, it must be protected from repressive tendencies that can undermine it. In areas such as psychiatry, the imposition of normative standards can lead to the marginalisation and discrimination of vulnerable groups, contradicting the principles of freedom and equality that underpin democratic societies. This concern becomes particularly clear when we consider the historical traces of psychiatry described above, as discussed by Foucault (1989) and Szasz (1994, 2000).

While I support the thesis that the definition of a mental disorder is inherently tainted with values, I firmly reject the notion that this value-laden nature necessarily implies a loss of objectivity that leads to oppression. Rather than arguing for abandoning the definition of a "mental disorder" or attempting to separate it entirely from values, I argue that the more appropriate response is to strive for some objectivity in the evaluative standards for categorising mental disorders. Achieving this kind of objectivity in these standards paves the way for a more objective definition of mental disorders that is open to pluralism and effectively prevents the arbitrary manifestations of oppression. The key to solving this challenge lies in the search for objectivity within the evaluative standards themselves. Once we gain an objective understanding of these standards, we create the potential for an equally objective determination of mental disorders. This approach serves as a safeguard against the forms of oppression that result from arbitrary or biassed assessments.

In the following section, I will first explore influential attempts to establish an objective understanding of evaluative standards within the Aristotelian framework, as articulated by Christopher Megone (1998, 2000) and Philippa Foot (2001). These approaches are significant because they represent early efforts to integrate values and objective norms, aiming to ground classifications on principled, non-subjective foundations. This examination is crucial for understanding whether such frameworks can avoid arbitrariness or oppression in the classification of mental disorders. However, I will critically evaluate and ultimately reject the Aristotelian approach, addressing key objections to its applicability in the context of contemporary psychiatric classifications. These objections raise serious concerns about the framework's ability to provide a truly objective and equitable definition of mental disorders. Following this, I will analyse George Graham’s (2013) Rawlsian-inspired approach, which presents a compelling attempt to define mental disorder within a framework of justice and fairness. Graham’s perspective offers valuable insights into the objective dimensions of disorder classification, highlighting the importance of fairness and impartiality in these processes. Finally, I will build on Graham’s definition of mental disorder by incorporating a new response inspired by Gerald Gaus (2011). This approach will centre on a specific form of public justification and epistemology, aiming to refine the conceptual tools available for classifying mental disorders.

The goal of this chapter is to demonstrate that, by striving for objectivity in evaluative standards and drawing from influential philosophical frameworks, we can enhance our understanding of mental disorders. This, in turn, can help lessen the risks of oppression and arbitrariness in their classification, paving the way for a more just and equitable approach to mental health in our society.

3.2. Section Two: Aristotelian replies

At the centre of the complicated debate about the classification of disorders, is a fundamental question: to what extent do values shape our understanding of what a disorder is? Christopher Megone (1998; 2000) has pursued an idea from Szasz, who claims that values play a central role in this determination. Megone disagrees with Szasz, however, as he does not see the inclusion of values as an obstacle to objectivity. Instead, he argues that a nuanced understanding of value standards can coexist with objectivity, drawing on Aristotelian ideas. Megone's theory, which emphasises rationality as central to human well-being, challenges Szasz's distinction between physical and mental illness. His theory centres on the fundamental importance of rationality for human flourishing.

However, as announced, the exploration of the Aristotelian answers does not end with Megone. I will also examine another Aristotelian perspective on the classification of disorders, presented by Philippa Foot (2001). Foot's theory of natural goodness offers a way of understanding human behaviour and moral virtue. Drawing on Aristotelian ethics combined with evolutionary theory, she attempts to provide a useful perspective for evaluating human behaviour.

Unlike Szasz, both Megone and Foot recognise that values matter, but they still strive to keep their evaluative criteria objective. From an Aristotelian perspective, both philosophers examine human well-being and the role of values and objectivity in understanding mental disorders. Megone questions the separation between physical and mental issues, while Foot's theory of natural goodness looks at human behaviour from a moral standpoint.

In the following subsections I will look more closely at these Aristotelian responses and emphasise the similarities between Megone's and Foot's approaches. I will begin with a detailed exploration of Megone's theory, followed by an examination of Foot's framework. I will then address some of the key weaknesses of the Aristotelian approach.

A. Christopher Megone: Understanding human nature, rationality and the concept of disorder[41]

In this subsection, I will examine Megone's (1998; 2000) theory, which challenges the claim that incorporating values into the classification of disorders undermines

[41] Ibid.

objectivity. Drawing on an Aristotelian teleological framework, he argues that disorders—both mental and physical—can be evaluated objectively by examining their impact on an individual's ability to fulfil their natural purpose or *telos*. In this framework, human flourishing is closely tied to the effective functioning of an individual, with rationality, social interaction, and other key human capabilities forming the basis of human well-being.

Megone emphasises that human nature, defined by species membership, plays a critical role in determining a person's inherent purpose. He explicitly links human flourishing to the *telos* of the species, arguing that fulfilling this purpose requires rational engagement and meaningful participation in the broader human community. As he states:

> *The Aristotelian account (…) provides a much broader context, thus relating the concept of illness to that of the human good as a whole. On this broader view, functionally explicable developments can only be understood in terms of the way such developments, or changes, contribute to a good human life. (…) Diseases or illnesses are bad (…) because they prevent the agent from exhibiting a fully rational life, which constitutes the human good on the Aristotelian picture* (Megone 1998: 14–15).

From the quotation we can see that the Aristotelian view links health and illness to the bigger picture of what it means to live a good human life. In this view, you can't fully understand things like illness or physical changes just by looking at how the body works (i.e. by looking at its functions). Instead, you need to consider whether these changes help or hinder a person's ability to live well. In this framework, diseases are considered "bad" because they prevent a person from living a life guided by reason — something Aristotle considered essential to being fully human and living a good life. Put simply, illness is bad not just because it affects the body, but because it hinders the kind of considered, rational life that Aristotle believed constituted human well-being.

For Megone, the human species' unique capacities, particularly rationality, guide the realisation of *telos*. Illness—whether physical or mental—is thus understood as an impediment to achieving human flourishing by disrupting rational capacities or other essential functions. This integrated approach presents a cohesive framework for understanding how disorders—whether physical or mental—negatively affect well-being. It does so by focusing on their shared impact on an individual's capacity to fulfil their natural purpose (*telos*), particularly the capacity for rationality and other essential human functions. This perspective treats the disruption of well-being caused by different types of disorders as fundamentally interconnected, rather than categorically distinct.

For example, if we consider depression as a mental disorder and paralysis as a physical illness, both are considered harmful in the Aristotelian framework of Megone because they impair a person's ability to lead a fully rational and flourishing life. Depression can impair a person's ability to think clearly, make decisions and engage in meaningful relationships or activities — undermining their rational agency and emotional balance, which are crucial to the fulfilment of their telos (natural destiny). Whilst paralysis is primarily a physical condition, it can prevent someone from participating in activities that contribute to their fulfilment, such as working, forming relationships or engaging in intellectual pursuits, by limiting their autonomy or independence. So rather than seeing mental and physical disorders as separate categories, Megone links them through their shared impact: Both disrupt the capacities (especially rationality and function) that define and enable a good human life.

Here Megone's view directly challenges Szasz's sharp distinction between physical and mental illness. Szasz contends that mental disorders lack the objective grounding of physical illnesses and are fundamentally value-laden, shaped by subjective societal judgments. In contrast, Megone argues that both physical and mental illnesses can be unified within the Aristotelian framework, as they impair an individual's capacity to flourish by undermining core human functions like rationality. Similarly, Megone's perspective contrasts with Foucault's (1989) critique of psychiatry. While Megone seeks to ground the classification of mental disorders in an objective understanding of human flourishing, Foucault questions whether such objectivity is possible. Foucault argues that concepts of normality and pathology are historically contingent and embedded in societal power structures. Psychiatry, according to Foucault, often serves to enforce social norms and control deviance under the guise of medical authority[42].

Megone's Aristotelian framework represents an effort to establish a universal evaluative standard for mental disorders, grounded in rationality and the intrinsic goals of human nature. This approach argues a systematic method for understanding how disorders affect well-being by considering their impact on rational capacities and other aspects of *telos*. However, Szasz's and Foucault's critique still remains: any

[42] The distinction between Megone and Szasz is primarily based on their views of rationality. Szasz rejects the idea that psychiatric judgments are rational, arguing that they are inherently evaluative and therefore subjective. Megone, on the other hand, seeks to demonstrate that psychiatric classifications can indeed be rational, relying on an Aristotelian synthesis of naturalism and normativism. However, Foucault's critique of psychiatry goes even deeper. He challenges not only the objectivity of psychiatric classifications but also the very notion of rationality as a fixed and universal foundation. By invoking rationality, Megone imposes a standard that may not be universally accepted. In the following discussion, I will explore how debates on rationality can be engaged in a way that respects individual perspectives while addressing the challenges posed by both Szasz's and Foucault's critiques.

framework for classifying mental disorders may inevitably reflect the cultural, political, and historical biases of the society in which it is constructed.

In summary, Megone's theory integrates mental and physical illnesses into a unified account, providing an objective lens to evaluate their impact on human flourishing. His emphasis on rationality as a core component of *telos* challenges Szasz's separation of mental and physical illnesses and offers an alternative to Foucault's scepticism about objectivity in psychiatry. Nonetheless, Megone's reliance on the Aristotelian tradition invites ongoing debate about whether such an approach can fully address the complexities of psychiatric classifications in a socially and historically contingent world.

B. Philippa Foot: Natural Goodness

Another Aristotelian perspective from which we can attempt to clarify the grounds for the concept of mental disorder and reinforce the impartiality of the criteria of evaluation can be found in the writings of Philippa Foot (2001). Her theory of natural goodness provides an Aristotelian framework for understanding human behaviour and moral evaluation in an evolutionary context. She argues that in order to judge human behaviour objectively, we must base our evaluations on biological standards that reflect the natural functions of living beings. In doing so, Foot (2001) integrates Aristotelian ethics with evolutionary theory, retaining the idea of function but rejecting any predetermined, metaphysical notion of purpose. Instead, she assumes that functions arise as evolutionary adaptations to specific environmental conditions and, once established, define the nature of a species and its well-being.

Foot's argument begins with the idea that goodness and defect in living organisms must be understood in relation to their species-specific life forms. She draws upon Michael Thompson's concept of "Aristotelian categoricals" or "natural history sentences," which describe the normative patterns of species-typical behaviour. These statements, such as "rabbits eat grass" (Foot, 2001: 28), do not merely indicate statistical regularities but express what is characteristic of a flourishing member of a given species. For example, a cactus with green, fleshy leaves that effectively store water is considered a well-functioning specimen, just as a deer that can run swiftly is judged to be a good deer because this ability is necessary for escaping predators. In this way, the background of a species determines the criteria by which its members are evaluated. Similarly, Foot extends this framework to humans, arguing that human virtues contribute to the well-functioning of our species, just as strong roots contribute to the health of a plant. Species evolve through variations and changes, some of which enhance survival and reproduction and stabilise over time. This process determines what works well within a species. As humans have evolved not through brute strength or speed, but through rationality and social co-operation, these abilities form the basis for assessing human well-being.

Crucially, Foot assumes that once a function has been established through evolutionary processes, it becomes a defining characteristic of the species and cannot be arbitrarily changed. Rationality, for example, is not just a random or an incidental human trait, but a species-specific ability that is essential for human flourishing. Our practical reasoning must take into account what we as humans need in order to be considered truly rational. Consequently, moral goodness is not an abstract ideal, but a reflection of what enables people to function well within the context of their evolved nature.

In this sense, Foot argues that the evaluation of human goodness follows the same form as the evaluation of the goodness of non-human animals and plants. The same evaluative structure applies to both the phrase "good roots" and "good disposition of human will" (Foot, 2001: 39). What distinguishes human goodness, however, is our ability to recognise and respond to reasons for our actions. While the lives of non-human animals are primarily focussed on self-preservation, development and reproduction, human life is also shaped by moral and rational considerations.

Thus, to judge whether a human action or trait is good, we must consider whether it aligns with the evolved nature of human life. Foot (2001: 34) provides an example from the animal world: "Because deer escape from predators by running, a certain degree of swiftness is required for a deer to be good qua deer." Likewise, human beings require rationality, social cooperation, and the ability to communicate in order to live well as humans. A lack of these capacities—such as the inability to use language or collaborate with others—constitutes a defect or disorder because it impairs an individual's ability to function according to the standards of human life.

Foot (2001: 55) emphasises that language is one such fundamental capacity:

> *But speech is crucial here in marking the difference between animals and humans. We know what an animal is going after only by what it does, whereas a child will be able to tell us. (...) When we say that human beings are able to choose on a rational ground as no animal can, it is because human action belongs in such surroundings, and so, ultimately, because humans use language not matched by anything in animal life.*

By linking human goodness to our natural capacities, Foot ensures that moral evaluations are neither arbitrary nor subjective. Instead, they are based on an objective understanding of human nature, rooted in the evolutionary functions that enable us to flourish.

Foot's framework provides a foundation for understanding mental disorders as forms of dysfunction. Since natural goodness is determined by how well an organism fulfils its evolved functions, any significant deviation from these functions is considered a defect. Just as a plant with roots that fail to absorb nutrients is defective because it

cannot sustain itself, a human being who lacks key cognitive or social capacities is considered to be impaired in fulfilling their life form's purposes.

However, Foot (2001) also acknowledges natural variation within species. Not every individual must exhibit *all* species-typical traits in order to function adequately. For instance, a tiger is typically described as having four legs, but an individual tiger with three legs does not undermine the general truth of this claim. Likewise, while some deviations from human capacities may not constitute dysfunction, others—such as the inability to reason or communicate—are significant enough to be classified as disorders. For instance, an individual born with a minor physical anomaly, such as a person with an extra finger, does not necessarily have a disorder, as this variation does not significantly impact their ability to function as a human being. However, an individual who is born without the capacity for language or the ability to form social connections is experiencing significant impairment because these functions are fundamental to human life.

By following Foot's theory, we could say that a mental disorder is not simply a socially constructed label, but an objective impairment of a person's ability to live according to their human nature. This approach strengthens the impartiality of diagnostic criteria by basing them on biological and evolutionary realities rather than cultural norms or subjective judgements.

In conclusion, Foot's theory of natural goodness provides a rigorous Aristotelian-evolutionary framework for evaluating human behaviour, morality, and mental health. By situating human goodness within the broader context of species-typical functioning, she offers an objective basis for determining what constitutes well-being and disorder. Her approach attempts to bridge the gap between Aristotelian ethics and evolutionary theory, rejecting the notion of preordained purposes while maintaining that evolved functions define what it means to live well.

While Aristotelian theories provide valuable insights into human development, they have also faced significant criticism. In the next section, I will explore these critiques and discuss the limitations and challenges of applying Aristotelian concepts to contemporary understandings of mental health and mental disorders.

C. Objections to the Aristotelian answers: Towards a synthesis of objectivity and pluralism

C.1. Criticising Megone's theory: the limits of the universal telos

As I mentioned earlier, while Megone's theory is truly important and admirable because it represents a shift in the debate about objective evaluative standards for defining mental disorders, I will argue that it is still not satisfactory enough.

I will begin with the criticism by Glackin (2016), who offers a nuanced critique of Megone's neo-Aristotelian account of disability. While Megone argues that judgements about illness are both factual and evaluative, Glackin claims that

Megone's approach overlooks the important role that social context plays in the evaluation of conditions. In particular, Glackin argues that Megone's view inappropriately blurs the distinction between facts and values. Glackin agrees with Megone's core assertion that illness is inherently evaluative and that such evaluations must take into account objective biological facts about the human condition.[43] However, Glackin criticises the assumption in Megone's account that the mere fact of a condition (such as immobility) automatically leads to a unitary judgement about its negative effects on humans. Instead, Glackin emphasises that such judgements depend on a variety of contextual social facts which can alter the evaluation of the same state. For example, immobility may be perceived as debilitating in certain environments, while neutral or even beneficial in others, as the experience of wheelchair users in wheelchair-accessible spaces shows. As described in the section on the objections to Nussbaum's approach, Glackin extends this critique to the example of deaf communities where deafness is not considered a disability in a different social context. In these cases, the condition itself is not inherently "bad" for the individual but is only viewed negatively due to social attitudes and structures. Glackin points out that when assessing the "badness" of a condition, the relevant background factors — such as social arrangements, environmental factors and cultural norms — must be taken into account. A change in the social or cultural context can therefore lead to different assessments of one and the same condition (2016: 1- 4).

In short, Glackin argues that while Megone is correct in asserting that evaluative judgements about illness are inextricably linked to the facts of the illness, he overlooks the wide range of social and contextual factors that influence the interpretation of those facts. From this, Glackin concludes that Megone's framework risks reinforcing discriminatory normative judgements that exclude individuals whose experiences deviate from conventional expectations. By focussing exclusively on the medical facts of a condition without considering its social implications, Megone's account may unintentionally ignore the role of social change in the reinterpretation of conditions or disabilities (2016: 1- 4)[44].

Another problematic point of Megone's theory is the assumption of a universal human *telos* — a specific function or purpose inherent in human nature. This assumption is reductive because it simplifies the vast and complicated spectrum of human experience to a single, unified goal. By assuming that there is an overarching purpose or function inherent in all people, the complexity of individual lives and identities is reduced to a narrow and predetermined path. Furthermore, this assumption is overly

[43] Just to remind, this discussion does not delve into the specific terminology of the philosophy of psychiatry. Instead, the focus is on establishing a method for defining objective evaluative standards. Therefore, in this context, the terms such as "illness," and "disorder" are used interchangeably to mean the same thing.

[44] This is a similar criticism to that of Begon (2023), which I presented in chapter one, section three, where I address the criticism of Nussbaum's approach.

prescriptive as it imposes a particular idea of what it means to live a good or meaningful life, potentially disregarding the unique aspirations, values and cultural backgrounds that shape each person's understanding of fulfilment. In a pluralistic society where individuals draw on a variety of cultural, philosophical and personal resources to define their own meaning, a singular, inherent human telos does not do justice to the complexity and richness of human diversity. A rigid, one-size-fits-all notion of human purpose can overlook the many different ways in which people understand prosperity and well-being. This leads to a tension between trying to impose a single vision of the good life and respecting the different beliefs of people and cultures. As a result, there is a risk that people whose lives or values do not fit into this universal vision will be excluded or undervalued, thereby ignoring the true complexity of human life.

A similar problem to the complexity of "function" is the complexity of rationality in the Aristotelian framework. While Aristotelian approaches emphasise rationality as central to human well-being, it could be argued that this focus oversimplifies the effects of mental disorders. As outlined in the first chapter, mental health encompasses a range of factors beyond rationality. These include emotional, social and cultural dimensions. Because Aristotelian theories focus predominantly on rationality, the complex impact of mental disorders on people's lives and experiences is not captured. I will use the example of people with borderline personality disorder (BPD) to illustrate how the Aristotelian approach can oversimplify the complexity of mental health by focussing predominantly on rationality.[45]

Take the example of a person with bipolar disorder, a group of mood disorders in which people experience episodes of depression - characterised by low mood, loss of pleasure and low energy — and episodes of mania or hypomania. Mania involves an overly elevated or irritable mood, increased energy and a low need for sleep, while hypomania has similar but milder symptoms (Phillips and Kupfer: 2013). Some warn of the specific nature of this disorder and that bipolar disorder can lead people who are not sufficiently cautious to irrational and deadly behaviour through productivity and euphoria (Weiner: 2011). From an Aristotelian perspective, in which rationality is seen as central to a good human life, this person's struggles could be understood primarily in terms of his difficulties in rationally controlling his or her thoughts and actions. In this view, bipolar disorder could be seen as a failure to maintain rational balance that impairs the ability to live a flourishing life. However, focusing only on rationality oversimplifies the experience. Bipolar disorder also involves intense emotional challenges that go beyond decision-making. It affects relationships, causing instability and fear of abandonment. Personal history and social factors play

[45] Later, in Chapter 5, I will provide further examples of the application of my own, more pluralistic model to mental disorders.

a big role too, and cultural expectations shape how the condition is expressed and understood. These aspects can't be explained by rationality alone.[46]

This raises a broader concern about using rigid philosophical or psychological frameworks to understand mental health. If we define well-being too narrowly—based only on rational function—we risk reinforcing harmful ideas about what counts as a good life. Such views might exclude or marginalize people whose lives don't fit into this ideal. It can lead to the mistaken belief that those with severe mental disorders are less capable of living fulfilling lives, overlooking the many different ways people find meaning and happiness despite their challenges.

In conclusion, while Megone's approach (1998; 2000) to mental health offers valuable insights, particularly by emphasizing the evaluative nature of illness, it remains inadequate in addressing the complexity of individual experiences. Glackin's (2016) critique highlights the importance of social context in shaping evaluations of illness, pointing out that Megone's framework risks oversimplifying the relationship between facts and values. Additionally, the assumption of a universal human telos within Aristotelian theories may marginalize individuals whose experiences or values deviate from this norm, thus failing to account for the diversity of human experiences. The overemphasis on rationality within these frameworks also overlooks the emotional, social, and cultural dimensions that are central to understanding mental health.

Ultimately, these criticisms underscore the need for a more pluralistic and inclusive approach that recognizes the multiplicity of ways in which people can lead fulfilling lives despite mental health challenges.

C.2. Criticising Foot's theory: the challenge of applying Aristotelian principles

Foot's theory, which is also based on Aristotelian principles, faces similar criticisms to Megone's theory. In particular, her concept of natural goodness—while offering a biologically and evolutionarily grounded framework for evaluating mental disorders—raises two key concerns: (1) the risk of reinforcing rigid normative standards, especially in areas such as sexuality and social inclusion, and (2) the limitations of her emphasis on rationality as the defining human characteristic, which fails to account for the diverse evolutionary strategies individuals may adopt to navigate their environment. I will now illustrate these two concerns with examples.

One of the main concerns with Foot's framework is that it risks reinforcing rigid and exclusionary norms by linking human goodness and defectiveness to natural functioning. By defining dysfunction (*defect*) as a deviation from species-typical

[46] In this context, Weiner (2011), for example, argues that we must not only focus on rational decision-making and control, but understand self-management as part of a broader, more flexible approach to agency and responsibility. This view accepts that people are not always in full control and that care and support are shared, ongoing processes — not just individual choices.

capacities, her theory may inadvertently prioritise conformity to established norms over an appreciation of the diversity of human experience. This has implications for areas such as sexuality, disability, and neurodiversity, where deviations from the majority experience are often pathologised rather than understood as natural variations within the human species. For instance, Foot's framework might struggle to accommodate variations in human sexuality that do not align with reproductive purposes. If one were to apply a strict evolutionary lens, non-heteronormative sexual orientations could be seen as deviations from natural functioning since they do not contribute to reproductive success. However, such an interpretation would fail to recognise the social, psychological, and relational benefits of diverse sexual orientations, as well as the fact that evolutionary success is not limited to direct reproduction but can also involve kin selection, social cohesion, and other adaptive strategies. This highlights the risk of Foot's theory imposing restrictive and outdated norms under the guise of objective biological evaluation.

A similar issue arises when considering people with severe cognitive disabilities. Under Foot's model, a person with profound intellectual disabilities or significant physical impairments—such as those caused by cerebral palsy—might be considered defective because their cognitive or physical abilities fall below the typical range of human functioning. However, such a classification overlooks the meaningful and fulfilling lives that individuals with disabilities can lead. A person with cerebral palsy, for example, may engage in advocacy, artistic expression, or form deep emotional bonds with family and caregivers, finding fulfilment and purpose through avenues not traditionally recognised as part of "natural functioning." It is visible from this case that various individuals can have different adaptive strategies. Moreover, in virtue of these different adaptive strategies, they can contribute in different ways to the evolutionary advantages of the species. By focusing too heavily on a fixed conception of normative functionality, Foot's theory risks devaluing alternative but equally valid ways of flourishing. Let's proceed with other illustrations of my criticism of Foot's theory.

A similar argument applies to neurodivergent individuals, such as those with autism. Evolutionary, social cohesion and community living are seen as fundamental to human survival. However, some individuals who prefer solitude or limited social interaction may find alternative ways to thrive. Consider a highly introverted autistic individual who struggles with conventional social interaction but excels in computational thinking, leading to groundbreaking advancements in technology. This person may defy Foot's criteria for natural goodness by not participating in typical human social behaviours, yet their unique cognitive strengths allow them to contribute significantly to society. This shows that evolution does not specify a single path for the success of an individual or its contribution to the evolutionary adaptations of species, but rather allows for a spectrum of adaptation strategies, which is not fully taken into account in the context of Foot's framework.

The case of John Nash[47], the renowned mathematician who lived with schizophrenia, further illustrates the limitations of Foot's theory. Aristotelian principles might evaluate Nash's life based on his ability to fulfil conventional social roles and rational functioning. Given his delusions and erratic behaviour, he could be classified as defective within Foot's framework. However, Nash developed an alternative strategy for managing his condition—rather than undergoing full medical treatment, which could have impaired his intellectual capacities; he learned to distinguish between real and hallucinatory experiences through rational analysis. For instance, he realised that one of his recurring hallucinations, a young girl, never aged overtime, leading him to conclude that she was not real. This strategic approach allowed him to function effectively while preserving his mathematical genius. The illustration, indeed, shows that rationality is a relevant characteristic. However, it also shows that conditions usually classified as defective adaptations, and, thus, mental disorders, do not need to be so, if individuals are able to find strategies to cope with them. Thus, Nash's case suggests that mental disorders are not always absolute defects, as he was successful in developing unique coping mechanisms that allow him to function despite his challenges. While I do not deny that some conditions could be classified as disorders and require medical intervention, others may be managed in ways that allow individuals to continue contributing meaningfully to society. Foot's (2001) framework does not easily accommodate such nuances, as it tends to classify deviations as outright defects rather than considering the possibility of alternative functional strategies.

The next illustration focuses directly on the master characteristic indicated by Foot as the defining feature of human goodness, i.e., rationality. While rationality is undeniably a crucial aspect of human life, her framework risks reducing human flourishing to a single evolutionary strategy, ignoring the fact that different individuals may adopt different adaptive strategies to navigate their environments. Imagine an emperor with visible features like those in Robert Graves' novel I, Claudius. Unlike his predecessors, who were assassinated because of their perceived political threat, Claudius survived because his physical impairments (such as his limp and speech difficulties) caused others to underestimate him and attribute limited rational abilities to him. Now, imagine that Claudius actually had limited rational abilities, in contrast to the reality portrayed in the novel. His hypothetical "defects" became an advantage in his particular environment, enabling him to avoid assassination and eventually ascend to the throne. This illustrates that traits that are typically seen as impairments can serve as successful evolutionary strategies in the right circumstances. Foot's theory, which focuses on species-wide norms rather than

[47] Here I present the interpretation of John Nash as portrayed in the 2001 film "A Beautiful Mind": https://www.imdb.com/title/tt0268978/

individual adaptability, ignores cases where what appears to be a defect in one context can be an advantage in another.

While Foot's theory of natural goodness provides a compelling attempt to ground the concept of mental disorder in objective biological and evolutionary realities, it faces significant limitations. Its reliance on normative species-wide standards risks reinforcing exclusionary and discriminatory perspectives, particularly regarding sexuality, disability, and neurodivergence. Additionally, its emphasis on rationality as the defining characteristic of human flourishing overlooks the diverse ways individuals can adapt and thrive in different circumstances[48].

A more flexible framework—one that recognises the plurality of evolutionary strategies and acknowledges the subjective dimensions of human flourishing—would offer a more nuanced and inclusive understanding of mental health. While Foot's insights into functional goodness remain valuable, they must be balanced with an appreciation for the variability and adaptability that define human experience.

C.3. Towards a synthesis: integrating objectivity and pluralism

Both Megone's and Foot's theories are based on Aristotelian principles and focus on evaluating mental disorders through the lens of natural functioning and human purpose. I will briefly summarise their commonalities.

They both emphasise that human health and flourishing are closely linked to the fulfilment of natural functions. Foot focuses on the concept of "natural goodness" and argues that human behaviour should be judged according to how well it conforms to these natural functions, such as rationality and social interaction. Similarly, Megone uses an Aristotelian teleological approach to argue that mental and physical disorders are harmful if they interfere with an individual's ability to fulfil their telos, particularly the capacity for rational thought and social engagement.

In both theories, the notion of what is "natural" or "normal" is central to the judgement of what is a disorder. They aim to provide objective standards for evaluating mental

[48] To remind, this is similar to the critique by Shane Glackin (2016) that I mentioned in the section on the critics of Nussbaum's approach in the first part of the dissertation. Namely, he (2016) critiques species-based approach by arguing that species are fluid and arbitrary categories rather than fixed entities, making species-typical capacities an unstable foundation for moral and functional norms. Drawing on evolutionary theory (Rachels, 1987), he contends that moral considerations should be based on individual characteristics rather than species membership. He argues that rights and moral considerations should be based on individual characteristics rather than species membership, as each person possesses unique qualities that cannot be captured by a general species-based standard. This criticism is particularly relevant to the concept of Foot, which runs the risk of unjustifiably privileging certain human capacities — such as rationality or social co—operation - over others. Glackin also challenges the assumption that certain capacities, such as hearing, are inherently superior, highlighting how Deaf communities create meaningful modes of communication. Using science fiction's concept of "remaking," he illustrates how human flourishing is context-dependent rather than tied to fixed biological traits. This was also emphasised in the section on the criticism of Megone's framework.

and physical disorders based on their alignment with the natural functions essential for human flourishing and reject subjective or arbitrary judgements. Because of this focus on natural functioning and universal standards, this type of theorising can lead to problems in accounting for different individual experiences and deviations from conventional norms.

To address these criticisms, it is important to explore alternative approaches that strike a balance between objectivity and respect for the diversity of individual experiences. One promising strategy for achieving this balance is Rawls' concept of public justification (2005). Rawls (2005) argues that norms and policies are only legitimate if they are justified through a public reasoning process that is accessible to all citizens, regardless of their social status or cultural background. This approach could mean that the criteria used to evaluate mental disorders should be subject to broad, inclusive consultation to ensure that they are fair and take into account different perspectives.

Building on the Rawlsian framework, in the next chapter I will examine George Graham's (2013) proposal that integrates Rawlsian principles into a multidimensional understanding of mental disorders. Graham's approach incorporates insights from different disciplines and perspectives, promoting a more integrative view of mental health. His framework addresses the complexities highlighted by Aristotelian approaches and seeks to emphasise objectivity while acknowledging the diversity of human experience.

3.3. Section Three: Graham's Rawlsian strategy[49]

As mentioned above, the evaluation and definition of mental disorders in psychiatry is a major challenge, as it requires a balance between objectivity and sensitivity to diverse perspectives. An innovative approach to overcoming these challenges is George Graham's (2013) application of the "original position", a conceptual tool borrowed from Rawls' theory of justice (1999: 15–19).

As a reminder, Rawls' original position is a thought experiment that aims to derive impartial principles by placing individuals behind a "veil of ignorance" where they are unaware of their personal circumstances or biases. Graham applies a similar methodology to explore evaluative standards that determine when a mental state should be categorised as clinically significant or a mental disorder. In doing so, he focuses on universal psychological capacities and employs a fairness-focused perspective. His goal is to avoid subjective bias and establish criteria that are fair and equitable and take into account different individual experiences.

Therefore, in this section I will explore the key psychological capacities that Graham (2013) identifies as fundamental to mental health and how his approach aims to ensure that evaluations of mental disorders is based on universally applicable, neutral

[49] The analyses and theses in this section were developed in collaboration with Elvio Baccarini and the JOPS research project.

standards. In addition, I will discuss the wider implications of this approach for psychiatric practise. It promotes a more inclusive and dialogic process in deciding which conditions warrant medicalisation, addressing ongoing debates about the balance between universal standards and individual experience in mental health assessment. However, I will discuss that the use of such an abstract and impartial methodology also has potential limitations, as it cannot fully capture the nuances of personal and cultural differences in mental health.

Graham's (2013: 150) approach, similar to Rawls' use of the original position, involves evaluating mental disorders through a lens in which we set aside specific personal details but retain an understanding of general human needs and circumstances. In this framework, we recognise some capacities as ones that we are "generally bound to need or care about, regardless of which particular goals [we] may have" (Graham 2013: 156). These are capacities that we absolutely need, regardless of how they vary for us. Graham believes that the loss of one or more basic psychological capacities increases the chance of losing a healthy quality of life if they are not moderate cases (Graham 2013: 164). And if they are disabled or impaired, we may find ourselves in a condition of mental disorder (Graham 2013: 152).

To operationalise this, Graham identifies seven basic psychological capacities that are crucial for maintaining mental health (Graham 2013: 157-159):

1. Bodily/spatial self-location – That is, physically locating oneself in order to know where we are and where we are in relation to other important objects.
2. Historical/temporal self-location – That is, to recognize or understand the past as our past and the future as our future.
3. General self/world comprehension – that we may comprehend ourselves and the universe to the practical level or degree necessary for life in a somewhat well-informed and educated manner. This includes understanding specific facts or situations.
4. Communication – It is important to know how to be both: listener and speaker; using the right words and having a sensitive ear.
5. Care, Commitment and Emotional Attachment or Engagement – to take care of other things and people besides yourself, to feel things.
6. Responsibility for self - to take care of ourselves as the person we are and the way we want to be.
7. Recognizing and Acting on Opportunities. – Recognize opportunities and decisions and act on them.

Among these essential psychological capacities is consciousness or phenomenal experience, which encompasses our capacity for conscious awareness. Equally important is our capacity to reason and respond to reason, as it enables us to make decisions, solve problems and navigate complex social and personal situations. In

fact, Graham states that mental disorders can be understood as an impairment of this very capacity for reason-responsiveness:

> *"If we are to successfully engage in the world, we need more than a capacity for conscious experience. We need more than the capacity to reason and be reason-responsive. We need capacities for performing activities that, in our being committed to them and aiming for their outcomes, help to make life structurally satisfying and enable us to function well in the world."* (Graham 2013: 156).

Graham explains that when an impairment in a person's capacity to respond to reason is harmful, this impairment contributes to the development of a mental disorder (Graham 2013: 136). In mental disorders, there is a mixture of causes — both without rational intentionality (which he calls "mere coils") and with rational intentionality (responsive reasons). He calls this phenomenon "interactive co-causation" between unreason and reason. Even in severe mental disorders where irrationality is prominent, the disorder still retains some elements of rationality so that it is not completely senseless (Graham 2013: 137).

In summary, Graham (2013) aims to establish criteria for identifying the basic mental capacities relevant to mental disorders by applying Rawls' methodology to avoid bias and ensure that judgements are fair and just, just as Rawls aims for impartiality when selecting principles of justice. However, the question arises as to whether the use of such an abstract and impartial methodology can fully capture the nuances of personal and cultural differences in mental health.

As we saw in the first chapter, Nussbaum, for example, has challenged these abstract methods by emphasising the importance of including specific human capacities and the concept of human dignity in the framework of justice. She argues that Rawls' approach neglects the need to consider what individuals can do and be in their lives, and not just their basic rights (Nussbaum: 2006). Amartya Sen has also criticised Rawls' framework, claiming that it is too theoretical and does not sufficiently address practical aspects of justice and welfare. Sen emphasises the need for a more pragmatic approach that considers the real contexts and capacities of individuals (Sen: 2009).

Similarly, I argue that Graham's use of an abstract, generalised approach to assessing mental disorders may fall short in practise. Just as Rawls' method has been criticised for focusing primarily on basic rights, freedoms and opportunities without fully capturing deeper personal values and aspirations, Graham's reliance on abstract criteria to determine mental disorders may not do justice to the complexity of individual experience. People need deeper self-knowledge of their values, as well as insights into the contexts in which they live, to recognise which capacities are truly relevant to them.

To further illustrate the problem of relying solely on common knowledge behind the veil of ignorance, let us consider a few hypothetical scenarios. Imagine an individual

who is blind but but has an extraordinary innate talent for music. If this person lives in a society where music is not valued or is entirely absent, their blindness might appear to be a significant limitation, with no relevant compensatory benefit. However, in a society that deeply values musical expression, the same individual might become a celebrated artist, with their blindness even enhancing their focus and ability in music due to heightened auditory sensitivity. Graham's model, which assumes a broad list of capacities for success—such as sensory abilities or motor skills—fails to take into account these context-dependent variations. By treating certain abilities as universally valuable without considering how their significance might shift depending on social and cultural environments, the model risks being too coarse-grained. It overlooks the reality that some traits, which might seem disadvantageous in one setting, could be strengths in another. This highlights a key shortcoming of Graham's approach: the assumption that we can determine what abilities are universally necessary for humans without a deeper understanding of what matters to individuals in their specific contexts. His list of capacities, chosen behind the veil of ignorance—where individuals do not know their specific circumstances—may be too abstract and overly general to adequately account for the complexities of real-world human experiences. Instead, a more fine-grained approach is needed—one that does not merely list generic human capacities but also considers the values, social structures, and individual aspirations that shape what is meaningful and beneficial in different contexts.

Another scenario might involve a person with an emotional impairment that manifests as a paralysing form of shyness towards individual of the gender they find attractive. This shyness prevents them from forming intimate relationships. For someone who is sexual and values intimate relationships as central to their well-being and personal fulfilment, this impairment is profoundly limiting. In contrast, for an asexual individual who does not prioritize or desire intimate relationships, the same emotional impairment may be far less significant or even irrelevant to their sense of a flourishing life. Graham's (2013) framework, rooted in generalized capacities for success, does not adequately address such variability in individual values and aspirations. It treats emotional capacities broadly without recognizing how specific impairments intersect with personal priorities. Decisions made within this framework, which overlook the importance of individual differences, risk failing to accommodate the diverse ways in which impairments affect people's lives depending on their worldviews and values.

The point of these examples is to show that Graham's generalized approach fails to account for the situational and personal relevance of certain capacities, leaving individuals and societies unable to fully appreciate or support unique talents and aspirations. By emphasizing only broad, abstract traits, the model risks neglecting the nuanced interplay between individual differences and specific contextual values. In other words, his model is not fully adequate because it imagines people who think only in terms of their common characteristics. However, disabilities can vary significantly across individuals with regard to their personal characteristics. For one

person, an impairment may be completely irrelevant, while for another, it may be profoundly crucial—depending on their worldview, values, and the circumstances in which they live[50].

Although I share Graham's basic ideas (e.g. defining a mental disorder by analysing unresponsiveness to reasons), I cannot fully support his proposal for the reasons mentioned above. The main problem I have with his approach is the depersonalization and the overlooking of specific individual evaluations that define what is truly important to people and take into account their unique characteristics. Therefore, I believe there needs to be a more sophisticated model—one that not only respects what individuals value but also opposes the imposition of sectarian views.

To address the challenge of balancing fairness and personal relevance in defining mental disorders, while avoiding sectarian impositions, I will propose a model in the next section based on specific kind of epistemology and public justification, inspired by Gerald Gaus (2011). Public justification ensures that the standards for defining mental disorders are based on the convergence of different perspectives rather than the imposition of a single viewpoint. This approach aims to create an inclusive and equitable system that recognises diverse values and experiences, ensuring the criteria for mental disorders are not shaped unfairly by one perspective. I will begin by introducing this model, drawing on the article by Baccarini and Lekić Barunčić (2023)[51], which explains how public justification can distinguish disorders from mere diversity. I will then address what it means to be in a state of disorder using a weak externalist justification inspired by the weak externalist epistemology of Gaus (2011). This approach emphasises that, when classifying conditions as disorders, we need to assess when a person is unresponsive to reason.

In other words, the first question is about defining and agreeing on what a mental disorder is in the general sense. The second question is about applying that definition to determine whether a particular person is affected by that specific disorder. In my detailed discussion, I will focus on whether a person's condition fulfils the criteria for being classified as a disorder – whether they are unresponsive to reasons - according to the general definition.

3.4. Section Four: The solution: A new approach to defining mental disorders

A. Justification of General Classifications

As mentioned above, I will first examine the detailed framework proposed by Baccarini and Lekić Barunčić (2023), which provides a method of public justification to determine whether certain conditions should be categorised as disorders rather than

[50] As shown in chapter one, in the first part of the dissertation in the critiques by Glackin (2016) and Begon (2023)

[51] This approach represents a collaboration with the JOPS project and is inspired by Gerald Gaus' theory of public justification (1996, 2011).

simply forms of diversity. The innovative aspect of their method lies in the emphasis on public justification as the key criterion. This principle is crucial for distinguishing between impairments that prevent a person from leading a dignified and fulfilling life and conditions that are natural variations within the spectrum of human experience. By elaborating a method for justifying classifications, the model protects against the imposition of inappropriate values or norms and contributes to the development of a justified medicalisation— that ensures that classifications are ethically and socially justifiable (Baccarini and Lekić Barunčić, 2023).

The inspiration for the proposed method lies in Gaus's (1996; 2011) theory on the convergence of public justification. Gaus claims that public justification is not achieved through consensus, as Rawls (1999; 2001; 2005) argues, but through the convergence of different reasons. This means that justification occurs when different individuals with different perspectives and reasoning processes independently come to overlapping conclusions that create a mutual tendency to agree. This qualified convergence of reasons means that all qualified individuals, or their representatives acting on behalf of unqualified individuals, agree.

To understand the relevance of this proposition, we must first clarify what "qualified" individuals means in this framework. Qualified individuals are those who possess basic reasoning skills, which encompass several key capacities. First, they must have the capacity to respond to reasons in their actions and beliefs. This involves an openness to rational deliberation and external evidence, enabling them to adapt their beliefs or behaviours when presented with sound reasons. Second, qualified individuals must demonstrate the competence to draw logical conclusions and to identify and resolve contradictions. This entails deriving consistent and coherent outcomes based on available information and addressing inconsistencies either in their own reasoning or in external data. In addition, qualified individuals need the capacity to develop and revise their conceptions of the good. This reflects a willingness to re-evaluate and adapt their values, aspirations, and goals considering new experiences, insights, or circumstances. Importantly, this process requires a level of intellectual openness—a readiness to reflect on and refine personal ideals when presented with reasons to do so. Finally, qualified individuals must possess the capacity to engage in reciprocal discussions about social norms. This includes facilitating meaningful communication, sharing diverse perspectives, and collectively navigating the complexities of social norms through respectful dialogue with others. By establishing clear criteria for qualified individuals, this method ensures that evaluative standards—such as those used to determine whether a condition constitutes a mental disorder—are grounded in principles broadly acceptable to all relevant parties. This prevents definitions from being imposed solely based on the perspectives of experts or dominant social groups.

The emphasis on inclusivity and public justification ensures that decisions about classifying conditions as disorders are neither arbitrary nor exclusionary. While

experts such as psychiatrists and clinicians play a crucial role in providing evidence-based insights, their contributions must also be publicly justifiable. In other words, these insights need to be defensible and acceptable to a diverse audience, including individuals whose lived experiences may challenge or enrich dominant narratives about a condition. This approach mitigates the risks of pathologising diversity or enforcing inappropriate norms by ensuring meaningful consideration of all perspectives, including those from marginalised or atypical groups (Baccarini and Lekić Barunčić, 2023). Aligned with Gaus's theory of convergence, it promotes legitimacy in defining mental disorders by grounding classifications in a consensus of justifiable reasons, avoiding the imposition of top-down definitions. Furthermore, it adopts an inclusive approach to assessing impairments, respecting the diversity of human experiences while focusing on genuine obstacles to well-being.

Another important aspect of the method is that defining disorders requires more than democratic validation; it also necessitates contextual validation, allowing expert bodies to provide insights, particularly in complex cases. This approach accounts for societal changes and evolving understandings of conditions, as illustrated by the neurodiversity movement. This movement challenges traditional views of autism as a deficit, advocating for societal adaptations rather than medicalisation. Disability should not be seen solely as arising from internal impairments but also from social and environmental barriers that exclude individuals from full participation in society. For example, autism is often associated with difficulties in communication, social interaction, and behaviour, which may impede societal functioning. However, these challenges often stem from societal barriers rather than the condition itself. If an autistic person experiences sensory overload in noisy or crowded environments, the issue may lie in the absence of sensory-friendly spaces rather than an inherent disability. Removing such barriers — through adaptations such as sensory-friendly environments or social skills training — can lead to the condition being seen as an expression of social injustice rather than a disability (Baccarini and Lekić Barunčić, 2023).

A two-stage process is proposed to justify the general classification of conditions. The first stage involves determining significance, where the characteristics of a condition are evaluated to establish whether they justify classifying it as a disability[52]—that is, whether the condition significantly impairs an individual's capacity to function in society. The second stage focuses on identifying causes. This entails assessing whether the deprivation of opportunities arises from the individual's impairment or from unjust social or economic conditions. As noted above, if exclusion or

[52] As already mentioned several times in the dissertation, the terms "disability"," "illness," and "disorder" are used synonymously in this dissertation. The reason for this is that the discussion does not focus on the specific terminology of the philosophy of psychiatry. Rather, the focus is on establishing a method for defining objective evaluative standards. Therefore, these terms are used in a general sense to convey the same meaning in the context of this thesis.

marginalisation is primarily due to societal structures, the condition may not qualify as a disorder but instead reflect social injustice (Baccarini and Lekić-Barunčić, 2023).

This approach, grounded in public justification and pluralism, ensures that the classification of conditions as disorders respects the complexity of human experiences and avoids pathologising diversity. It offers a more inclusive, fair, and adaptive method for determining what constitutes a disorder, as opposed to simply labelling differences as disabilities without adequate justification. I am in favour of this proposed model of public justification because it serves as a safeguard against oppression, addressing the first part of Szasz's and Foucault's challenge that I discussed earlier.

The aim of this section was to present and justify a new approach to defining mental disorders—an approach based on public justification and consideration of pluralism. Drawing on the work of Baccarini and Lekić Barunčić (2023) and Gaus' theory of public justification (1996; 2011), I have outlined a framework that guards against the imposition of external values and ensures that the classification of disorder is justifiable, inclusive, and responsive to societal change. This model addresses Szasz and Foucault's critique of the oppressive potential of psychiatry by ensuring that definitions respect the autonomy and diversity of the individual. Furthermore, this general framework for classification has served me as the foundation for determining whether a specific individual is experiencing a mental disorder, once the general characteristics of mental disorders have been established.

In the next section, then, I will move from this justification of the general classification of disorders to the more specific question of whether a particular person is in a state that meets the agreed definition of a disorder.

B. A weakly externalist model of justification for the determination of mental disorders [53]

So far, I have examined the broader question of when a condition qualifies as a mental disorder—for example, whether autism or homosexuality should be classified as such. Now, the focus shifts to a different but equally important issue: determining whether a specific individual is actually in a condition of disorder. Even if a condition is generally recognised as a disorder, it remains crucial to assess whether an individual meets the criteria for that diagnosis in a meaningful way. This shift in focus moves from defining disorders in general to considering their application in particular cases, raising questions about subjectivity, context, and the role of individual experience in psychiatric evaluation.

[53] The ideas for this approach were developed in collaboration with my supervisor, Elvio Baccarini, and the JOPS research project. More specifically, Baccarini and Lekić-Barunčić present it in their work (2023) and the main idea is part of a co-authored paper with Baccarini and Shane Glackin.

To clarify this, consider the following example. Anxiety is generally harmful, but does a person with a specific phobia necessarily have a disorder? For instance, having anxiety because of public speaking is not necessarily a disorder. Someone may simply have little interest in socializing and be perfectly content working from home. Similarly, a lack of motivation to work or socialise is often seen as a symptom of a disorder, such as depression. However, this is not always the case—if someone has a clear reason for their condition, such as an irredeemable loss, their withdrawal may be a natural and understandable response rather than a sign of mental disorder. While phobias can be classified as disorders in general, the key question remains: is an individual experiencing a phobia necessarily in a state of disorder?

A further illustration is suicide by some authors defined as a symptom of disorder or disorder itself (Maung 2022). The main argument here is that conditions should not be considered disorders if they are justified by the person's own reasons. To support this, the analysis will be grounded in Gaus's (1996, 2011) theory of weakly externalist epistemology, which emphasizes the importance of evaluating an individual's responsiveness to their own core values and reasoning. By assessing whether a person is capable of acting in accordance with their personal beliefs and commitments, we can distinguish between conditions that genuinely undermine autonomy and those that do not.

Weakly externalist epistemology preserves individual autonomy while identifying genuine impairments that may hinder a person's ability to live according to their principles and conception of a good life. This distinction is crucial—without it, there is a risk of mixing voluntary deviations from social norms with behaviours that indicate an underlying disorder. Failing to make this distinction could lead to the pathologisation of diversity and an erosion of personal freedom. In this context, grounding psychiatric diagnosis in an individual's system of reasons is essential. This approach aligns with Szasz's (1960, 1994) call to respect human diversity, ensuring that psychiatric evaluations remain sensitive to both individual autonomy and the broad spectrum of human experience.

In other words, the main question is: how do we assess a condition in light of a person's system of reasons? The "system of reasons" includes their beliefs, values, preferences, emotions, and interests—everything they consider significant and meaningful. For example, we might ask*: Is public speaking a relevant consideration within a person's system of reasons? Is their lack of motivation to work and socialize justified by their system of reasons? Is suicide justified, or is continuing to live justified, according to their system of reasons?* These questions highlight the importance of evaluating conditions not in isolation but in relation to the individual's own values and perspective.

Thus, the central concept here is unresponsiveness to reasons. The main question is: does establishing that a person is in a condition of disorder follow from their system

of reasons? For example, does the need to socialize or work follow from a person's system of reasons? Does apathy or indifference towards life align with their system of reasons?

Crucially, this system of reasons requires a degree of reflectivity to ensure coherence. This reflective process—whether or not it involves deep philosophical contemplation—helps individuals align their values, preferences, and beliefs in a way that is meaningful and consistent. Here, it is important to note that systems of reasons are not static; they are dynamic and can be revised in the light of new experiences, insights or changed circumstances. However, any revision must include reflection to maintain coherence within the system. Impulsive reactions or fleeting instincts alone do not constitute a meaningful revision of one's system of reasons. While this approach focuses on the individual's reasons, it remains "externalist" in that it does not require the individual to consciously comprehend all the implications of their value system. The key point is that their system of reasons forms the basis for justification, even if they are not aware of all its components. The important aspect is whether their system of reasons, even if not fully conscious, enables meaningful, coherent decision making. In the third chapter, I will discuss reflectivity in more detail in relation to the work of Michel DePaul (as summarised by Baccarini, 2007). The question will be how to determine which values are relevant, considering that a person's system of reasons is dynamic. More specifically, in the next chapter I will explain how we can deal with this dynamic.

Before continuing, I must emphasise that in applying this method I will focus on *self-regarding* cases, which is why it is crucial to distinguish between self-regarding and *other-regarding* cases crucial when assessing the unresponsiveness to reasons in different contexts. In this explanation, I draw on the distinction made by John Stuart Mill and further elaborated by Baccarini 2013[54]. Self-regarding cases involve actions or behaviors that affect only the individual's legitimately personal sphere, while other-regarding cases directly impact the legitimate interests or rights of others. For example, a person who chooses to live in a tiny house and minimise their material possessions might be seen as eccentric or impractical. However, as long as they find fulfilment and do not cause harm to others, their decision should not be pathologised. Similarly, someone who spends several hours a day playing video games may be judged as unproductive, but intervention is only justified if the behaviour results in harm to the legitimate interests of others, such as neglecting responsibilities or deteriorating the health of people for whom the person is responsible. In these cases, the individual's internal system of reasons is sufficient for evaluation. On the other hand, other-regarding behaviors, like reckless driving, directly endanger others and

[54] In this discussion, I primarily follow Baccarini (2013: 6–15); https://www.academia.edu/5651503/Mill_udzbenik, who builds on John Stuart Mill's distinction, which has been further elaborated by Berger (1984), Crisp (1997), Rawls (2007), Riley (1998), Ten (1990) and Gaus (1981).

violate their rights, requiring intervention based on shared standards of justification. Mill's discussion of liberty connects with this distinction, as he emphasises the need for individual freedom while recognising the role of society in protecting others from harm. According to Mill, society can only legitimately interfere with the freedom of the individual when it is a matter of preventing harm to others. He formulates this principle in the form of two maxims: one allows freedom for actions that only affect the individual, the other allows intervention when these actions harm others. Mill also emphasises that personal discomfort or negative feelings alone, such as disapproval of alcohol consumption or consensual same-sex relationships, do not justify social intervention. His utilitarian approach advocates the maximisation of social happiness but also links individual freedom with personal development and the pursuit of truth. For Mill, personal development and freedom are essential for both individual flourishing and social progress. Freedom provides space for individuals to cultivate their unique talents and contribute to the betterment of society (Baccarini: 2012 6-15). In other words, in cases where an individual's behaviour affects others, the concept of public justification becomes important. As already mentioned, public justification is about the exchange of reasons between individuals, which presupposes that the reasons are valid in interpersonal contexts. Here, unresponsiveness to reasons must be judged on the basis of shared principles and standards of justification and not only on the basis of the individual's internal system of reasons. Thus, while self-regarding cases require evaluation based on personal values and commitments, in other-regarding cases, public justification plays a crucial role in determining whether unresponsiveness signifies a disorder.

In summary, by integrating a weakly externalist model of justification and focusing on self-regarding cases, the proposed framework offers a nuanced, individualised approach to the challenges posed by the main criticism of the second part of this dissertation of the lack of objective evaluative standards in psychiatry. This approach ensures that mental disorders are evaluated based on an individual's own system of reasons, while avoiding the imposition of external values.

In the following sections of the dissertation, I will use examples to illustrate how unresponsiveness to reasons can be better understood and analysed by proposing the consideration of additional criteria such as temporal relevance. The aim of the next chapter is to examine the temporal dimension of the individual system of reasons, in particular how it develops over time and how personal experiences influence decision-making. The chapter will deal with two central questions. The first is how an individual's priorities and system of reasons change over time. The second is how important it is to understand a person's life history and experiences when it comes to interpreting their behaviour and reasoning, because symptoms alone are not enough to understand the full context of their situation. The chapter also emphasises the need for a dynamic and reflective approach when assessing personal reasons and recognises that people's systems of values and beliefs evolve through their life experiences. It

discusses the challenge of identifying which values are relevant at any given time and how to navigate the fluidity of personal reasoning. The aim is to emphasise the balance between life experience and reflection and to suggest that growth and change should be considered in a broader, more inclusive context. After this chapter, I will also show in the next chapters how the weak externalist justification can be applied to psychiatric diagnosis.

3.5. Section Five: Conclusion of Chapter Three

In chapter three, I addressed the complex and nuanced issue of defining mental disorders within a framework that respects individual values and reason-responsiveness while protecting against potential oppression. Several important points emerged from this exploration.

Firstly, I acknowledged that the definition of a mental disorder is value-laden. I began by introducing the challenge posed by the antipsychiatry movement, specifically the views of Szasz and Foucault. The central issue is the difficulty in distinguishing between emotional distress stemming from a person's unique system of reasons (such as beliefs, preferences, emotions, and other factors they consider fundamental) and a genuine mental disorder. This value-laden aspect has historical significance, as demonstrated by the categorisation of homosexuality as a mental disorder, which highlights the potential for oppressive power dynamics within psychiatry. This underscores the importance of seeking objective evaluative standards. I analysed two responses to this challenge: the first being the Aristotelian approach, and the second, Graham's Rawlsian-inspired approach.

The Aristotelian perspective suggests that understanding the function of an individual in the context of the human species is essential. This framework places rationality at the centre of human function and contributes to the discussion of mental disorders by emphasising the importance of rationality for a fulfilling life. However, I disagree with this view, as I believe that focusing exclusively on rationality can overlook other important aspects of well-being and personal fulfilment.

The chapter continues with George Graham's Rawlsian approach, who calls for a model of public justification that seeks to take into account the reasons of all parties and honour pluralism. This model seeks to prevent the imposition of sectarian views in defining mental disorders and provides a mechanism that ensures objectivity while respecting individual values. However, I argue that this approach is not entirely satisfactory; as I believe, it still does not take into account the deep-rooted societal biases and power imbalances that can influence what is considered a mental disorder.

I have presented a new solution to define and diagnose mental disorders – a specific method of public justification and specific method of epistemology inspired by Gerald Gaus (1996; 2011). Specifically, I have argued that Baccarini and Lekić Barunčić's (2023) framework for classifying mental disorders inspired by Gerald Gaus (2011) provides a more inclusive, ethically defensible approach that respects individual

autonomy and societal diversity. By grounding the classification of disorders in public justification, the framework ensures that decisions are based on a convergence of reasons rather than imposed norms. Their framework also safeguards against pathologising diversity by emphasising the importance of considering both societal and individual factors when determining whether a condition qualifies as a disorder.

The proposed epistemological method, which involves a weakly externalist justification inspired by Gaus (2011) and focuses on self-regarding cases allows for nuanced and personalised judgements that prevent the imposition of external values. Ultimately, I aim for this solution to provide an adaptive and socially responsive model for identifying mental disorders, ensuring that the classification process is fair, just, and sensitive to the complexities of human experience.

As mentioned earlier, the following sections of this dissertation will demonstrate the application of the proposed weak externalist model of justification. First, I will explore how the passage of time, life stages, and changing circumstances influence the diagnosis and understanding of mental disorders, highlighting the crucial role of the temporal dimension in understanding unresponsiveness to reasons in this context. Next, I will apply the weak externalist model of justification to analyse the rationality of suicide, considering both individual reasoning and external societal influences. I will then investigate impostor syndrome through the weak externalist justification lens, examining personal reasoning and societal pressures. Finally, I will apply weak externalist justification to a broad range of mental disorders, evaluating its potential as a framework for diagnosing mental disorder, ensuring objectivity, and respecting individual and societal diversity.

4. CHAPTER FOUR: THE TEMPORAL DIMENSION OF UNRESPONSIVENESS TO REASONS IN MENTAL DISORDERS: A DYNAMIC APPROACH TO SYSTEMS OF REASONS

This chapter examines the crucial role of the temporal dimension in understanding unresponsiveness to reasons within the context of mental disorders. Building on existing frameworks that identify unresponsiveness to reasons as a key feature of mental disorders (Graham 2013; Dembić 2023), I argue that incorporating the temporal aspect allows for a more nuanced approach to defining and assessing these conditions.

When considering the temporal dimension, we come up against two central issues. The first concerns the idea that a person changes over time. To illustrate this, we can look at the mythological figure of Hecuba, the queen of Troy, who was the wife of King Priam and the mother of many children, including Hector, Paris, Cassandra and Polyxena. Her story, especially during and after the Trojan War, is a profound tragedy. In a prophetic dream before the birth of her son Paris, Hecuba gave birth to a firebrand that was to burn Troy. According to the seer Aesacus, this dream was a warning that the city would fall because of her unborn child. Priam then gave the order to abandon Paris on Mount Ida, but he survived and was raised by shepherds. Eventually, Paris made his way back to Troy, where he kidnapped Helen and helped to start the war. Hecuba endured much as queen during the Trojan War and lost many of her children, including Hector, Troy's best warrior, who was slain by Achilles. Troilus, another of her sons, was also killed. She could only watch as Troy was dominated by the Greeks despite her futile attempts to defend her city and her family. After the destruction of the city, Hecuba was taken as a slave by the Greek conquerors. In Euripides' tragedy *Hecuba*, she is portrayed as a woman who endures unbearable pain and seeks revenge. Polyxena, one of her daughters, was offered as a sacrifice at the tomb of Achilles. She had given King Polymestor of Thrace custody of her youngest son, Polydorus, but he was killed for his wealth. Hecuba blinds Polymestor and murders his sons as ruthless retribution after discovering Polydorus' murder[55]. By examining Hecuba before and after the war, we can reflect on what was most important to her at different points in time and what constituted her system of reasons. Observing the shifts in her actions suggests that a person's system of reasons is dynamic. The first key question, therefore, is: what exactly defines her system of reasons?

The second challenge posed by the temporal dimension concerns the individuality of symptoms. One of the classical positions in the philosophy of psychiatry argue that if a person exhibits certain symptoms, then we can claim they are in a specific mental state.[56] However, I argue that by introducing the temporal dimension, we cannot fully

[55] https://en.wikipedia.org/wiki/Hecuba_(play) Accessed on 10.03.2025.

[56] For instance, there are debates regarding neopositivism in the articles: Aragona, M. "Neopositivism and the DSM psychiatric classification. An epistemological history. Part 1: Theoretical

comprehend a person's state by merely observing their symptoms at a given moment. More precisely, regardless of the symptoms, we cannot determine a person's condition without also considering their broader life story. For example, if we do not take Hecuba's full life story into account, we might quickly conclude that she is suffering from a mental disorder. However, when we consider her circumstances in their entirety, we may instead recognise that her mental state is a rational response to extreme grief and suffering. Thus, the second key issue is that we cannot fully understand an individual's responses unless we take a more holistic perspective. In this context, we may acknowledge that some states may initially appear to indicate a mental disorder—such as difficulties in work performance, which many consider important. However, to confirm this, we would have to assume that work ability is an objectively overriding reason that applies to everyone, regardless of their personal values. This assumption is rejected because individuals have the right to form their own systems of reasons based on what is personally meaningful to them. This leads to two key arguments: (a) identifying symptoms as possible signs of mental disorder requires understanding them in the context of what matters to the individual, and (b) this perspective varies from person to person. In some cases, what seems like depression might actually be sadness or unhappiness rather than a clinical condition.

The discussion in this chapter contributes to psychiatry from a philosophical perspective by addressing the attribution of values. However, this leads us to the core question and objective of this chapter: how do we determine which values are relevant? The answer is that relevant values are those that are significant within a person's system of reasons. The challenge, as previously noted, is that this system is dynamic. The next question, therefore, is: how do we navigate this dynamism? I will argue that the answer lies in recognising that people change their reasons throughout life. However, not every change necessarily indicates a shift in their entire system of reasons. At times, an individual may simply be acting impulsively. To illustrate the dynamic nature of systems of reasons, I will draw on the work of Michael DePaul, as summarised by Baccarini (2007).

I will proceed as follows: I begin by briefly introducing the role of the temporal dimension in understanding unresponsiveness to reasons in mental disorders. This will establish my key argument: a person's system of reasons is dynamic and cannot be fully understood without considering how it evolves over time. I will then explain why incorporating the temporal dimension provides a more nuanced understanding of a person's system of reasons when assessing whether they are in a state of mental disorder. Next, I will explore how individuals change over time and how this affects their system of reasons. I will examine how circumstances reshape a person's priorities, values, and actions, raising the question: if a person's values and reasons

comparison." *History of Psychiatry* 24.2 (2013): 166-179; Tekin, Şerife. "Participatory interactive objectivity in psychiatry." *Philosophy of Science* 89.5 (2022): 1166-1175.

shift over time, how should we know if they are unresponsive to reasons? Folowing this, I will address the challenge of accounting for changes in a person's system of reasons over time. I will argue that not all changes indicate deep shifts—some may be impulsive or temporary.

Temporal relevance refers to the role that time plays in understanding and assessing mental disorders. In the context of mental health, it involves considering how a person's responses to various stressors, challenges, or emotional states evolve over time. This concept is critical because mental conditions are not static; they change in response to both internal and external factors. The duration, progression, and fluctuations of symptoms[57] are essential in determining whether an individual's unresponsiveness to reasons reflects a temporary reaction to life circumstances or a more enduring and persistent disorder.

As mentioned earlier, an individual's system of reasons—encompassing their beliefs, values, preferences, emotions, and interests—is inherently dynamic and evolves in response to new experiences or changing circumstances. However, for any revision of this system to be meaningful, it must involve reflection and a degree of coherence. Impulsive reactions or fleeting instincts, while part of human experience, do not amount to genuine change in how one perceives or engages with the world. This perspective shifts the focus away from merely identifying and addressing symptoms at a specific moment in time. Instead, it emphasises understanding the broader track of an individual's mental health—how symptoms emerge, evolve, and respond to interventions over time. By adopting this approach, it becomes possible to gain a more nuanced understanding of mental disorders, one that acknowledges the fluid nature of human experience and the ongoing interplay between internal systems of reasoning and external influences. This temporal and developmental perspective encourages a deeper appreciation of the complexities of mental conditions, moving beyond static or momentary assessments toward a framework that respects the dynamic processes shaping individual well-being.

As noted, mental conditions are not static; they can evolve and change over time. These changes are crucial in understanding how mental health issues are diagnosed, treated, and managed. The same individual may exhibit distress at different points in their life, and how that distress is interpreted can vary depending on factors such as their stage of life, current life circumstances, or external stressors they may be experiencing at the time. In some cases, it may be that an adolescent showing signs of anxiety is going through developmental changes and that the anxiety they are experiencing is a normal, albeit temporary, response to the challenges of adolescence.

[57] The study by Wittchen, Hans-Ulrich, et al. "The waxing and waning of mental disorders: evaluating the stability of syndromes of mental disorders in the population." Comprehensive psychiatry 41.2 (2000): 122-132. have shown that "the symptoms and syndromes as well as the diagnoses of mental disorders wax and wane over time".

In contrast, if an adult shows similar symptoms of anxiety but struggles with these feelings for several years, this may be an indication of a more persistent problem, such as generalised anxiety disorder or another anxiety disorder.

This distinction between transient and enduring mental states is central to the diagnosis of mental disorders. While some symptoms may appear to be transient, their persistence or recurrence over time can be a strong indicator of an underlying disorder. What may initially seem like a temporary reaction to life events such as the loss of a loved one, a significant job change, or a traumatic experience can eventually develop into a chronic condition that requires intervention. Let me clarify with an example. Imagine a person who, after a major life change such as losing a job, experiences temporary anxiety and difficulty making rational decisions. During this time, it may be difficult for them to maintain their usual level of functioning, and they may be emotionally distressed, finding it challenging to make decisions. This state of unresponsiveness to reason is likely a reaction to acute stress and can be classified as a temporary impairment. While this state can cause significant short-term distress, it is generally seen as a normal adjustment to a major life event. But if the individual persists in having serious trouble controlling their emotions and making logical decisions for a long time, even after getting help or stress-reduction techniques, this could indicate a more serious issue. A mental disorder like major depressive disorder or generalised anxiety disorder may be indicated by persistent challenges with rational decision-making that continue to affect day-to-day functioning and general well-being. To investigate the underlying causes and choose the best course of action in such situations, more thorough and ongoing clinical evaluation would be required.

The temporal dimension also involves understanding how long an individual's unresponsiveness to reasons persists in relation to their system of reasons—what they value and how they interpret their experiences. This requires considering not just the immediate symptoms, but also the broader context of their values, beliefs, and priorities. Take, for instance, someone who temporarily withdraws from work following a personal crisis, such as the death of a close family member. In this case, the individual might take a leave of absence to cope with grief and adjust to the loss. Once the immediate crisis has passed, and the person returns to their usual activities, this temporary impairment is generally seen as a normal response—a "reason-responsive reaction according to their system of reasons"—and not indicative of a mental disorder. This brief period of withdrawal is consistent with a healthy adjustment process.

However, if the withdrawal continues beyond what would be considered a normal period for recovery, and particularly if the individual has responsibilities they value within their system of reasons—such as caring for young children or fulfilling work commitments—this extended period of dysfunction may signal a deeper issue. If, after several months, the person still struggles to meet their obligations and their ability to function effectively in personal and professional domains remains significantly

impaired, this ongoing difficulty in responding to reasons may suggest a more profound mental health issue. For example, it could point to conditions like major depressive episodes or prolonged grief, both of which would benefit from further clinical evaluation.

When assessing mental disorders, therefore, both the changing system of reason of the individual and the wider impact on their health must be considered. This method recognises that people's responses to causes are not set in stone and can vary over time. For example, a depressed person may initially find it difficult to respond to reasons because they feel unworthy. However, their short-term unresponsiveness to reasons should not be immediately categorised as a chronic condition if they later change their viewpoint and resume important activities. The initial impairment can be considered a temporary reaction and not a chronic disorder if the person's capacity to respond to reasons gradually improves, their distress decreases because they adopt a more optimistic attitude, and they resume their functioning activities. Now, the question arises: how should we approach the dynamic nature of a person's system of reasons concerning the temporal dimension? More precisely, how do we determine which values within a person's system of reasons are relevant? To address this question, I will draw on an analogous discussion about reflective equilibrium presented by Baccarini in his 2007 work.

Baccarini (2007) summarises Michael DePaul's discussion[58], stating that DePaul (1987) identifies two possible versions of wide reflective equilibrium: conservative and radical. The conservative version represents the standard position—the method functions as an algorithm used to eliminate conflicts between beliefs while preserving as many accepted beliefs as possible. In principle, the only reason to change an initial belief is to resolve inconsistencies, meaning that the conservative version of wide reflective equilibrium maintains the individual's original moral stance. On the other hand, the radical version of wide reflective equilibrium allows for changes in beliefs even if the reasons for these changes are independent of the need to establish coherence among them. The researcher is permitted to change their mind; in other words, moral transformation is accepted. The method guides the researcher to develop their inquiry under the most favourable conditions for moral judgments—conditions that provide the richest inputs, extensive opportunities to identify and correct inconsistencies in moral judgments, the chance to interact with other reasoning faculties, and, perhaps most importantly, the opportunity to mature. While the goal of the conservative version is merely coherence among initial beliefs, the radical version aims to improve the researcher's moral judgments. The conservative approach seeks only to systematise our moral beliefs, whereas the radical approach aspires to enhance our epistemological standing, making it possible for individuals who begin with

[58] For more details, see M. DePaul, *Two Conceptions of Coherence Methods in Ethics*, *Mind*, 1987, pp. 466–467.

entirely flawed moral beliefs to reach valid conclusions. DePaul argues that the second strategy can be justified in a straightforward way—by highlighting the absurd implication of the first: that at the very outset of moral inquiry, we already possess fully developed cognitive abilities. Furthermore, the radical version of wide reflective equilibrium is the only one that takes seriously the element distinguishing wide reflective equilibrium from narrow reflective equilibrium—the necessity of considering all alternative theories in the process of achieving reflective equilibrium. For the conservative wide reflective equilibrium, this merely means that a person should accept the theory that best aligns with their considered moral judgments. However, the radical version acknowledges that a new theory can provide grounds for a profound shift in moral perspective, for example, by demonstrating that most of one's considered moral judgments were mistaken (Baccarini 2007: 53–54).

The point is that if we evaluate a person's *system of reasons* solely based on their originally existing beliefs and reasons—seeking only for coherence among them—then if those beliefs are flawed or "rotten," they will remain so indefinitely. However, if we take into consideration the advantage of life experiences, we can reshape our system of reasons. For instance, imagine a person who holds racist beliefs. If we assess their moral reasoning solely based on this aspect without considering their life experiences, we miss the potential for growth. Suppose this individual is exposed to artwork such as *American Generation X* or has positive real-world interactions with people of colour. These experiences could lead them to gradually revise their beliefs, challenging their initial biases. Over time, these encounters might cause them to question their racism, integrate new perspectives, and ultimately transform their worldview. However, it is necessary to emphasise that life experiences can also be *regressive*, leading someone to adopt more prejudiced views rather than overcoming them. Consider a person who initially has an open-minded and humanitarian outlook, believing in helping all people regardless of race. Suppose this person volunteers as a doctor in a community where the population belongs to a different racial or ethnic group. If they encounter hostility, cultural misunderstandings, or negative interactions, they might start to generalise these experiences and develop racist attitudes. This kind of negative experience could reinforce stereotypes rather than dismantle them. This is where *critical reflection* becomes essential. Life experiences alone are not enough; they must be synthesised with thoughtful self-examination. The open-minded doctor in the second example should pause and ask themselves: *Do I really want to become racist because of a single bad experience? Am I considering the broader circumstances?* A reflective individual would recognise that one negative encounter does not define an entire group of people and would strive to avoid allowing such experiences to distort their moral reasoning.

Thus, the key argument is that in a system of reasons, the most important reasons and values are those that an individual retains after thoughtful reflection. Beliefs and attitudes should not be accepted or rejected purely based on coherence with existing

views or single experiences. Instead, they should be subjected to critical examination, ensuring that they are shaped by both life experiences and rational reflection. This approach allows for moral and intellectual growth rather than stagnation or regression. It is necessary to clarify and connect this epistemological part with the question of when a person is in a condition of mental disorder. What DePaul's discussion[59] demonstrated is that a system of reasons can be in evolution, meaning that it is not epistemologically legitimate to only change through a coherentist method, but rather, a person, through their life experiences, can come to new knowledge. It has also been shown that a person should not succumb to every change that arises from life circumstances but should critically examine them. The same applies to psychiatry. For example, a person may experience negative events and, as a result, develop a new perspective – such as extreme pessimism or resignation. However, the person should also take a reflective stance to reconsider whether these changes are acceptable to them. In other words, they should ask themselves whether it truly aligns with their values to be resigned? Do my life experiences otherwise confirm the attitude to resign, or is this just a result of a particular life experience? Do I know any people who were resigned but managed to overcome it? However, if, after reflection, the person finds they are unable to cope with this change, they have undergone a revision of their system of reasons. Imagine an entrepreneur for whom their career was the most important thing, but after the collapse of their career, they realized that they actually wanted to devote themselves to their family and lead a quieter life. A revision has occurred.

In conclusion, this chapter has explored the crucial role of the temporal dimension in understanding unresponsiveness to reasons within the context of mental disorders. By highlighting the dynamic nature of a person's system of reasons, I have argued that mental states cannot be fully assessed by examining symptoms in isolation; rather, they must be understood in light of an individual's evolving life story and values. The shift in a person's priorities over time, as illustrated through Hecuba's tragedy, demonstrates that one's reasons and actions are not static but are deeply shaped by personal history and circumstances. I have also discussed the limitations of classical approaches that focus solely on coherence among beliefs and symptoms, advocating instead for a more holistic perspective. This involves recognizing the diversity of individual systems of reasons and considering both their internal dynamics and external influences. Central to this view is the idea that a person's values are not fixed, but subject to change—sometimes impulsively or temporarily—requiring thoughtful reflection to determine their relevance in any given context.

Following Baccarini (2007), who draws on Michael DePaul's dynamic wide reflective equilibrium, I have shown that individuals can refine their reasoning

[59] For more about DePaul's discussion see: DePaul, Michael R. *Balance and refinement: Beyond coherence methods of moral inquiry*. Routledge, 2006.

through critical reflection and life experiences, which enables growth and change. It is not enough to evaluate someone's mental state based on static beliefs or isolated experiences; rather, we must account for their capacity for change, their reflective judgment, and the context in which they form their reasons. A more complex and compassionate approach to psychiatry and mental health is ultimately provided by a more nuanced understanding of unresponsiveness to reasons, one that acknowledges the fluidity of human values and places an emphasis on both autonomy and the potential for development and recovery.

5. CHAPTER FIVE: APPLICATION OF THE WEAK EXTERNALIST JUSTIFICATION TO MENTAL DISORDERS

In the first section of the final chaper, I examine the application of the weak externalist model of justification to two different but related mental health phenomena: suicide and impostor syndrome. Both phenomena pose a particular challenge to traditional understandings of rationality and require a more nuanced approach that takes into account the complex interplay between personal reasoning and external influences. I will argue that the proposed model provides a valuable lens through which to examine how mental conditions, which are deeply rooted in emotional, existential and contextual factors, affects an individual's capacity for reason-responsiveness. The illustration of suicide is relevant in the present discussion, because it is associated with mental disorders. According to debates, it has been both argued that it is a symptom of a disorder (such as depression), or that it may be considered a disorder itself, as argued by Maung (2022).

The first part of the section delves into the case of suicide, drawing on Christopher Cowley's (2006) analysis to highlight the limitations of conventional rational frameworks. Cowley's work illuminates the emotional and existential dimensions of suicide, which often challenge purely rational evaluation. By applying a weak externalist model of justification to detect unresponsiveness to reasons, a deeper understanding of how personal suffering and external contextual factors interact to influence suicidal ideation will be gained, offering a more sympathetic approach to this tragic phenomenon.

The second part of the section shifts to impostor syndrome, a psychological condition characterized by persistent feelings of inadequacy despite clear evidence of competence. Drawing on the insights of Katherine Hawley (2019), I show how the weak externalist model of justification, inspired by the weak externalist epistemology of Gaus (2011), provides a nuanced framework for evaluating the rationality of impostor feelings. This approach considers the role of social, cultural and temporal factors in shaping self-perceptions, allowing for a more complete understanding of the rationality behind these self-doubts and how they can be addressed.

By applying weak externalist model of justification to both suicide and impostor syndrome, this chapter underscores the importance of recognizing the complexities of individual reasoning in the context of mental health. Through this lens, I aim to show that proposed model offers more flexible and context-sensitive approach to understanding mental disorders, paving the way for more effective interventions and support.

5.1. Section One: The Rationality of Suicide

In this setion, I examine how the weak externalist model of justification sheds light on the complexity of mental states, especially as they relate to suicide. Suicide has a strong emotional and existential component that goes against conventional ideas of

logical assessment. We can better grasp how psychological difficulties impact rationality by analysing the complex interactions between individual reasoning and contextual factors by using the weak externalist model. This approach expands our understanding of mental disorders and provides a basis for extending the model to other conditions. To ground this discussion, I draw on Cowley's (2006) analysis of suicide, which highlights the limits of conventional rationality in addressing the emotional and existential dimensions of such acts. I will argue that the proposed model of justification in the context of suicide provides a more nuanced lens for interpreting and dealing with the complexities of mental health.

Rationality in the context of suicide requires careful definition. As discussed in the first part of the dissertation, traditional notions of rationality involve deliberation and the alignment of beliefs and actions with available reasons. Principles such as self-interest, coherence, and foresight typically guide this process. However, the existential gravity of suicide complicates these frameworks. Unlike everyday decisions, suicide eliminates all future experience, making cost-benefit analyses and considerations of long-term outcomes inadequate (Cowley 2006). This emphasises the need to rethink the way we use rationality to understand suicide, because traditional ideas often miss the profound emotional and existential dimensions involved. The morality and rationality of suicide have long been topics of discussion among philosophers. It was seen through the prism of virtue by Plato and Aristotle in ancient times. Plato addressed suicide and the immortality of the soul in the *Phaedo*. According to Aristotle's *Nicomachean Ethics*, suicide is immoral since it undermines society and contradicts leading a decent, moral life. Opinions changed later, during the Enlightenment. In his *Groundwork of the Metaphysics of Morals*, Kant strongly opposed suicide, arguing that it betrays the duty we have to humanity and to ourselves. Hume, on the other hand, supported suicide by emphasising personal freedom and dismissing religious objections to it. By taking a different approach and shifting the conversation from morality to existential despair, Schopenhauer saw suicide as a response to the inevitable suffering of life. Today, however, the question is no longer whether suicide is moral, but whether it is rational.[60] Cowley's (2006) work exemplifies this shift, highlighting the limitations of traditional frameworks in understanding the emotional and existential struggles preceding suicide.

Societal and cultural contexts also shape perceptions of suicide and influence individual reasoning. Historically, attitudes toward suicide have ranged from moral condemnation in religious doctrines to romanticisation in 19th-century literature, such as Goethe's *The Sorrows of Young Werther*.[61] Contemporary viewpoints frequently associate mental disorders with suicide, emphasising psychological problems and depression while ignoring socioeconomic or cultural causes. Such limited framing

[60] https://plato.stanford.edu/entries/suicide/ Accessed on 10.03.2025.

[61] https://plato.stanford.edu/entries/suicide/ Accessed on 10.03.2025.

undermines our comprehension of the larger environment in which suicidal thoughts can occur and obscures a variety of causes, such as societal constraints, chronic pain, or financial hardship. Isolation can be made worse by the shame associated with suicide, especially in societies where it is frowned upon, as is the case in orthodox religious communities. On the other hand, some cultures—like Japan, which has a long-standing tradition of *seppuku*—may implicitly encourage some types of suicide. In the past, seppuku was regarded as a noble deed in Japan, strongly linked to the samurai's principles of unselfishness, responsibility, and devotion. The social and cultural fabric of the era was strongly influenced by the ritualised practice of seppuku, in which samurai willingly killed themselves to protect their dignity or avoid dishonour. Zen Buddhism, which assisted the samurai in overcoming their fear of death and viewing it as an affirmation of life's purpose rather than an act of nihilism, strengthened this cultural framework. However, when Japan entered periods of peace and modernisation, the practice of seppuku changed throughout time and became less associated with the warrior code and more associated with a variety of causes, including shame, sadness, protest, or despair. The cultural worship of suicide continued even after seppuku was outlawed in the late 19th century, particularly during periods of national catastrophe like World War II and the kamikaze bombings. Even though seppuku has decreased in frequency in contemporary Japan, suicide is still traditionally perceived through the prism of honour and responsibility, particularly in situations like *inseki jisatsu* (suicide of responsibility) (Pierre 2015). These cultural differences highlight the importance of understanding the influence of social context on individual decision-making, as attitudes towards suicide are shaped not only by mental health issues but also by deep-rooted cultural, historical and social factors. The weak externalist model of justification accounts for these dynamics by recognising the interplay between personal reasoning and external influences. For instance, systemic inequalities like poverty or discrimination can lead to despair that distorts reasoning. Someone experiencing workplace discrimination may internalise feelings of inadequacy, even when external factors are the root cause. Similarly, social isolation among elderly individuals in Western societies underscores how contextual factors shape reasoning. By addressing both internal reasoning and external circumstances, the weak externalist model of justification offers a nuanced framework for understanding and responding to suicide.

The weak externalist model of justification, which acknowledges an individual’s system of reasons while taking into account external contextual influences, is in line with Cowley's (2006) critique of the insufficiency of purely rational evaluations to capture the depth of such struggles. He advocates for a broader perspective that incorporates philosophical, psychological, and existential factors. This dual emphasis is especially helpful in comprehending how a person's responsiveness to reasons is impacted by both emotional states and outside conditions. For example, as Cowley (2006) states emotional reactions such as horror and pity, often seen as secondary to rational judgement, reflect a deeper sensitivity to the tragedy of suicide. These

emotions capture the existential weight of the act in ways that purely rational frameworks cannot (Cowley 2006). For individuals with suicidal thoughts, personal reasons often stem from emotions, past traumas, or despair and may not align with external rationality, shaped by societal norms or public concerns. The weak externalist model accommodates this divergence, recognising personal reasons while allowing them to be challenged or overridden by broader considerations, such as concerns about the value of life or health policy imperatives.

As mentioned in the previous chapter on the role of temporal relevance in relation to an individual's reason-responsiveness, temporal factors are also crucial in understanding the rationality of suicide, as a person's past experiences, present circumstances and future prospects significantly influence their reasoning processes. Feelings of hopelessness or despair can distort temporal judgement and lead people to overemphasise immediate suffering while undervaluing the potential for future improvement. Take, for example, the case of Nikolina, who is overwhelmed by her circumstances. She has recently experienced several devastating personal tragedies, including the death of a close family member and the loss of her job. Her deep sense of loss has led her to believe that her suffering is unbearable and irreparable. This perspective produces a generalised pessimism about the future that overshadows any potential for improvement. In assessing Nikolina's situation, weak externalist justification provides a valuable framework. This approach recognises that her feelings and beliefs are deeply personal and based on her lived experiences but also acknowledges the influence of wider contextual factors. This allows us to evaluate whether their reasons are justified in their current context or whether they could be challenged or overridden by external considerations. In contrast to more rigid views that would immediately categorise Nikolina's condition as a mental disorder based solely on her sadness and suicidal thoughts, the weak externalist justification promotes a nuanced understanding of her reasoning. For example, if Nikolina's beliefs about her circumstances are supported by valid evidence — such as a long period of severe suffering, a lack of effective social or institutional support, or concrete barriers to recovery — her reasoning could be considered a rational response to her situation. If, on the other hand, there are accessible forms of support, such as family carers, professional therapy or community resources, and she does not acknowledge or avail herself of these options, her reasoning may reflect a more complex problem that requires intervention. Weak externalist justification also helps to distinguish between cases where a person's response to reasons is contextually justified and those where it is distorted by internal or external factors. Nikolina's deep despair may appear at first glance to be an extreme reaction, but when viewed through the lens of weak externalism, it could be understood as an appropriate response to her immediate context. This framework respects the complexity of individual experience and allows for a compassionate and multi-layered assessment of reasoning processes. By applying this approach to cases such as Nikolina's, we can better distinguish between cases in which suicidal ideation is based on justifiable responses to adverse

circumstances and those in which it results from distorted reasoning. This distinction is crucial in providing appropriate support, as it ensures that interventions address not only the person's internal state but also the external factors contributing to their distress.

Because it provides a more flexible and nuanced approach, the weak externalist model of justification is also more accommodating of pluralism and various viewpoints when evaluating rationality than the strong externalism type. Strong externalism frequently ignores the subjective and complex character of human reasoning in favour of strict external standards to determine rationality. In contrast, weak externalism recognises this complexity and permits more flexibility, which is crucial in situations like Nikolina's, where contextual elements and personal reasons are deeply interconnected. Weak externalism avoids the drawbacks of prescriptive methods that miss the nuances of real experience by recognising the plurality of individual's systems of reason.

In Nikolina's case, for example, strong externalism might disregard the personal significance of her emotional state and impose a uniform standard that does not consider her subjective experience. In contrast, weak externalist model of justification respects her unique system of reasons while evaluating the extent to which her reasoning aligns with or diverges from external factors such as available support systems or alternative solutions.

In conclusion, the weak externalist model of justification ensures that rationality is assessed in a way that recognises the lived experiences of individuals and provides a more effective framework for understanding complex situations. By integrating personal and contextual dimensions, weak externalism provides valuable guidance for addressing issues such as suicidal ideation and developing interventions that respect the individuality of those affected.

5.2. Section Two: Impostor Syndrome

In this section, I apply the weak externalist model of justification to examine impostor syndrome, a phenomenon characterised by persistent feelings of fraud and inadequacy despite apparent success. I'll proceed as follows: I'll describe the imposter syndrome first, then analyse it using the weak externalist model of justification and look at social and cultural factors. The discussion draws on the findings of Hawley (2019) in her work "What is impostor syndrome?" as well as illustrative cases such as that of a person named Slavica to show how the weak externalist model of justification provides a nuanced understanding of the rationality of impostor syndrome.

Let's start with a more precise definition of impostor syndrome. Impostor syndrome refers to the pervasive feeling of not deserving one's achievements, accompanied by a strong fear of being exposed as a fraud. As Hawley (2019) explains, these feelings often persist despite clear evidence of competence, which distinguishes impostor syndrome from occasional self-doubt. It manifests itself in the form of a fear of

failure, a reluctance to internalise success and a constant worry that others will expose one's perceived inadequacy. While many people feel like an impostor temporarily, impostor syndrome becomes problematic when these feelings become chronic and significantly interfere with daily functioning, self-esteem and career advancement. Hawley (2019) also highlights that imposter syndrome disproportionately affects individuals from marginalised groups — women, people of colour and LGBTQ+ individuals — who face systemic challenges such as prejudice, discrimination and lack of representation. These external pressures intensify personal feelings of inadequacy, highlighting the dual influence of internal experiences and societal factors and making impostor syndrome a compelling argument for the application of a weak externalist model of justification. Applying it to impostor syndrome involves examining whether feelings of fraudulence and inadequacy are reason-responsive or distortions of reason-responsiveness. Weak externalist model of justification, as previously discussed, acknowledges the individual's system of reasons while recognising the role of external influences. In this framework, impostor syndrome can be seen as either a justified response to certain contextual factors or as a departure from reason-responsiveness requiring intervention.

For example, consider Slavica, a successful software engineer who constantly doubts her competence despite receiving praises and promotions. Weak externalist model of justification allows us to explore whether Slavica's feelings reflect a reasonable reaction to external conditions, such as systemic biases in her workplace, or whether they indicate an inability to appropriately respond to evidence of her competence. If Slavica's workplace fosters a culture of hyper-criticism, competition, or implicit bias, her feelings of inadequacy might be rational responses to these external factors. Additionally, if Slavica belongs to a marginalised group within her industry, societal stereotypes and a lack of representation might contribute to her impostor feelings. In such contexts, her internal doubts align with external realities, making her reasoning justified. On the other hand, if Slavica continues to feel like an impostor despite receiving consistent positive feedback and supportive mentorship, one might assume that her feelings are not entirely reason responsive. In such cases, her doubts could stem from internal insecurities rather than an accurate assessment of her situation. Recognising this distinction helps clarify when impostor syndrome is a rational reaction to contextual factors and when it reflects deeper psychological challenges requiring support.

As mentioned above, societal and cultural pressures play a pivotal role in shaping impostor syndrome. Historical and systemic inequities often create environments where certain groups are more likely to experience impostor feelings. Women in male-dominated professions, for instance, frequently encounter implicit biases that undermine their confidence. Similarly, people of colour may face microaggressions and stereotypes that reinforce feelings of inadequacy. These external pressures can intensify internal doubts, further influencing how individuals assess their own

abilities. Hawley (2019) highlights how societal narratives, such as the idealisation of perfection or meritocracy, further compound impostor syndrome. These narratives often set unrealistic expectations, making individuals feel that their achievements are insufficient or undeserved. Weak externalist model of justification addresses these dynamics by recognising the interplay between personal reasoning and external influences. It challenges reductive views that attribute impostor syndrome solely to internal dysfunction, instead highlighting the role of societal and cultural factors in distorting self-perception and reasoning.

Temporal factors also play a significant role in understanding impostor syndrome. An individual's perception of their past achievements, current circumstances, and future potential can influence the persistence and intensity of impostor feelings. Weak externalism's sensitivity to temporal relevance and dynamic system of reasons offers valuable insights into this phenomenon. For instance, individuals with impostor syndrome often downplay their past successes, attributing them to luck or external factors rather than their abilities. This distorted view undermines their confidence in future endeavours, creating a cycle of self-doubt. Social and professional environments that prioritise immediate results or reinforce unattainable standards can further skew temporal judgment, intensifying the syndrome. In Slavica's case, if her workplace consistently devalues past accomplishments or imposes unrealistic expectations for future performance, her impostor feelings might be rational responses to these pressures. However, if she dismisses evidence of her competence despite a supportive environment, her reasoning may be less aligned with contextual realities. Weak externalism helps to distinguish between these scenarios and provides a framework for considering the temporal aspects of impostor syndrome.

By applying the weak externalist model of justification, it becomes possible to analyse the rationality of the impostor syndrome in a more differentiated way. This approach respects the subjective nature of individual reasoning while taking into account the influence of external and temporal factors. Through the lens of weak externalism, feelings of deception and inadequacy can be understood either as justified responses to external pressures or as deviations from reason-responsiveness that require intervention.

Hawley's analysis emphasises the importance of considering both personal struggles and systemic influences when assessing impostor syndrome. The weak externalist model of justification provides a balanced framework for this task, guiding both philosophical enquiry and practical interventions aimed at alleviating the effects of impostor syndrome on individuals and society. In the next section, I will apply this model to depression to further illustrate the application of the weak externalist model of justification in understanding mental states.

5.3. Section Three: Application of the model to depression [62]

To further illustrate how the weak externalist model of justification works, I will provide specific examples following the discussion of cases about suicide and imposter syndrome. By examining these examples, I hope to illustrate how the models can be applied in practise. I begin with the example of depression, drawing on Graham's (2013) distinction between cases where depression is considered a medical disorder and those where it is not, based on the individual's responsiveness to reasons.

As mentioned in the introduction to the proposed model of weak externalist justification as a solution to Szasz's challenge, I agree with the definition of mental disorders in which one of the key indicators of a disorder is an impairment of a person's capacity to respond to reasons (as argued by authors such as Graham (2013) and Dembić (2023)). This means that when depression (or a condition exhibiting all the outward symptoms of depression) is a reason-responsive reaction to certain events and facts, it cannot be classified as a clinical case of mental disorder, as the person's capacity to respond to reason remains intact. Graham (2013) cites St Augustine as an example of the case where the person is not in a state of mental disorder, i.e. is responsive to reasons.[63]

Graham (2013) argues that Augustine's experience is not one of clinical depression, but of philosophical despair as defined by Richard Garrett (see Garrett 1994: 74). According to Graham, Augustine's depression was due to philosophical reasons that were initially convincing and not refuted. Unlike someone with a mental disorder, Augustine's reactions were justified within his own system of reasons, which comprised beliefs based on the facts available to him, uncontested values and sound deliberation. Augustine's depression was thus weakly externally justified by his system of reasons. Augustine was weakly externalistically responsive to reasons because he overcame his fear of the meaninglessness of life by changing his system of reasons through a religious-spiritual transformation. By finding a higher purpose in faith in the Christian God, Augustine's philosophical reflection led to a significant change in his mental state. In other words, he reacted to the depression by adapting his reasoning, which led to a coherent change in his attitude. Graham (2013) explains that Augustine's situation is different from that of someone who is unhappy because of Addison's disease[64]. Augustine's unhappiness was associated with a deep search

[62] Depression is discussed in more detail in a co-authored paper with Shane Glackin and Elvio Baccarini, which is a product of the joint HRZZ project JOPS.

[63]This chapter does not aim to provide a historical account of Augustine. Instead, it focuses on the exploration of the figure as presented by Graham (2013). Furthermore, the examples discussed were developed in collaboration with the JOPS project, including presented works with my supervisor Elvio Baccarini.

[64] According to Chakera and Vaidya (2010), Addison's disease is defined as a rare, chronic endocrine disorder characterised by inadequate production of adrenal hormones due to destruction or dysfunction

for meaning and a fear that life was not worth living, whereas Addison's disease causes unhappiness through neurochemical changes that are not influenced by reasoning or arguments. Similarly, deep sadness due to events such as grief, which is a reasonable response to personal loss, is not considered a disorder. Augustine's grief over the death of a friend is an example of this. His emotional attachment and awareness of the loss made his grief a justified response based on his own belief and value system. Augustine's condition was therefore weakly externally justified and showed that he was reacting to his own reflections and emotional context (Graham: 2013).[65]

As mentioned in previous chapters, I argue that the weak externalist model of justification is a crucial addition to ensure that the resulting diagnoses adequately account for value pluralism among individuals. With the weak externalist epistemology model of justification, psychiatry can respect a person's perspective and avoid epistemic injustice while still diagnosing a mental disorder that the person may not accept. Epistemic injustice, as Miranda Fricker defines it, occurs when someone is wronged specifically in their capacity as a knower, often due to prejudice or a failure to take their testimony seriously (Fricker, 2007). For example, in the context of mental disorder, a person experiencing atypical thought patterns might be dismissed or discredited because their perspective is seen as irrational, leading to a failure to address their actual needs or understand their experience. This approach involves firstly recognising the person's system of reasons, secondly identifying a disorder as a condition in which the person's capacity to respond to their own reasons is impaired and thirdly ensuring that this impairment is harmful to the person and has a physiological component. In this way, psychiatry can diagnose a mental disorder in a way that respects the individual's own system of reasons.

When a person's system of reasons justifies a profoundly negative state, as in the case of Augustine, that state should be understood as a justified response to their reasons rather than as a disorder. It reflects the individual's own reflections and circumstances and is not inherently pathological. This view allows me to address the criticisms I raised in the section on Graham's (2013) view on recognising the diversity of personal values. As mentioned earlier, Graham's (2013) analysis may not fully capture the extent of values that different individuals hold. I will now provide further examples to demonstrate how this respect for diverse values can be more effectively accounted for through the weak externalist model of justification.

---

of the adrenal glands. The disease often manifests with symptoms such as fatigue, weight loss and hyperpigmentation and requires lifelong hormone replacement therapy for treatment.

[65] The relationship between grief and clinical depression is complex and controversial. Some view grief as a possible precursor to a disorder, others as a disorder in its own right. Graham (2013) suggests that grief, particularly in the case of Augustine, is not necessarily a disorder but a natural response to loss. As the main aim is to illustrate how the weak externalist model operates, I do not focus on resolving these debates, but rather characterise grief as overlapping with depression in a broad conceptual sense.

Consider the cases of Monika and Lovro, who are both deeply distressed about their failed careers. Monika's sense of meaning in life is tied closely to her professional success. Although she has other interests, the loss of her career feels so profound that nothing else can compensate for it. However, her reaction aligns with her system of values and, as such, cannot be classified as a disorder. She is an enthusiastic reader and writer whose life revolves around her passion for literature. When a major publisher rejects her manuscript, she is deeply distressed. This rejection feels like the collapse of her lifelong dreams and identity, leading to a state of deep sadness that affects her ability to maintain her career and personal relationships. Despite this, her reaction is consistent with her system of values and aspirations—literature has always been her central focus, something of fundamental value to her. Her condition reflects her personal reflections and, therefore, cannot be categorised as a disorder. Lovro, on the other hand, values his role as a father more than his professional goals. He is also deeply distressed when an important project he has been working on fails spectacularly. However, this despair leads him to neglect his family and social responsibilities, causing significant problems in these areas. His reaction distracts from his fundamental values, and he struggles to change his condition, despite recognising its impact. Unlike Monika, Lovro neglects his other fundamental values in his system of reasons such as the value of being a father. This despair shows that he is unable to effectively manage his response and sacrifice his most important values. Therefore, Lovro's condition is not adequately justified by his values and is better categorised as a disorder. In these cases, the weak externalist epistemology model of justification is useful because it considers each person's unique values and considerations. It distinguishes between responses that are consistent with a person's values and those that interfere with their most important aspects of life, while respecting individual differences in priorities and experiences.

To accurately diagnose clinical depression, it is also important to recall the point made in the introduction to the model: systems of reasons are not rigid. The mutability of systems of reasons underscores the dynamic way in which individuals form and adapt their beliefs and reactions to life events. However, not every epistemic or emotional reaction can be regarded as a well-founded reason. For a reaction to qualify as a valid reason, it must be underpinned by reflection, careful consideration, and coherence with the individual's broader system of beliefs (as shown in the discussion of DePaul in chapter three). This distinction is critical because it helps differentiate between justified reactions to life events and signs of a potential mental disorder. If a person feels a deep sense of despair or a complete lack of a sense of life, this may be a sign that they are not responding to reasons, which could indicate clinical depression. Such states require careful investigation to determine whether they are due to a genuine change in beliefs or whether they are the manifestation of unresponsive patterns that affect well-being. It is important to respect the autonomy of the individual and not to

make paternalistic judgements, but to have a dialogue with the individual. The aim should be to support them to identify whether their current state reflects a justified new system of beliefs or whether it is an unresponsiveness to reasons that could benefit from further reflection and support. This weak externalist model of justification, combined with the identification of reason-responsiveness as an indicator of a mental disorder, respects personal autonomy while providing a way to understand and potentially resolve the underlying issues.

In conclusion, I would like to summarise what I have done. In this section, I have demonstrated how the weak externalist model of justification provides a nuanced framework for understanding mental disorders by distinguishing between justified responsive states and those indicative of clinical conditions. By analysing cases such as Augustine, Monika and Lovro, I have shown how this model accounts for the diversity of individual values and the mutability of systems of reasons. This approach respects personal autonomy while addressing unresponsiveness to reasons as a key indicator of mental disorders. The argument is that weak externalist model of justification not only helps differentiate between justified and pathological responses but also provides a basis for meaningful psychiatric dialogue that avoids epistemic injustice. It ensures that diagnoses are grounded in the individual's own system of reasons while maintaining the objectivity necessary to identify harmful or unresponsive patterns. The proposed model encourages a collaborative and reflective process that empowers individuals to re-evaluate their perspectives and adapt their systems of reasons, promoting resilience and mental well-being. This conclusion forms the basis for further exploration of how the model can be applied to a wider range of mental health problems to ensure that diverse perspectives are recognised and respected in clinical practise.[66]

5.4. Section Four: Application of the model to anxiety disorders

Building on the previous application of the proposed model to depression, I will now examine its relevance and findings when applied to the example of anxiety disorders. As with depression, the model provides a framework for understanding whether an anxiety response should be categorised as a mental disorder, based on the individual's responsiveness to their own system of reasons. Before illustrating specific cases, I will draw on the findings of a respected psychologist, Jerome Kagan (2017), who has contributed significantly to the understanding of the factors that can lead to anxiety-related problems in individuals. In his research, Kagan (2017) emphasises that individual differences in temperament — such as increased reactivity or heightened sensitivity to environmental stimuli — can predispose some individuals to anxiety disorders. His studies show that children with a temperament characterised by behavioural inhibition (characterised by shyness, caution and avoidance of unfamiliar

[66]bDepression is discussed in more detail in a co-authored paper with Shane Glackin and Elvio Baccarini, which is a product of the joint HRZZ project JOPS.

situations) have a higher risk of developing anxiety disorders later in life. Kagan's longitudinal studies show how early temperament and experiences shape vulnerability to anxiety and that these traits can lead to increased susceptibility to anxiety problems. These studies have shed light on how a person's temperament and early life experiences can significantly influence their susceptibility to anxiety disorders. His findings provide a deeper context for understanding how the weak externalist model of justification can be enriched by considering both biological and environmental factors in anxiety disorders.

Anxiety disorders are often characterised by excessive or irrational fears and worries that interfere with daily life. To classify such conditions as disorders under the weak externalist model, we must determine whether the anxiety represents a reason-responsive reaction or indicates an impairment in the capacity to respond to reasons. To illustrate this, we can draw on temperament, which, according to Kagan, plays a significant role in the development of anxiety. Temperament interacts with an individual's system of values and beliefs, shaping how they interpret and respond to potential threats or challenges. Consider the case of Tanja, a PhD student who experiences severe social anxiety. Her fear of being judged negatively or making mistakes in social settings prevents her from participating in group activities, negatively impacting her academic performance and social life. Tanja's anxiety is closely tied to her values of being accepted and successful in social contexts. On one hand, her anxiety can be seen as a reaction to her deeply rooted values of social acceptance and achievement. However, it becomes problematic when it disrupts her ability to make sense of her life and pursue her goals effectively. While Tanja's anxieties align with her personal values, they also interfere with her overall capacity to respond to reasons that support her well-being and personal aspirations. This unresponsiveness, compounded by the neurological effects of chronic anxiety, suggests that her condition may be classified as an anxiety disorder. Tanja's heightened sensitivity, as suggested by Kagan's findings (2017), could stem from temperamental traits that predispose her to anxiety. If her heightened reactivity is indeed linked to temperament, her anxiety reflects not only her values but also an inherent sensitivity. This distinction helps clarify whether her condition arises from a predisposing temperament or represents an impairment in reason responsiveness. Understanding the role of temperament offers valuable insight into how Tanja's social anxiety develops and persists, as well as its broader implications for her functioning.

Let me examine two further examples of anxiety disorders and their relation to reason responsiveness and temperament. The first case is Milivoj, a middle-aged man who experiences overwhelming anxiety about his health, despite the absence of medical evidence suggesting serious illness. His constant worry about potential health issues leads him to avoid doctor's appointments and excessively control his health, disrupting his daily life and relationships. Milivoj's health anxiety is rooted in personal fears, such as a fear of mortality and a need for security. His anxiety,

however, goes beyond a reasonable response to genuine health concerns and instead manifests itself in a pervasive and irrational fear. This type of attitude prevents him from engaging in everyday activities and maintaining healthy relationships, reflecting an impairment in his capacity to respond appropriately to reasons. Milivoj's condition suggests a neurologically based incapacity to respond to reasons consistent with the characteristics of an anxiety disorder. His heightened concern for his health, while linked to personal fears, may also be influenced by his temperament, which predisposes him to heightened sensitivity. As Kagan's (2017) research highlights, Milivoj's anxiety may stem in part from an inherent temperament that shapes how he perceives and responds to health-related concerns. Incorporating Kagan's findings helps distinguish between an impairment of reason responsiveness and an exacerbation of pre-existing sensitivities.

The second example is Magda, a successful manager who experiences chronic anxiety about various aspects of her life, including work, family, and finances. Her anxiety is pervasive, interfering with her ability to focus, make decisions, and enjoy life. Despite recognising the disproportionate nature of her worries, Magda struggles to control her anxiety effectively. Magda's anxiety is a response to areas of life that are deeply important to her, such as professional success and family well-being. However, it has reached a level that severely impacts her daily functioning and overall quality of life. Her inability to address the underlying reasons, even while acknowledging the irrationality of her anxiety, points to a neurologically based and harmful condition. This aligns with the weak externalist model for diagnosing an anxiety disorder. Kagan's work (2017) on behavioural inhibition provides further insight into Magda's case. Her pervasive anxiety, which affects multiple areas of her life, can be interpreted through the lens of behavioural inhibition. If Magda's temperament includes a high level of behavioural inhibition, her anxiety could reflect an extension of her inherent character traits rather than a mere reaction to specific life events. Thus, it is important to consider that her anxiety may not only stem from her values but also represent a manifestation of her innate sensitivity to stressors. The case of Milivoj and Magda shows how temperament and reason responsiveness can interact in understanding anxiety disorders and offers a nuanced perspective on the origins and effects of such conditions.

Now, let us consider an example in which a person's anxiety, although significant, cannot be categorised as a disorder because it reflects a reason-responsive reaction rather than an impairment of the capacity to respond to reasons. Gabrijela, a talented cellist, was selected to perform in a prestigious concert hall. Although she has always enjoyed playing the cello, the importance of this performance and the high expectations of the audience cause her great anxiety. In the run-up to the event, Gabrijela suffers from symptoms such as nervousness, insomnia and increased stress. Despite these challenges, her anxiety is closely linked to her system of reasons and values, in particular her commitment to deliver an outstanding performance and her

fear of not living up to her own high expectations. Gabrijela's anxiety can be understood as an appropriate and reason-responsive reaction rooted in her personal values. Her desire to shine in the performance and her self-imposed quality standards make her reaction coherent and understandable. Her anxiety is not a mental disorder because it does not reflect a broader impairment in her capacity to manage her life or respond to reasons. Instead, it is context-specific and directly related to the unique stresses of this significant event. Thus, while her anxiety is severe, it does not interfere with her ability to function in other areas of her life. It is time and context specific and arises from the particular circumstances of preparing for the concert. As mentioned, and what is most important, her anxiety is consistent with her values and does not undermine her overall capacity to effectively fulfil her goals and daily tasks. Gabrijela's case demonstrates the importance of distinguishing between justified emotional reactions and clinical mental disorders. Her reaction to the upcoming performance, while challenging, is not indicative of an underlying disorder. Rather, it is a heightened but reason-responsive emotional state in the face of an important event. According to the weak externalist model, her anxiety reflects her personal values and context and does not require a psychiatric diagnosis. In Gabrijela's case, appropriate support might include performance preparation strategies and stress management techniques rather than psychiatric intervention. This approach respects her personal values and recognises the nuanced interplay between her emotions and the meaning she attaches to the event. It emphasises the need for a diagnostic framework that distinguishes between appropriate emotional responses and genuine impairments in reason responsiveness.

In conclusion, in this section I have shown how the weak externalist model provides a nuanced framework for understanding anxiety disorders by assessing whether they are due to an impairment in reason responsiveness or a heightened but reason-aligned reaction to personal values. The vivid examples of Tanja, Milivoj and Magda were used to illustrate the applicability of the model. In these cases, their anxiety impairs their capacity to respond effectively to reasons, interferes with their daily functioning and corresponds to the characteristics of an anxiety disorder. This understanding is enriched by the findings of Kagan (2017) on the role of temperament, who highlights how behavioural inhibitions and innate sensitivities can predispose individuals to anxiety disorders. By integrating these biological and environmental factors, the weak externalist model ensures that diagnoses take into account the complexity of individual vulnerabilities. Conversely, Gabrijela's case demonstrates that rection of anxiety, when proportionate to personal values and particular circumstances, is not a disorder. Her anxiety is a reason-responsive reaction that reflects her high aspirations and commitment to achievement without undermining her overall capacity to manage her life. This approach emphasises the importance of distinguishing between justified emotional reactions and genuine mental disorders. The weak externalist model respects the interplay between personal values, temperament and environmental context and advocates tailored support strategies rather than unnecessary psychiatric

labelling where appropriate. This nuanced perspective promotes an empathetic and accurate understanding of anxiety, avoiding the over-pathologisation of reason-responsive behaviours while effectively addressing genuine impairments.

5.5. Section Five: Application of the model to obsessive-compulsive disorder (OCD)

In applying the weak externalist model to obsessive-compulsive disorder (OCD), it is important to determine whether the condition is a reason-responsive reaction or an impairment in the capacity to respond to reasons. As in the applications before, I argue that the proposed model's framework allows for a nuanced understanding of OCD by examining how well a person's compulsions fit with their system of reasons and whether they interfere with their overall capacity to function effectively.

Obsessive-compulsive disorder is characterised by persistent, intrusive thoughts (obsessions) and repetitive behaviours or mental acts (compulsions) that are performed to relieve the distress caused by these obsessions. Symptoms often include excessive checking, cleaning, or counting and are usually driven by irrational fears or doubts[67]. OCD can significantly interfere with a person's daily life and functioning. The weak externalist model of justification helps to distinguish between cases in which OCD symptoms are consistent with a person's values and those in which the symptoms reflect an impairment in their capacity to respond effectively to reasons.

To illustrate the usefulness of the model, consider the following case. Marijan is a software developer who is very afraid that his computer might be infected with a virus. This anxiety causes him to perform frequent and time-consuming security checks on his devices. Although his behaviour may seem exaggerated, Marijan's actions are in line with his personal values of data security and the protection of sensitive information, as well as his prudent character in general. Marijan's reaction, while extreme, is reason-responsive because it is based on his values and his concern for data integrity. He recognises the importance of security in his professional and personal life, and his obsessive controls, while excessive, are consistent with his value of ensuring the protection of data. Besides, it does not impair his working performance. In this case, Marijan's behaviour, although causing uneasiness, does not reflect an impairment in his overall capacity to respond to reasons. Therefore, his condition cannot be categorised as a mental disorder, but rather as coherent with the values he endorses.

On the other hand, let us consider the case of Eva. Eva is a student who suffers from severe OCD symptoms characterised by obsessive fears of infection. Not only does she ritualistically wash her hands several times a day, which harms her and causes discomfort in social relationships, but also prevents her from any social interaction,

---

[67] https://www.psychiatry.org/patients-families/obsessive-compulsive-disorder/what-is-obsessive-compulsive-disorder Accessed on 10.03.2025.

including attending classes, which severely affects her academic and social life. Despite her negative assessment of the irrational nature of her compulsions and their negative impact on her well-being, Eva finds it impossible to control her behaviour. Her compulsive symptoms reflect an impairment in her capacity to respond to reasons. Eva's fear of contagion is not only exaggerated, but also interferes with her capacity to carry out daily activities and maintain relationships. Her incapacity to control her compulsions and their harmful effects on her functioning indicate that she is unable to effectively manage her response to reasons. This case fulfils the criteria of the weak externalist model for the diagnosis of a mental disorder. Eva no longer responds to her own reasons and values. In addition, there are neurological and psychological components that interfere with her capacity to function in daily life. Her symptoms go beyond a reasoned reaction and represent a significant impairment in her capacity to respond to reasons, suggesting OCD as a mental disorder.

The cases of Marijan and Eva illustrate the difference between compulsive behaviours that are consistent with a person's values and beliefs and those that indicate a broader impairment in their capacity to effectively control their responses. Marijan's compulsive behaviour, while extreme, does not prevent him from performing his job successfully and is consistent with his cautious attitude. While his actions cause some discomfort, they do not interfere with his overall capacity to function in other areas of his life. Therefore, Marijan's condition is considered a non-harmful, reason-responsive reaction and not a mental disorder. In contrast, Eva's case shows how OCD symptoms can develop into a significant impairment. Her inability to control them and their severe impact on her daily life and relationships demonstrate a breakdown in Eva's capacity to respond to reasons. This impairment, combined with other components of her condition, fulfils the criteria for a diagnosis of mental disorder under the weak externalist model.

To summarise, the weak externalist model effectively differentiates between OCD symptoms that are consistent with personal values and those that reflect an impairment in reason responsiveness. By applying this model, we can better understand and diagnose OCD, ensuring that the diagnosis respects individual perspectives while recognising the functional impairments associated with the disorder.

5.6. Section Six: Application of the model to eating disorders

Eating disorders are complex psychological conditions characterised by persistent disturbances in eating behaviour, which can significantly impact a person's physical health and emotional well-being. The most common eating disorders include anorexia nervosa, bulimia nervosa, and binge eating disorder (American Psychiatric Association, 2013[68]). These conditions are often associated with maladaptive eating

[68] https://www.psychiatry.org/patients-families/eating-disorders Accessed on 10.03.2025.

habits and intense concerns about body weight and shape. To explore the relevance of the weak externalist model in the context of eating disorders, I will examine three illustrative cases: one with anorexia nervosa, one with bulimia nervosa and one with binge eating disorder.

Consider the case of Antonia, who suffers from anorexia nervosa. Antonia is an influencer who severely restricts her food intake due to an intense fear of gaining weight. Despite being underweight and experiencing significant physical and psychological distress, she persists with extreme calorie restriction and excessive exercise. Despite the negative impact of her behaviour on her health and the fear of gaining weight being unfounded, Antonia feels compelled to continue these restrictive behaviours and finds it difficult to exert control over her eating habits. Antonia's condition illustrates an impairment in her capacity to respond to reasons. Her extreme dietary restrictions and compulsive exercise are not rationally justified or grounded in her personal values or meaningful goals. Instead, they reflect a profound disconnection from reason-responsiveness. Her eating disorder undermines her ability to maintain her health and effectively engage in daily life. This fits the criteria for a diagnosis of a mental disorder, as Antonia's symptoms signify a severe impairment in her capacity to respond to reasons and manage her well-being in a reason-responsive manner.

Let us look at the next case of Petar, a bodybuilder who follows a strict diet to stay in top shape for his sport. Petar's diet plan includes a carefully controlled calorie intake and rigorous meal planning aligned with his goal of optimising performance. Although his eating habits are very restrictive, they are a rational response to his personal and professional commitment to peak athletic performance. Petar's eating behaviour is consistent with his values and goals, and despite the strict diet, he maintains his overall physical health and functionality. Petar's adherence to a rigid diet plan, while extreme, is not indicative of a mental disorder as defined by the weak externalist model. His eating habits are a reasonable response to his values of achieving athletic success and do not interfere with his capacity to respond to other reasons or maintain his general well-being. Petar's situation is therefore an example of an exaggerated but reason-responsive reaction rather than a pathological condition.

Let us now consider the case of Veronika, who suffers from binge eating disorder (BED). Veronika often eats large amounts of food in a short period of time and feels like she is losing control. She often feels ashamed and guilty about her eating behaviour and tries to control her weight through dieting and excessive exercise, but these strategies are unsuccessful. Although her eating behaviour is excessive and harmful to her health, Veronika finds it difficult to control her cravings and regulate her eating behaviour. Her binge eating episodes and subsequent emotional distress reflect an impairment in her capacity to respond to reasons. Her behaviour is not consistent with her values of maintaining health and well-being; instead, it represents a significant breakdown in her capacity to control her responses to emotional and

situational triggers. Her incapacity to control her eating habits and the pervasive stress caused by her behaviour demonstrate significant functional impairment, suggesting that her condition meets the criteria for a mental disorder.

The application of the weak externalist model to eating disorders highlights the importance of distinguishing between reason-driven responses and impairments of reason responsiveness. In Antonia's case of anorexia nervosa and Veronika's case of binge eating disorder, their symptoms reflect a significant impairment in their capacity to control their responses to reasons, which affects their overall health and functioning. In contrast, Petar's case shows a heightened but reason-responsive reaction that is consistent with his athletic goals. This model helps to distinguish between behaviours that are consistent with personal values and those that are indicative of a pathological condition, providing a nuanced approach to understanding and diagnosing eating disorders. By respecting individual perspectives while recognising functional impairments, the weak externalist model provides a comprehensive framework for treating eating disorders in a way that acknowledges both personal values and clinical needs.

5.7. Section Seven: Conclusion of Chapter Five

In this chapter, I have examined the theoretical applications of the weak externalist model to the understanding and diagnosis of mental disorders. The weak externalist model offers a nuanced approach that balances individual autonomy, reason-responsiveness, and the influence of external factors in the assessment of mental health conditions. By focusing on the individual's capacity to respond to reasons, the model strikes a delicate balance between respecting personal values and recognising impairments in functioning. This chapter began by examining the application of weak externalist justification to two significant mental health phenomena: suicide and impostor syndrome. By examining the rationality of these conditions through the lens of weak externalism, I have demonstrated the value of this framework for a more nuanced understanding of the interaction between individual reasoning and emotional, existential and contextual influences.

The first section explored the rationality of suicide, drawing on the insights of Christopher Cowley (2006) to critique traditional rational frameworks that often neglect the emotional and existential dimensions of this act. Using examples such as Nikolina's case, I have shown how a weak externalist justification respects personal systems of reasoning while evaluating the influence of external and temporal factors. This approach not only challenges rigid evaluations of rationality, but also promotes compassionate and context-sensitive interventions for those experiencing suicidal ideation.

In the second section, weak externalist justification was applied to impostor syndrome. Here, Katherine Hawley's (2019) analysis was used to highlight the interplay between internal doubts and external societal pressures. Using cases such as

Slavica's, I have shown how weak externalism can distinguish between justified responses to systemic prejudice and distorted thinking that requires intervention. This perspective emphasises the importance of considering both individual and structural factors to understand and alleviate impostor syndrome.

The following case studies explored in more detail throughout the chapter emphasise the flexibility of the model and demonstrate how it can be applied to a variety of mental health conditions, including depression, anxiety disorders, obsessive-compulsive disorders and eating disorders. The model encourages a deeper understanding of mental health that honours the complexity of personal experience while taking into account the objective impairments that may hinder a person's ability to respond effectively to reasons.

Looking to the future, there are several promising avenues for further research into the weak externalist model of justification. A key first step is refining diagnostic criteria since mental health issues frequently manifest differently in different cultures and social settings. Development of diagnostic instruments that consider these variances would help the model to be more inclusive, therefore guaranteeing its applicability to a greater spectrum of individuals and settings. While recognising cultural particularities, defining universal markers of impairments in reason-responsiveness will help to make the paradigm fairer and more flexible. Moreover, combining the weak externalist paradigm with neuroscientific results could greatly improve its empirical basis. Learning both the neurological and physiological causes of problems in reason-responsiveness—that is, the brain areas connected to decision-making, self-regulation, and emotional processing—could bring great clarity. Stronger support for the theoretical underpinnings of the paradigm as well as more accurate, evidence-based assessments and treatments would come from this multidisciplinary approach. Moreover, the weak externalist approach can guide the creation of personalised therapy plans. Mental health treatment can get more tailored and successful by matching it with an individual's beliefs, goals, and system of reasons. In addition to being clinically effective, this method could ensure that mental health therapies are profoundly respectful of the patient's identity and agency. Lastly, taking into account the weak externalist model's wider ethical ramifications could change public perceptions and mental health policies. The model's emphasis on diversity and reason-responsiveness may help lessen the stigma associated with mental health and increase public awareness of the complex connection between individual values and mental health. Its use could also guide the distribution of resources, giving priority to therapies that maintain personal liberty while addressing severe impairments.

In summary, the future development of the weak externalist model holds the promise of refining its diagnostic criteria for greater cultural inclusivity, integrating it with neuroscientific research to provide empirical grounding, creating personalised interventions aligned with individual values, and addressing ethical considerations

that shape public policy and societal attitudes toward mental health. These efforts will not only strengthen the model's theoretical foundation but also ensure its practical relevance in advancing mental health care on a global scale. By fostering a more pluralistic approach to mental health, the weak externalist model has the potential to transform clinical practice and contribute to a more compassionate, equitable society.

CONCLUSION OF PART TWO

In the second part of this dissertation, I critically analysed the extent to which psychiatry can develop objective evaluative standards that respect individual autonomy and uphold the principles of freedom and equality. Drawing on a historical and philosophical critique of psychiatric practises, this investigation questions whether the discipline can reconcile objectivity with respect for individuality and personal rights. In a world where psychiatric classifications are often used to categorise and control people deemed 'mentally disordered', it is crucial to examine whether diagnostic standards unintentionally reinforce social inequalities or undermine autonomy. This analysis has explored critical dimensions of understanding, defining, and diagnosing mental disorders, culminating in the development and application of a nuanced framework based on the weak externalist model of justification. Across chapters, I have emphasised the importance of integrating individual autonomy, reason-responsiveness, and external influences into a model that is both ethically defensible and practically applicable.

Chapter 3 of the dissertation outlined the fundamental challenge of defining mental disorders within a framework that avoids oppressive dynamics while respecting individual values. It critically engaged with critiques of Szasz (1960, 1961; 1994; 2000) and Foucault (1989) and the value-laden nature of psychiatric definitions, showing how historical misclassifications such as that of homosexuality reveal the dangers of subjective bias. By evaluating the responses of the Aristotelian and Rawlsian approaches, the chapter laid the groundwork for an alternative model that reconciles objectivity and pluralism. Ultimately, a framework inspired by Gerald Gaus (2011) was proposed that relies on a specific public justification approach (Baccarini and Lekić Barunčić 2023) and a weakly externalist justification model inspired by Gaus'weak externalist epistemology to protect against oppressive classifications while honouring diversity and individual autonomy.

In chapter 4, I emphasised the importance of the temporal dimension for understanding unresponsiveness to reasons in mental disorders. Mental states should be assessed in the context of an individual's evolving life history and values, rather than as isolated symptoms. I have criticised classical approaches that focus exclusively on the coherence of beliefs. I argue in favour of a more holistic perspective that takes into account both internal dynamics and external influences. Drawing on Michael DePaul's dynamic, wide-ranging reflective equilibrium, I have shown how individuals can refine their reasoning through reflection and life

experiences, enabling growth. A nuanced understanding of unresponsiveness to reasons emphasises autonomy and the potential for recovery, offering a more compassionate approach to psychiatry and mental health.

Chapter 5 demonstrated the utility of the framework by applying the weak externalist model of justification to specific mental health phenomena such as suicide and impostor syndrome. The analysis of suicide criticised traditional rationality models and offered a compassionate and context-sensitive alternative that takes into account emotional, existential and temporal factors. Similarly, the study of impostor syndrome highlighted the interplay between systemic biases and personal reasoning and emphasised the ability of the framework to account for structural and individual factors contributing to mental health problems. These discussions emphasised the limitations of internalist models and the value of a weak externalist model of justification for a nuanced understanding of psychological states. The chapter has also enabled the application of the weak externalist model of justification to a broader range of mental disorders such as depression, anxiety disorders, obsessive-compulsive disorders and eating disorders. By focussing on an individual's capacity to respond to reasons, this chapter illustrated the versatility of the model and its ability to balance respect for personal experience with recognition of functional impairment. Detailed case studies were used to demonstrate how the model recognises the complexity of mental disorders and provides a framework that is both rigorous and adaptable.

To summarise, the second part of this dissertation has established the weak externalist model of justification as a robust and solid framework for understanding mental disorders. By integrating temporal, individual and societal factors, this model overcomes the limitations of traditional approaches and provides a comprehensive method for diagnosis and intervention. This framework not only respects the diversity of human experience, but also provides a pathway for equitable and effective mental health care. It ensures that classifications and treatments are fair and objective and take into account the complexity of the human experience.

CONCLUSION

This dissertation has critically explored various philosophical and practical questions in psychiatry and extension regarding individuals who are not reasonable and rational. As I have shown, for concerns of coherence and reasonableness this extension needs to embrace non-human animals, as well. The first part was focused on inclusion in justice. The second part on how to treat fairly individuals with psychiatric symptoms.

The first part engaged deeply with Nussbaum's capabilities approach and contrasted it with Rawls's theory of justice, highlighting the limitations of Rawlsian principles in addressing the needs of individuals unable to engage in rational deliberation. Although Nussbaum's framework offers a compelling alternative, its reliance on a fixed list of capabilities and species-based norms risks excluding diverse perspectives on flourishing. Drawing on theorists such as Badano, Richardson, Freeman and Stark, this dissertation proposed the Ideal Reasonable Agents (IRAs) model as a more inclusive framework. The IRAs model transcends traditional membership-based concepts of justice by ensuring principles are justified for all individuals, including those unable to participate directly in the process. This transition, then, requires a further elaboration of reasoning about justice. This includes a distinction between ideal and real-world justice, due to the extension of individuals included in questions of justice. I have advocated for pragmatic strategies to progressively realise ideals of fairness and dignity for all living beings in real-world.

The second part focused on psychiatry, critically analysing whether it can develop objective evaluative standards that respect individual autonomy and uphold freedom and equality. Building on the critiques of Szasz and Foucault, it examined the potential of psychiatric classifications to reinforce power imbalances and marginalise people who are considered "mentally disordered". In response, a framework inspired by weak externalist epistemology and public justification was proposed. This pluralistic approach respects the diversity of individual reasoning while ensuring consistency and fairness in diagnostic standards. By integrating temporal, societal, and individual factors, this model promotes a nuanced understanding of mental disorders that honours autonomy and avoids oppressive dynamics. Its application to phenomena such as suicide, impostor syndrome and various disorders such as depression, anxiety disorders, obsessive-compulsive disorder and eating disorders has shown that it has the potential to transform psychiatric practise and promote equitable mental health care.

Together, these two parts argue for frameworks of justice and mental health care that are both theoretically robust and practically inclusive. By addressing the inadequacies of existing models, this dissertation advocates for a pluralistic, context-sensitive approach to justice and psychiatry that recognises the diversity of experiences and respects individual dignity. It offers a vision for a more equitable and compassionate society, where justice principles and mental health practices are designed to uphold

the rights and well-being of all individuals—human and non-human alike—while addressing the structural inequalities and resource constraints of the modern world. This work ultimately seeks to contribute to a more just and dignified future for all.

LITERATURE

1. Abbey, Ruth. "Rawlsian Resources for Animal Ethics." *Ethics and the Environment* 12, no. 1 (2007): 1-22.
2. Alvaro, Carlo. "Ethical Veganism, Virtue, and Greatness of the Soul." *Journal of Agricultural and Environmental Ethics* 30, no. 6 (2017): 765-781.
3. Baccarini, Elvio. Moralna spoznaja, *Izdavački centar Rijeka*, Rijeka (2007).
4. Baccarini, Elvio, and Kristina Lekić Barunčić. "Public Justification, Evaluative Standards, and Different Perspectives in the Attribution of Disability." *Philosophies* 8, no. 5 (2023): 87. https://doi.org/10.3390/philosophies8050087.
5. Badano, Gabriele. "Political Liberalism and the Justice Claims of the Disabled: A Reconciliation." *Critical Review of International Social and Political Philosophy* 17, no. 4 (2014): 401-422.
6. Barnes, Elizabeth. *The Minority Body: A Theory of Disability*. Oxford: Oxford University Press, 2016.
7. Batavia, Andrew I. "The New Paternalism: Portraying People with Disabilities as an Oppressed Minority." *Journal of Disability Policy Studies Studies*, *12*(2), 107-113. https://doi.org/10.1177/104420730101200208 (Original work published 2001)
8. Begon, Jessica. *Disability Through the Lens of Justice*. Oxford: Oxford University Press, 2023.
9. Berkey, Brian. "Prospects for an Inclusive Theory of Justice: The Case of Non-Human Animals." *Journal of Applied Philosophy* 34, no. 5 (2017): 679-695. https://doi.org/10.1111/japp.12163
10. Blease, Charlotte. "The Duty to Be Well-Informed: The Case of Depression." *Journal of Medical Ethics* 40, no. 4 (2014): 225-229.
11. Bok, Hilary, 'Keeping Pets', in Tom L. Beauchamp, and R. G. Frey (eds), *The Oxford Handbook of Animal Ethics*, Oxford Handbooks (2011; online edn, Oxford Academic, 1 May 2012), https://doi.org/10.1093/oxfordhb/9780195371963.013.0029, accessed 07.03.2025.
12. Brown, Cheryl, and Stacy Miller. "The Impacts of Local Markets: A Review of Research on Farmers Markets and Community Supported Agriculture (CSA)." *American Journal of Agricultural Economics* 90, no. 5 (2008): 1296-1302.
13. Brownlee, Kimberley. *Conscience and conviction: The case for civil disobedience*. OUP Oxford, 2012.
14. Chakera, Ali J., and Bijay Vaidya. "Addison Disease in Adults: Diagnosis and Management." *The American Journal of Medicine* 123, no. 5 (2010): 409-413.
15. Cisney, Vernon W., and Nicolae Morar, eds. *Biopower: Foucault and Beyond*. Chicago: University of Chicago Press, 2020. Available at: https://cupola.gettysburg.edu/cgi/viewcontent.cgi?article=1090&context=books

16. Claassen, Rutger. "Capability paternalism." *Economics & Philosophy* 30.1 (2014): 57-73.
17. Clark, Judy MacArthur. "The 3Rs in research: a contemporary approach to replacement, reduction and refinement." *British Journal of Nutrition* 120.s1 (2018): S1-S7.
18. Cleary, M., S. West, D. Visentin, M. Phipps, M. Westman, K. Vesk, and R. Kornhaber. "The Unbreakable Bond: The Mental Health Benefits and Challenges of Pet Ownership for People Experiencing Homelessness." *Issues in Mental Health Nursing* 42, no. 8 (2021): 741-746.
19. Clifford, Stacy. "Making Disability Public in Deliberative Democracy." *Contemporary Political Theory* 11, no. 2 (2012): 211-228.
20. Clifford, Stacy. "The Capacity Contract: Locke, Disability, and the Political Exclusion of 'Idiots'." *Politics, Groups, and Identities* 2, no. 1 (2014): 90-103.
21. Cochrane, Alasdair, Robert Garner, and Siobhan O'Sullivan. "Animal Ethics and the Political." *Critical Review of International Social and Political Philosophy* 21, no. 2 (2018): 261-277.
22. Cooper, Rachel. "Disease." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 33, no. 2 (2002): 263-282. Available at: https://raptorresearchfoundation.org/wp-content/uploads/2023/02/Techniques_Manual_Chapter-17.pdf
23. Cowley, Christopher. "Suicide Is Neither Rational Nor Irrational." *Ethical Theory and Moral Practice* 9 (2006): 495-504.
24. Dantzer, Robert, and Pierre Mormède. "Stress in Farm Animals: A Need for Reevaluation." *Journal of Animal Science* 57, no. 1 (1983): 6-18. https://doi.org/10.2527/jas1983.5716.
25. Davis, Lennard J. *The disability studies reader*. Routledge, 2016. https://cyberspacerobinson.org/courses/crit-theory-materials/packet/davis.pdf
26. De Waal, Frans BM, and Angeline van Roosmalen. "Reconciliation and consolation among chimpanzees." *Behavioral Ecology and Sociobiology* 5 (1979): 55-66. https://thebrainissocool.com/wp-content/uploads/2024/03/fransdewaalreconciliationandconsolation.pdf
27. Dembić, Sanja. "Mental Disorder: An Ability-Based View." *Philosophy and the Mind Sciences* 4 (2023).
28. Fernandes, J., Blache, D., Maloney, S. K., Martin, G. B., Venus, B., Walker, F. R., ... & Tilbrook, A. (2019). Addressing animal welfare through collaborative stakeholder networks. *Agriculture*, *9*(6), 132. Available at: https://www.mdpi.com/2077-0472/9/6/132
29. Foot, Philippa. *Natural Goodness*. Oxford: Oxford University Press, (2001).
30. Foucault, Michel. *Madness and Civilization: A History of Insanity in the Age of Reason*. New York: Random House. (1989).
31. Fournier, Angela K., and E. Scott Geller. "Behavior analysis of companion-animal overpopulation: A conceptualization of the problem and suggestions for

intervention." *Behavior and Social Issues* 13 (2004): 51-69. https://link.springer.com/content/pdf/10.5210/bsi.v13i1.35.pdf

32. Fraser, Nancy. "Foucault on Modern Power: Empirical Insights and Normative Confusions." *Praxis International* 1, no. 3 (1981): 272-287. Available at: https://www.academia.edu/86461202/Foucault_on_modern_power_Empirical_insights_and_normative_confusions
33. Freeman, Samuel. "Contractarian Justice and Severe Cognitive Disabilities." In *Disability and Practice*, edited by Thomas Hill and Adam Cureton, 174-203. Oxford: Oxford University Press, 2018.
34. Fricker, Miranda. *Epistemic Injustice: Power & the Ethics of Knowing*. Oxford University Press, 2007.
35. Friedrich, J. *Depression and Critique: Towards a Political Theory of Mental Health*. PhD diss., University of Oxford, 2021. Available at: https://ora.ox.ac.uk/objects/uuid:37aca423-2b4c-4b21-bac7-ad298e13e9b1/files/d2801pg775
36. Garner, Robert. "Animals, Politics and Justice: Rawlsian Liberalism and the Plight of Non-Humans." *Environmental Politics* 12, no. 2 (2003): 3-22. https://doi.org/10.1080/0964401042331308164
37. Garner, Robert. "Rawls, Animals and Justice: New Literature, Same Response." *Res Publica* 18, no. 2 (2012): 145-165. https://link.springer.com/article/10.1007/s11158-011-9173-z
38. Garrett, Richard. "The Problem of Despair." In *Philosophical Psychopathology*, 73-89. 1994.
39. Gaus, Gerald F. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. Oxford: Oxford University Press, 1996.
40. Gaus, Gerald. "A Tale of Two Sets: Public Reason in Equilibrium." *Public Affairs Quarterly* 25, no. 4 (2011): 305-325. https://doi.org/10.2307/40473414.
41. Glackin, Shane N. "Three Aristotelian Accounts of Disease and Disability." *Journal of Applied Philosophy* 33, no. 3 (2016): 311-326. https://doi.org/10.1111/japp.12114.
42. Glannon, Walter. "Psychopathy and Responsibility." *Journal of Applied Philosophy* 14, no. 3 (1997): 263-275. https://doi.org/10.1111/1468-5930.00062 ; Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-5930.00062?casa_token=MdQjItAqihEAAAAA:PVPfnFuOYleZwVIux8P0YUGE2ZJdC9_du1dywTF-GD64_tk-iorz1RI4ZJu7mY33xFpvHyA4QKsG
43. Graham, George. *The Disordered Mind: An Introduction to Philosophy of Mind and Mental Illness*. Routledge, 2013.
44. Hartley, Christie. "Disability and Justice." *Philosophy Compass* 6, no. 2 (2011): 120-132. https://doi.org/10.1111/j.1747-9991.2010.00375.x
45. Hartley, Christie. "Justice for the Disabled: A Contractualist Approach." *Journal of social philosophy* 40.1 (2009).

46. Hartley, Christie. "Two Conceptions of Justice as Reciprocity." *Social Theory and Practice* 40, no. 2 (2014): 409-432.
47. Hawley, Katherine. "I—What Is Impostor Syndrome?" *Aristotelian Society Supplementary Volume* 93, no. 1 (2019): 69-91.
48. Hursthouse, Rosalind. "Virtue ethics and the treatment of animals." (2011). Available at: https://tomwilk.net/wp-content/uploads/2019/08/Virtue-Ethics-and-the-Treatment-of-Animals-Hursthouse.pdf
49. Kagan, Jerome. *Five Constraints on Predicting Behavior*. Cambridge, MA: MIT Press, 2017.
50. Kant, Immanuel. *Critique of Practical Reason*. Indianapolis: Hackett Publishing, 2002.
51. Kant, Immanuel. *Fundamental Principles of the Metaphysic of Morals*. 1785. English translation by Thomas Kingsmill Abbott. New York: Cosimo Classics, 1988.
52. Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Cambridge: Cambridge University Press, 2012.
53. Kittay, Eva Feder. "At the Margins of Moral Personhood." *Ethics* 116, no. 1 (2005a): 100-131.
54. Kittay, Eva Feder. "Equality, Dignity, and Disability." In *Perspectives on Equality: The Second Seamus Heaney Lectures*, edited by Mary Ann Lyons and Fionnuala Waldron, 23-44. Dublin: The Liffey Press, 2005b.
55. Kittay, Eva Feder. "When Caring Is Just and Justice Is Caring: Justice and Mental Retardation." *Public Culture* 13, no. 3 (2001): 557-579. https://doi.org/10.1215/08992363-13-3-557.
56. Kittay, Eva Feder. *Love's Labor: Essays on Women, Equality, and Dependency*. Routledge, 1999.
57. Korsgaard, C. M. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: Oxford University Press, 2018.
58. Kymlicka, Will, and Sue Donaldson. "Animal Rights, Multiculturalism, and the Left." *Journal of Social Philosophy* 45, no. 1 (2014): 116-135. Available at: https://vegstudies.univie.ac.at/fileadmin/user_upload/p_foodethik/Kymlicka__Will__Donaldson__Sue_2014._Animal_Rights__Multiculturalism__and_the_Left._Journal_of_Social_Philosophy.pdf
59. Kymlicka, Will, and Sue Donaldson. *Zoopolis: A Political Theory of Animal Rights*. Oxford: Oxford University Press, 2011.
60. Lisboa HM, Nascimento A, Arruda A, et al. Unlocking the potential of insect-based proteins: Sustainable solutions for global food security and nutrition. *Foods*. 2024;13(12):1846. Available at: https://www.proquest.com/scholarly-journals/unlocking-potential-insect-based-proteins/docview/3072324676/se-2. doi: https://doi.org/10.3390/foods13121846.
61. Lock, Andrew, et al. "Resisting anorexia/bulimia: Foucauldian perspectives in narrative therapy." *British Journal of Guidance & Counselling* 33.3 (2005): 315-

332. Available at: https://www.researchgate.net/profile/David-Epston/publication/241444694_Resisting_anorexiabulimia_Foucauldian_perspectives_in_narrative_therapy/links/561bc09908aea80367242d84/Resisting-anorexia-bulimia-Foucauldian-perspectives-in-narrative-therapy.pdf
62. Martinić, Iva, and Elvio Baccarini. "Capabilities and Justice for People Who Lack the Capacity for Reason and Rationality." *Filozofska istraživanja* 43.3 (2023): 495-507. Available at: https://hrcak.srce.hr/file/455892
63. Martinić, Iva. "Politika i Neljudske Životinje: Držanje Pasa na Lancu." *Političke Analize: Tromjesečnik za Hrvatsku i Međunarodnu Politiku* 10, no. 37 (2021): 32-37.
64. Martinić, Iva. *Specizam: Poremećaj Cjelokupnog Društva Kojim Se Opravdava Iskorištavanje Ne-Ljudskih Životinja*. PhD diss., University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy, 2020.
65. Maung, Hane Htut. "Mental disorder and suicide: what's the connection?." *The journal of medicine and philosophy: A forum for bioethics and philosophy of medicine*. Vol. 47. No. 3. US: Oxford University Press, 2022. Available at: https://academic.oup.com/jmp/article-pdf/47/3/345/45251983/jhab015.pdf
66. McGuire, Jonathan, Robyn Langdon, and Martin Brüne. "Moral Cognition in Schizophrenia." *Cognitive Neuropsychiatry* 19, no. 6 (2014): 495-508. https://doi.org/10.1080/13546805.2014.928195.
67. McMahan, Jeff. "Cognitive Disability, Misfortune, and Justice." *Philosophy & Public Affairs* 25, no. 1 (1996): 3-35.
68. McMahan, Jeff. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press, 2002.
69. Megone, Christopher. "Aristotle's Function Argument and the Concept of Mental Illness." *Philosophy, Psychiatry, & Psychology* 5, no. 3 (1998): 187-201.
70. Megone, Christopher. "Mental Illness, Human Function, and Values." *Philosophy, Psychiatry, & Psychology* 7, no. 1 (2000): 45-65.
71. Melis, Alicia P., Brian Hare, and Michael Tomasello. "Chimpanzees recruit the best collaborators." *Science* 311.5765 (2006a): 1297-1300. Available at: https://science.umd.edu/faculty/wilkinson/BIOL608W/Melis2006Science.pdf
72. Melis, Alicia P., Brian Hare, and Michael Tomasello. "Engineering cooperation in chimpanzees: tolerance constraints on cooperation." *Animal Behaviour* 72.2 (2006b): 275-286. Available at: https://www.eva.mpg.de/documents/Elsevier/Melis_Engineering_AdapBeh_2006_1555101.pdf
73. Mendez, Mario F., Eric Anderson, and Jill S. Shapira. "An Investigation of Moral Judgement in Frontotemporal Dementia." *Cognitive and Behavioral Neurology* 18, no. 4 (2005): 193-197. Available at: https://www.academia.edu/download/82242123/mendez-anderson-shapira-2005-an-investigation-of-moral-judgement-in-frontotemporal-dementia2.pdf

74. Mercer, Gary. "Dignity and Disability: Toward a Relational Approach." *Philosophy and Public Affairs* 45, no. 3 (2017): 189-213. Available at: https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1216&context=philosophy_theses
75. Morell, Virginia. "Minds of Their Own: Animals Are Smarter Than You Think." *National Geographic* 213, no. 3 (2008): 36-61. Available at: http://www.mreroh.com/student/apdocs/NERVOUS/Brain%20-%20Sense/Mind%20of%20their%20Own%20-%20Nat%20Geo.pdf
76. Murray, Rowan. "The capability approach, pedagogic rights and course design: Developing autonomy and reflection through student-led, individually created courses." *Journal of Human Development and Capabilities* 25.1 (2024): 131-150. Available at: https://www.tandfonline.com/doi/pdf/10.1080/19452829.2023.2261856
77. Nussbaum, Martha C. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Belknap Press of Harvard University Press, 2006.
78. Nussbaum, Martha C. *Nature, function and capability: Aristotle on political distribution. Oxford Studies in Ancient Philosophy, Supplementary Volume* (Vol. 6), 145–184. Oxford: Clarendon Press, 1988. Available at: https://changingminds.org/explanations/needs/nussbaum_capabilities.htm
79. Nussbaum, Martha C. *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press, 2000.
80. Oliver, Kelly. "Service Dogs: Between Animal Studies and Disability Studies." In *Disability and Animality*. Routledge (2020): 111-128. Available at: https://www.academia.edu/download/65575728/Disability_and_Animality_Published_Book.pdf#page=130
81. Oosterlaken, Ilse. "Design for development: A capability approach." (2009): 91-102. Available at: https://direct.mit.edu/desi/article-pdf/25/4/91/1714696/desi.2009.25.4.91.pdf
82. Ormandy, Elisabeth H., et al. "Animal research, accountability, openness and public engagement: Report from an international expert forum." *Animals* 9.9 (2019): 622. **https://doi.org/10.3390/ani9090622**
83. Phillips, Mary L., and David J. Kupfer. "Bipolar disorder diagnosis: challenges and future directions." *The Lancet* 381.9878 (2013): 1663-1671.
84. Pierre, Joseph M. "Culturally sanctioned suicide: Euthanasia, seppuku, and terrorist martyrdom." *World journal of psychiatry* 5.1 (2015). URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4369548/
85. Plannthin, Drude-Katrine. "Animal ethics and welfare in the fashion and lifestyle industries." *Green Fashion: Volume 2* (2016): 49-122. https://doi.org/10.1007/978-981-10-0245-8_3
86. Rachels, James. "Darwin, Species, and Morality." *The Monist* 70, no. 1 (1987): 98-113. Available at: https://www.jstor.org/stable/pdf/27903016.pdf?casa_token=0nw0untCWjUAAA

AA:x4TVRZe_Lwpvx2cdjQ6OvKTuBZlN-IrQoLzP0rvm0wgNr6HggQFcKbrlPfQhEhUGG6mSfATYipN6mhTkJZwOyj_pyksXN1Z_8R8f4LxPS_Fk5vh8Sw

87. Rajapakse, Rakhitha. "Redefining Abnormality through a Critical Examination of Societal Influences and Biases in Psychological Diagnoses." *Available at SSRN 4916006* (2024): https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=4916006
88. Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
89. Rawls, John. *A Theory of Justice*. Revised ed. Cambridge, MA: Harvard University Press, 1999.
90. Rawls, John. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press, 2001.
91. Rawls, John. *Political Liberalism*. Columbia University Press, 1995.
92. Rawls, John. *Political Liberalism*. Expanded ed. New York, NY: Columbia University Press, 2005.
93. Regan, Tom. "Duties to Animals: Rawl's Dilemma." *Ethics and Animals* 2, no. 4 (1981): 4. Available at: https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1095&context=ethicsandanimals
94. Regan, Tom. *The Case of Animal Rights.* University of California Press, 1983.
95. Richardson, Henry S. "Rawlsian Social-Contract Theory and the Severely Disabled." *The Journal of Ethics* 10 (2006): 419-462.
96. Robeyns, I. (2017). *Wellbeing, freedom and social justice: The capability approach re-examined*. Open Book Publishers.
97. Robeyns, Ingrid. "The capability approach: a theoretical survey." *Journal of human development* 6.1 (2005): 93-117. Available at: https://www.tandfonline.com/doi/pdf/10.1080/146498805200034266?casa_token=T5nZ23vpX2gAAAAA:FggC8TaqqMXTliPujCd6LJfJzk4vDIdJrvq37mRiUp0JKUFEJilsJwTzpNBQvOr8JAVJBLuFfos
98. Rowlands, Mark. "Contractarianism and Animal Rights." *Journal of Applied Philosophy* 14, no. 3 (1997): 235-247. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-5930.00060?casa_token=FdxaJtU8pjAAAAAA:6oszzC-S2wG17NzMSMq-6cO8UdWtl2HG6IXR9eUvhDyP6-P_AMiLwzLRitvlZfpMjoyE6gjTLg0l
99. Sanders, Clinton R. "Actions Speak Louder Than Words: Close Relationships Between Humans and Nonhuman Animals." *Symbolic Interaction* 26, no. 3 (2003): 405-426. https://doi.org/10.1525/si.2003.26.3.405.
100. Schramme, Thomas. "Capable Deliberators: Towards Inclusion of Minority Minds in Discourse Practices." *Critical Review of International Social and Political Philosophy* (2021): 1-24. Available at: https://www.tandfonline.com/doi/pdf/10.1080/13698230.2021.2020550

101. Schulz-Weidner, Nelly, et al. "Dental Treatment Under General Anesthesia in Pre-School Children and Schoolchildren with Special Healthcare Needs: A Comparative Retrospective Study." *Journal of Clinical Medicine* 11, no. 9 (2022): 2613. https://doi.org/10.3390/jcm11092613.
102. Schuppert, Fabian (2014). *Freedom, Recognition and Non-Domination. Studies in Global Justice*. doi:10.1007/978-94-007-6806-2
103. Sen, Amartya. (2004) 'Capabilities, Lists, and Public Reason: Continuing the Conversation', *Feminist Economics*, 10(3), pp. 77–80. doi: 10.1080/1354570042000315163.
104. Sen Amartya. *The idea of justice*. London: Penguin Books, 2009. Available at: https://www.jsscacs.edu.in/sites/default/files/Files/The_idea_of_justice_Amartya_Sen.pdf
105. Sen, Amartya. "Human rights and capabilities." *Journal of human development* 6.2 (2005): 151-166. Available at: https://www.tandfonline.com/doi/pdf/10.1080/14649880500120491?casa_token=APLf98w5RVIAAAAA:0V6oNqFyrrx1quD3OoXGKnpOu5qODBm-FgCiU9sy71mki19c2M1dzXXZOHM0ydp2zL6n16b2qVg
106. Singer, Peter. "Speciesism and moral status." *Metaphilosophy* 40.3-4 (2009): 567-581. Available at: https://www.jstor.org/stable/pdf/24439802.pdf?casa_token=MrdX8JjszEIAAAAA:pBbormVcVtTZ3cBXZ6WVYMpO6BROvh7kOGDlumoSSPWqtMm-6v6K9g0THDKqIj-MpchS-2AkRT_Kd9xclR6wjKh2m7nQVhOxD5JUiCawkWyz_IiD8g
107. Singer, Peter. *Oslobođenje Životinja*. Zagreb: IBIS Grafika, 1998.
108. Singer, Peter. *Praktična Etika*. KruZak, 2003.
109. Singer, Peter. *The expanding circle*. Oxford: Clarendon Press, 1981.
110. Sinnott-Armstrong, Walter, ed. *Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Vol. 3. MIT Press, 2008.
111. Stark, Cynthia A. "How to Include the Severely Disabled in a Contractarian Theory of Justice." *Social Philosophy and Policy* 24, no. 2 (2007): 142-165. https://doi.org/10.1017/S0265052507070774.
112. Sunstein, Cass R., and Martha C. Nussbaum, eds. *Animal Rights: Current Debates and New Directions*. Oxford University Press, 2004.
113. Szasz, Thomas S. "Second Commentary on 'Aristotle's Function Argument'." *Philosophy, Psychiatry, & Psychology* 7, no. 1 (2000): 3-16.
114. Szasz, Thomas. "Mental Illness Is Still a Myth." *Society* 31, no. 4 (1994): 34-39.
115. Szasz, Thomas. "The Myth of Mental Illness." *American Psychologist* 15, no. 2 (1960): 113-118. Available at: https://www.narcissisticabuserehab.com/wp-content/uploads/2022/03/TheMythOfMentalIllnessThomasSzasz.pdf

116. Taylor, Sunaura. *Beasts of Burden: Animal and Disability Liberation*. The New Press, 2017. Available at: http://edelweiss-assets.abovethetreeline.com/THEN/supplemental/Beasts%20of%20Burden%20excerpt.pdf
117. Watnick, Valerie J. "The Business and Ethics of Laying Hens: California's Groundbreaking Law Goes into Effect on Animal Confinement." *BC Envtl. Aff. L. Rev.* 43 (2016): 45. Available at: https://core.ac.uk/download/pdf/71468023.pdf
118. Weiner, Talia. "The (un) managed self: Paradoxical forms of agency in self-management of bipolar disorder." *Culture, Medicine, and Psychiatry* 35 (2011): 448-483. Available at: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/s11013-011-9231-1&casa_token=uxtvheRNu_UAAAAA:jZV7R2_EfDWTSUurGcXkX1hvwT9OaWol10bbC5H-rOCiW0i32jvigc3CdDHzYkqjvTBlpMCa4HWZULg
119. Wilkinson, Greg. "Political Dissent and 'Sluggish' Schizophrenia in the Soviet Union." *British Medical Journal (Clinical Research Ed.)* 293, no. 6548 (1986): 641-644. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC1341504/pdf/bmjcre..
120. Wittchen, Hans-Ulrich, et al. "The waxing and waning of mental disorders: evaluating the stability of syndromes of mental disorders in the population." Comprehensive psychiatry 41.2 (2000): 122-132. Available at: https://core.ac.uk/download/pdf/236368519.pdf
121. Wolff, Jonathan. *Ethics and Public Policy: A Philosophical Inquiry*. New York: Routledge Press. (2011).

Internet links:

1. https://politicalnotmetaphysical.wordpress.com/2016/07/01/basic-issues-can-rawlsians-offer-a-plausible-account-of-disability-justice/ Accessed on: 08.07.2023.
2. https://medicine.missouri.edu/centers-institutes-labs/health-ethics/faq/personhood
   Accessed on: 03.08.2023.
3. Slayton, Kelly, Alexander Grigorievskiy, and Live Statistics. "DAVID REIMER AND THE GENDER EXPERIMENT." https://wiki2.org/en/David_Reimer Accessed on: 15.08.2024.
4. https://lambdalegal.org/case/brandon-v-richardson-county/ Accesed on 3.3.2025.
5. https://en.wikipedia.org/wiki/Brandon_Teena Accesed on 3.3.2025.
6. https://ascot-meats.com/ethical-meat-consumption-a-guide-to-conscious-eating-habits/ Accessed on 07.03.2025.

7. https://wyss.harvard.edu/technology/human-organs-on-chips/ Accessed on 07.03.2025.
8. https://www.stellamccartney.com/gb/en/sustainability/fur-free-fur.html?srsltid=AfmBOoqtc9UYULQjegQbhi_Er2BgoR4qJhjZn35X3Ag_Et1yN9jyuIvO Accessed from Stella McCartney's official website on 07.03.2025.
9. https://plato.stanford.edu/entries/suicide/ Accessed on 10.03.2025.
10. https://en.wikipedia.org/wiki/Hecuba_(play) Accessed on 10.03.2025.