



# **On The Significance of Reason, Finitude and Vulnerability in Moral Constructivism**

**Francisco Lisboa**

A thesis presented in partial fulfillment of the requirements for the degree of

Master of Philosophy

Supervised by: PhD, Susana Cadilha

**Faculdade de Ciências Sociais e Humanas**

**Universidade NOVA de Lisboa**

**October, 2022**

Research project funded by FCT and attributed by IFILNOVA.

## **Acknowledgements**

*To my beloved mother, who has been endorsing my value from day one.*

*To my loving friends and partner, in whom I always found comfort and support.*

*To my cherished supervisor, for the ever-helpful guidance.*

*To FCT and IFILNOVA, whose funding allowed full dedication to this project.*

*“Your judgement judges you and defines you.”*

*-Jean-Paul Sartre.*

## **Abstract**

The present dissertation is a work of philosophical inquiry on the moral implications of ontological vulnerability and finitude within a moral constructivist framework of ethics and morality. Specifically, the aim is to understand whether these concepts, in conjunction with practical reason, can help ground objective moral norms in a way that is stronger than standard constructivist approaches thus far, where only practical reason is accounted for. The theoretical framework chosen - moral constructivism - must first be elaborated on, analyzing where it stands in the realism/anti-realism debate as a metaethical theory. Establishing this general framework leads me to employ a chronological study of its main contenders, to better understand the intricacies and divisive aspects of each approach. Doing so helps me make better sense of why each of them has been unable to vindicate moral objectivity through the use of practical reason alone, in the case of Kantian constructivism, and the subsequent uprising of Humean constructivism, which abandons the project of vindicating moral objectivity altogether. Accepting that the practical standpoint of moral reasoning is incomplete with pure reason alone, but rejecting that moral objectivity must be forfeit, I propose that finitude and ontological vulnerability, together with practical reason, can ground moral objectivity under a substantive claim to embodied autonomy, establishing a moral principle of non-exploitation of another's vulnerability. Reason grounds moral agency, whereas vulnerability grounds moral subject-ability / "subjectness." From this convergence, morality is created.

### **KEYWORDS:**

Morality, Constructivism, Realism, Anti-Realism, Normativity, Reason, Agency, Vulnerability.

**Note:** I will refer to singular "agent" and "subject" with the usage of "they/them" pronouns.

# Table of Contents

<b>Introduction</b> .....	<b>1</b>
1. The Problem of Morality .....	2
2. Metaethical Moral Constructivist Framework .....	5
<b>Chapter I: Kantian Moral Constructivist Framework</b> .....	<b>8</b>
1. Rawlsian Constructivism .....	9
2. Onora O’Neill’s “Modal” Constructivism .....	17
3. Korsgaard’s Constitutivist Constructivism .....	27
<b>Chapter II: Humean Constructivism</b> .....	<b>37</b>
1. Asserting Anti-Realism against Kantian Constructivism .....	37
2. How To (Try To) Vindicate Objectivity Through Moral Constructivism .....	39
<b>Chapter III: Practical Reason, Finitude and Vulnerability</b> .....	<b>44</b>
1. Finitude and Ontological Vulnerability .....	44
2. Moral Agency and Moral Subjectness .....	48
<b>Conclusion</b> .....	<b>58</b>
<b>Bibliography</b> .....	<b>63</b>

## Introduction

Constructivism as a metaethical theory is generally understood as a project which ought to be anti-realist in kind<sup>1</sup>. It establishes that *value* is something conferred to upon the world by *valuing* creatures and hence it does not exist independently of them. This means that moral values cannot be externally imposed upon or ascertained by rational agents from external sources. Valuing agents are responsible for both creating and attributing value to everything in the world, and this applies to moral values as well. Understood this way, morality becomes the product of creative reasoning, the application of practical reason by reflective rational agents who, faced with a (moral) problem, must reason and reach a (moral) solution.

Against a background of dogmatic moral theories, this characterization of moral constructivism is in one way liberating, since rational agents must now take up an active role in the construction of moral norms, as opposed to a passive stance where external moral facts act as the source of objective moral norms. On the other hand, however, moral constructivism now faces the problem of trying to find a legitimate source for so-called objective moral norms. If they can no longer be inferred from an external source such as God or Nature, what then can ground morality in an objective way? What separates moral reasoning from instrumental practical reasoning? And even within moral reasoning, what stops us from falling into subjectivism or relativism? Is it even possible to reach objective moral norms within constructivism?

My (rather ambitious) claim is that it *might* be. While mainstream definitions of objectivity understand it in opposition to subjectivism, where the former “designates the matter outside the mind”, and the latter “the mental and inner sphere of thought”<sup>2</sup>, Kantian constructivists have sought objective value in an anti-realist setting by resorting to formalist procedures of abstraction from subjective points of view, claiming that an objective value is one which is universally shared. In this sense, it is objective because it is not subjectively mind-dependent. From the general framework identified, I expose central features common to the different approaches. First, there are “standards of correctness” for practical reasoning and second, practical reason is ultimately responsible for moral thinking if one is indeed committed to be coherent with the reasoning for action.

---

<sup>1</sup> See Rawls 1980, Korsgaard 1996, Street 2012, Engstrom 2013.

<sup>2</sup> A thorough analysis on ontological and epistemological objectivity can be found on Paolo Savoia’s *Objectivity: Philosophical Aspects* (2015).

It follows that Kantian constructivists take practical reason to be the foundation of morality and self-sufficient for the determination of objective moral norms<sup>3</sup>. This way, Kantian Constructivism tries to balance the seemingly irreconcilable positions of anti-realism and reflective construction of objective moral norms by resorting only to reason to establish substantive moral values. After studying the works of John Rawls, Onora O’Neill, and Christine Korsgaard and their subsequent “faults.” I will explore Sharon Street’s criticisms of Kantian Constructivism as a whole and her own proposal of Humean Constructivism. This effort, combined with valuable input from scholars such as Carla Bagnoli and Paul Formosa, will ground my argument that, siding with Street, pure practical reasoning seems insufficient to ground objective moral norms whilst rejecting her proposal that morality then comes down to each agent’s substantive set of values. In this fashion I reject subjectivism, while agreeing that moral reasoning cannot be thought of in purely rational terms which do not account for the agents in question, specifically, their limitations. I subscribe to Carla Bagnoli’s proposal that in seeking objective moral norms, moral constructivism must willingly account for the subjects’ embodiment and ontological vulnerability stemming from their finitude. I argue that Reason cannot be self-sufficient for the determination of objective moral norms, despite being a necessary condition for it. Morality cannot be solely grounded in practical reasoning because moral reasoning must address the moral subjects’ finitude. Instead, morality must be grounded in something else, namely our finitude, depicted in our vulnerability and overall contingencies, which make morality a rational necessity, and for this reason, vulnerability must play a significant role in the determination of moral norms, which can indeed be objective.

### **1. The Problem of Morality**

To undergo a study of ethics, or morality, is to partake in an inquiry on one of the most primordial questions of the history of humankind. Essentially, we can encapsulate it in the question of “what is the *right* thing to do?” Of course, this question, despite its lexical simplicity, holds behind it numerous conceptual and philosophical difficulties responsible for the lack of consensus, millennia after the first Greek philosophers like Plato and Aristotle began devoting themselves to it. One of such difficulties is, of course, the definition of “right,” or “good.” To be able to characterize certain actions as such, we

---

<sup>3</sup> I hope to question this self-sufficiency and prove that moral reasoning must also rely on the objectivity of the fact of finitude and ontological vulnerability.

must first come to terms with what makes those actions “right” and “good,” which is to say, to have a clear definition of what the criteria of rightness and goodness is. This invariably forces us to dive into the philosophy of value: is goodness an intrinsic characteristic of things? Can anything *in and by itself be good*? Or is that a characteristic or value *attributed* to it by something (or someone) else? This division leads philosophers in different approaches to morality and what grounds it. Those who opt for the former answer are called “*moral realists*” whereas those who choose the later are deemed “*moral anti-realists*.” If one is a moral realist, then they share an understanding that *there are* moral facts, that things or actions can be *objectively good*, independently of what our subjective understanding and evaluation of them may be. Alternatively, for a moral anti-realist, things have no inherent independent value fixed upon themselves; rather, they require the attribution of value from a third party (for instance us, human beings as valuing agents), then their value is entirely dependent on the considerations of valuing creatures, and therefore subject to considerations and changes in the perception of those who engage in valuing activities. If we follow this line of thought, then it seems that for moral realists, value is a mind-independent property, implying that there are objective moral norms *independent* of the considerations and actions of moral agents. On the other hand, moral anti-realists take moral norms and value to be *dependent* on the considerations and actions of moral agents, and therefore *subjective*, or at least seeking different degrees of epistemological objectivity through intersubjective validity<sup>4</sup>.

Further, there is also disagreement regarding the place of Reason vs. Sentiment in moral theories, leading to tensions “*between intuitionists vs. naturalists; cognitivists vs. non-cognitivists; realists vs. anti-realists.*” (Engstrom 2013, p. 139). According to Engstrom, all these dichotomies can be boxed within the debate between dogmatic rationalism and skeptical empiricism<sup>5</sup>. This discussion pertains to the questions of which element do we base our moral judgments on, and which of them can or should guide our moral actions? Despite some moral constructivists like Carla Bagnoli and Sharon Street opening space to integrate Sentiment in moral reasoning (in the form of a moral feeling of respect for others and the moral law in Bagnoli’s case<sup>6</sup>, and through the impossibility of separation from subjective considerations and substantive set of values from moral

---

<sup>4</sup> See Savoia, 2015.

<sup>5</sup> Because the focus of this work is on moral constructivism, I will not focus any more on this debate other than to say specify that constructivism is meant to be a third approach entirely.

<sup>6</sup> For more on the moral feeling of respect, see Carla Bagnoli’s *Constructivism About Practical Knowledge* (2013).

reasoning in the agent's practical standpoint for Sharon Street<sup>7</sup>), what all moral constructivists can agree on is that the construction of norms entails – or must necessarily entail – *standards of correctness*; in other words, there must be rational criteria of correctness for our moral judgments and considerations. This ties directly with the issue of the cognitive power of rational agents, who can access both theoretical and practical knowledge. For moral constructivists specifically, practical knowledge is of paramount importance and is dependent on cognition as “*spontaneous, a self-consciously active power, in whose exercise diverse concepts are employed together in a single act of judgment*” (Engstrom 2013, p. 144), unlike theoretical knowledge where cognition is *receptive* and *passive* about knowledge of things. Practical knowledge deals with the question of how one ought to act, and as such the object of the judgment is the convergence of both agent and action. Practical knowledge – obtained by cognition in the form of practical reasoning – holds implications for the identity of the agents themselves. In short, there is a connection between “practical agency and deliberation,” just as there is a relation between “judgment and action,” in the sense that practical deliberation (practical reasoning) establishes the ownership and authorship of an action (Bagnoli 2013, p. 164). Bagnoli identifies an important convergence between Anscombe's practical cognitivism and Korsgaard's constructivism in that there is a special normative relation of agential authority between the agents and their actions<sup>8</sup>, which she defines as a moral feeling:

“The moral feeling of oneself as a practical subject is the feeling of acting on the basis of reason. It is this feeling of being susceptible to reason that establishes the special relation of authorship between agents and their action. [... It] conveys practical knowledge as knowledge of oneself as a practical subject [while at the same time acting] as a deliberative constraint that regulates self-love and self-interest.” (Bagnoli 2013, p. 177-178)

In this way, moral action holds implications for the identity of the agents themselves. There is an intimate relation of authorship between the agent and their action, and it is why the agent must find ways to justify and vindicate it. Moral constructivism takes this exercise and portrays it through a so-called practical standpoint where agents evaluate both their reasons for action and their moral weight and implications for other subjects. The characterizations of this standpoint and the agents themselves mark the different

---

<sup>7</sup> The role of Sentiment is not that relevant for Street's work and central claim. Rather, I am simply suggesting it could be accepted as something which grounds some of the subjective values that agents will endorse within the practical evaluative standpoint in Humean Constructivism, to be explained further in Chapter II.

<sup>8</sup> This relation of authorship is incredibly important specifically for both O'Neill's and Korsgaard's constructivist contentions, as well as my own.

contentions of moral constructivism, as well as the conclusions that follow from each of them.

## 2. Metaethical Moral Constructivist Framework

The term “moral constructivism” was first introduced by John Rawls in 1980 in “*Kantian Constructivism in Moral Theory*”. Branded *Kantian* constructivism, this first contention of moral constructivism had several features which, of course, have not been able to resist the test of time, leading to many criticisms. In fact, Rawls himself changed his position over time, moving *justice as fairness* from a more *metaethical* to more *political* position (O’Neill 1989, p. 22), something which I will address in-depth in the first chapter. Since then, several kinds of moral constructivism have been developed by different philosophers, most of them Kantian (O’Neill 1989; Korsgaard 1996), with other variants such as Humean (Street 2012), Habermasian (Rees 2020) and even Hegelian (Laitinen 2016).

As I mentioned before, what all kinds of moral constructivism have in common is the contention that moral norms are *not* independent of the moral agents, metaphysically fixed and waiting to be simply apprehended, but must instead be *constructed* by the moral agents themselves. Due to the limited space and time at my disposition, for the purposes of this thesis I will choose to focus solely on the Kantian and Humean approaches. I will do so because the literature on Kantian constructivism - being far more extensive than the other kinds due to its earlier genesis - allows for a more detailed and consistent analysis of the arguments presented over the years by its different contenders. At the same time, the Kantian constructivist commitment to proving the existence and bindingness of *objective moral norms* makes it extremely interesting to reflect upon. In turn, Humean constructivism is born as the antithesis of the Kantian project, admitting that if moral norms are a construction of rational moral agents, they can at most have limited intersubjective validity, since each agent’s substantive set of values come into play, making it the perfect contender to measure the Kantian constructivist project against. From the analysis of both the merits and faults of each approach, I hope to draw interesting conclusions which may ground my own proposal.

On Lafont’s understanding of the metaethical debate traditionally split between moral realists (typically cognitivists) concerned with objectivity against anti-realists (typically expressivists) who accept subjectivity, Humean constructivism seems to have an

irrevocable commitment to the later, while Kantian constructivists specifically are much harder to place in either category. On the one hand, Kantian constructivism rejects moral realism insofar as it claims that judgments do not correspond to “*a pre-given moral order, heteronomously imposed on us independently of our practical reason*” (Lafont 2004, p. 28) but at the same time they reject abandoning the claim to objectivity which, for her, makes them closer to cognitivists than expressivists. Another typology which can be used to understand the distinct metaethical commitments of the relevant moral constructivisms I selected is the one employed by Paul Formosa in his article “*Is Kant a Moral Constructivist or a Moral Realist?*” (2011). Formosa splits moral constructivists between “all the way down” and not “all the way down.” While the former are relativistic because they consider that the content and authority of the procedure is constructed through an *actual* (individual or collective) act of sheer willing/agreeing, the latter take non-relativistic views, since they do not rely on *actual* but *hypothetical* acts of willing and agreement. In this latter approach, not “all the way down” constructivists take the content and authority of the law to be independent of *actual* acts of willing. Despite this difference, both kinds defend a cognitivist success theory, and he specifically labels Kantian constructivism as such moral realist in a weak sense of the term. Because they believe that there is a rational procedure out of which we can construct valid maxims, meaning that objective value can come from a procedure which applies practical reason, he holds that constructivism must *imply* weak moral realism<sup>9</sup>. In seeking to ground objective moral truth from a procedure which filters the possible agreements of agents, it must endorse intrinsic value of rational agency. This seems so because of the tension between the formulas acting as either a *procedure* which *discovers* the moral norms, or as *procedures* which *ground* the moral norms. Indeed, Street’s criticisms of constructivist proceduralism (which she ascribes to the works of Rawls and Korsgaard) follow this same concern. She holds that constructivist procedures do not uncover normative truth but rather constitute the emergence of the norms themselves, and as such, she points out that “*the notion of procedure is ultimately merely a heuristic device, whereas the philosophical heart of the position is the notion of the practical point of view and what does or doesn’t follow from within it*” (Street 2010, p. 366). Her critique towards proceduralist approaches thus rests on the fact the characterization of the procedure

---

<sup>9</sup> It is worth noting that in this article Formosa claims that Kant can be read as both a moral realist (through the Second Formula of the Categorical Imperative) or a moral constructivist (using the First and Third Formulas of the Categorical Imperative).

invariably influences and shapes the kind of norms that could arise, and yet proceduralist constructivists (typically Kantian ones) try to justify the existence of objective moral norms by using their so-called “impartial” procedures. While they claim that these are stripped from any preconceived values or assumptions about the will and nature of human beings, she disagrees.

Following this critique, she proposes a *practical standpoint characterization* in opposition to the *proceduralist characterization*. Instead of having a procedural account of normativity, she claims that we must accept that every agent – despite knowing or not what value consists of – is a valuing creature who can therefore deem anything valuable or not. At the moment of norm-making, these valuing agents enter a *practical point of view* or *practical standpoint*, where they measure their substantive reasons against one another. In this approach, the agents are not tasked with stripping themselves of their subjectivity, their considerations, or their own values to accommodate a proceduralist view which asks them to think impartially, objectively, and universally according to what everyone should think. Rather, it merely describes the process through which one measures their reasons. In short, there aren’t any “*substantive normative assumptions about what anyone should or ought to do;*” it simply describes “*what is constitutively involved in the attitude of valuing or normative judgment itself*” (Street 2010, p. 367). As such, she argues that while there are *standards of correctness* that apply to the rational deliberation to endorse normative claims, subjective substantive values are *entailed* within the practical standpoint, and thus considered and accounted for. By stripping constructivism from a proceduralist approach, Street reclaims anti-realism in her Humean constructivism. Furthermore, proceduralism “*where the practical point of view is given some substantive characterization*” (Street 2010, p. 367) leads only to a “restricted constructivist view”, according to Street. This is especially true in the case of Rawls’s proposal, which deliberately limits the set of shared values endorsed by agents<sup>10</sup>. Alternatively, where “*the truth of a normative claim consists in that claim’s being entailed from within the practical point of view, where the practical point of view is given a formal characterization*” (Street 2010, p. 369), we identify “*metaethical versions of constructivism*”. In these cases, the procedure does not impose any substantive restrictions on the agents’ evaluative practical standpoint. While Humean constructivism is of this kind and argues that no moral conclusions can follow from a merely formal point

---

<sup>10</sup> More on this on Chapter I.

of view, Kantian constructivism of the kind of Korsgaard still endorses that moral conclusions do follow, namely because of the value of humanity no matter the given starting set of values of each agent (thus embracing the sort of weak moral realism described above)<sup>11</sup>.

This overview is helpful in determining what is common and different between the kinds of moral constructivism. I hope it gives a general idea about the kind of topics that will be addressed over the course of this thesis and where some divergences can be found. The following chapters follow the different contentions introduced in this first section closely. Hopefully, this introduction can resonate with the reader throughout this work making it easier to keep a mental account of the metaethical standings of each theory as it is discussed.

## **Chapter I**

### **Kantian Moral Constructivism**

In this chapter I shall go over the earliest and most prominent Kantian constructivist approaches, following the development of these theories chronologically. Since John Rawls first coined the term, his work and its subsequent critical analysis marks the starting point of my study. The criticisms addressed to this view will mainly follow Onora O'Neill's commentary on why Rawls's work better fits a contractarian rather than constructivist designation, as well as how the agents are idealized in the procedure which grounds the agreement. Focusing on her criticisms will allow me to transition to her own approach, where the importance of recognizing the plurality of rational finite agents is stressed, as well as the issue of scope. All of this, along with critical remarks about her theory, will be addressed within the second section of this chapter. Finally, in the third section, I will go over Christine Korsgaard's own contention of Kantian constructivism, featuring her original constitutivist argument for the bindingness of moral action to preserve our practical identities. I will leave the main criticisms addressed to her view to

---

<sup>11</sup> The argument that from a formal characterization of the practical point of view no substantive moral truths may follow will be extremely relevant for my argument and I will address it on the end of Chapter II.

be discussed in the following chapter, since they help understand the emergence of Humean Constructivism.

The methodology of chronological analysis chosen enables a depiction of the shift from initial contentions of pragmatic or practical agreement in what concerns issues of political fairness and social justice to issues of everyday interpersonal ethics. To resort to T. M. Besch's (2009) terminology, I hope to show how Rawls's constructivism can be labeled as more of a "political" brand of constructivism, whereas O'Neill's (and later Korsgaard's) can be understood as "more strictly Kantian". This pertains mostly to the kind of practical problems addressed above. In the case of Rawls, those would be the structure of a just society, and consequently the moral justness of its institutions, issues which are political in kind and limited in scope (to the citizens of said societies, or their representatives). While O'Neill also seeks to address issues of political justice, her insistence on the scope of affected people by our actions enters the ethical domain of agent-to-agent interactions, something which Korsgaard also does, hence the designation of "more Kantian in kind." Autonomous norm-making, trying to maximize the scope of agents affected by our actions, the inherent value of our identity as human beings, and practical reason and its internal standards of validity will be pivotal points for this chapter.

### **1. Rawlsian Constructivism**

As the pivotal author in moral constructivist approaches to morality, John Rawls's contribution is undeniable as a starting point from which several interpretations and criticisms sprung, and thus valuable from an evaluative standpoint to understand how different basilar concepts came to be and evolved over time. It is important to note that Rawls's most prominent works, beginning with *A Theory of Justice* (1971) are directed towards how one can come up with the best principles of justice for a hypothetical just society. The devices and procedure worked out were later reused to fit moral theory, in *Kantian Constructivism in Moral Theory* (1980), where his view of what would be Kantian moral constructivism was first developed. Over the years, however, in other works such as *Political Liberalism* (1993), the initial commitment to universal agreement was dropped in pursuit of a much more political approach that, presupposing a shared democratic, pluralist, and liberal conception of society, assumed much more about the nature and beliefs of the rational and reasonable agents considered. This evolution signals the depart from a more metaphysical approach to a more politically situated, practical one. Nonetheless, I find it crucial not to skip a thorough analysis of Rawls's.

## 1.1 Pure Proceduralist Constructivist

“Rawlsian” constructivism<sup>12</sup> is a strand of moral constructivism which is not only “proceduralist,” but specifically labeled by himself as a “pure proceduralist approach.” Moral norms are to be constructed and derived from a kind of procedure, and it is the procedure itself that justifies the norms conceived and grants them their legitimacy. This Kantian form of procedural constructivism “*specifies a particular conception of the person as an element in a reasonable procedure of construction [...] within this procedure persons characterized as rational agents of construction specify, through their agreements, the first principles of justice.*” (Rawls 1980, p. 516). This proceduralist approach envisions agents - considered *both* reasonable and rational as well as constrained by devices such as the original position and the veil of ignorance - to (to the best of their ability) impartially assess norms which would be acceptable to all. If consensus on certain norms through the procedure is reached, then it must mean that those norms are objectively *fair* for all the parties involved in the procedure. Through this procedure and agreement over fair principles of justice, Rawls claims to reject strong moral realism since the moral norms are not independent of our ratification. In fact, the norms reached are grounded in our agreement, in our shared cognition and understanding. Rawls’s goal is that the devices and constraints which he implements in his approach guarantee that the procedure itself allows for a construction of norms that are not simply based on intersubjective and heteronomous agreement, but that they are able to ensure the hypothetical agreement which would be reached would be based solely on rational and reasonable autonomous deliberation:

“It seems better to say that in constructivism first principles are reasonable (or unreasonable) than that they are true (or false) – better still, that they are most reasonable for those who conceive of their person as it is represented in the procedure of construction. [...] Objectivity is to be understood by reference to a suitably constructed social point of view, an example of which is the framework provided by the procedure of the original position. [...] The essential agreement in judgments of justice arises not from the recognition of a prior and independent moral order, but from everyone’s affirmation of the same authoritative social perspective.” (Rawls 1980, p. 571)

Going as far as to reject the notion of moral truth or falseness of the principles endorsed within the procedure, instead endorsing that they would be reasonable and thus fair, it

---

<sup>12</sup> In sharing a lot of features with Rawls’s *A Theory of Justice* (1971) where he labels the theory as a hypothetical social contract, this approach fits the label of contractualist better than constructivist. A detailed review on such a criticism can be found on O’Neill’s *Constructivism vs. Contractualism* (2003), and I will briefly discuss it ahead.

seems clear that Rawls's commitment to anti-realism is vindicated, and resorting to Paolo Savoia's (2015, p. 97) terminology, one can argue that this model would fit within the category of achieving *modest objectivity*. In arguing that there are no moral facts, no prior and independent moral order which ought to be recognized, he endorses that norms must be constructed, devised, and elaborated on, through a process which preserves the rational autonomy of every deliberative party involved. An important remark is that while he refers to moral constructivism, the terminology employed of principles of justice refers to the structure of the basic political institutions of a "well-ordered society." I take this to be relevant since the level of application of political norms differs from that of moral norms. While the former pertains to institutional-individual dynamics, the latter pertains to individual-individual dynamics. In being directly concerned with the aspect of socio-political organization at the institutional level, it resorts to a particular conception of a "well-ordered society and of a moral person." The moral person thus conceived has, according to Rawls, two kinds of autonomy: 1) full autonomy, which is experienced in everyday lives where citizens choose the best course of action to pursue their ends, and 2) rational autonomy, to be applied in the formulation of the principles of justice under Rawls's procedure (see Rawls 1980, p. 520). Another key aspect is that in this paradigm the deliberation and agreement is not made by *actual* individuals but reasonable and rationally autonomous "representatives" in an original position instead. They are reasonable insofar as they understand that social cooperation must rest in "fair terms of cooperation", articulating the idea of "reciprocity and mutuality". Simultaneously, they are rational insofar as each participant is trying to advance what contributes to their own rational advantage<sup>13</sup>.

The procedure thus consists in an original position where these rational and reasonable agents (the representatives of individuals in a society) are placed behind a veil of ignorance. Being rationally but not fully autonomous, and behind a veil of ignorance, they must reason without being aware of their own predicaments, such as their "*place in society, their class position, or social status, [...] their fortune in the distribution of natural talents and abilities, [...] their own conception of the good, that is, their particular final ends [...] their own distinctive psychological dispositions and propensities, and the like.*" (Rawls 1980, p. 522). This ensures their equal standing as free rationally autonomous agents with moral capacity, stripped of their own subjective

---

<sup>13</sup> In the sense of being mutually disinterested as mentioned in *A Theory of Justice* (1971).

contingency and thus unable to advance their own desired ends. From this equal standing the validity of the principles adopted as fair is guaranteed, hence the name of the theory ‘*justice as fairness*.’ Without these limitations on knowledge about themselves and their place in the world, the parties would be capable to determine which conditions would suit them best according to their own characteristics and in turn exploit that knowledge to their own advantage in norm-making. The agreements made would be based on heteronomy rather than autonomy, which understood in Kantian terms refers to a subversion of purely rational thinking to the precepts, desires, and self-interest characteristic of beings fully aware and fully autonomous who could try to one-up the fellow parties. Their “ignorance,” in turn, allows for fairness in the deliberation where purely rational autonomy leads the parties to deliberate only in terms of minimal self-interest, guaranteeing that the principles endorsed would never allow them to be in an extremely dire social position. This ensures a minimal threshold of fair living conditions, with necessary freedoms and justice that the system must uphold for all, in the logic of the *maximin* principle.

In a utilitarian critique made by John C. Harsanyi (1975), the maximin principle is stated to be discriminatory toward rich or not-extremely-poor people; but I take this to be misplaced since the consensus on principles is made *before* each person knows their socio-economic status. The agreement only tries to minimize the damage to the worst-off people, since no one wants the possibility to be in a terrible position when it comes down to practical reality. This way, in Rawls’s envisioned just society, even the worst-off people would be able to agree with the principles that allowed for their situation as *just*. In Rawls’s terms, the “good” is understood as each person’s conception of it, while the “right” is what is fair for everyone. Thus, in the original position through the veil of ignorance, reasonableness “*subordinates the Rational*,” ensuring that a “*priority of the right over the good is characteristic of Kantian constructivism*.” (Rawls 1980, p. 532).

In short, in “*pure procedural justice*,” the deliberation within the original position defines the principles of justice. They do not exist independently and are ratified by the procedure. By rejecting the existence of independent and antecedent criteria of morality, Rawls explicitly rejects moral realism and commits to a form of anti-realism which rests only on the cognitive reflections of the parties, unaware of their own personal biases. Ensuring the rational autonomy of each party also ensures that the agreement is not heteronomous. The essential aspect of Rawls’s procedural proposal is the heuristic

importance of the original position and veil of ignorance. However, this process needs to be complemented by a third method: *reflective equilibrium*.

## 1.2. On Reflective Equilibrium

Before addressing reflective equilibrium, I want to expand on the issue of Rawls's approach being focused on socio-political norms to be agreed upon by agents. Recall that the parties in the moment of deliberation are placed behind a veil of ignorance and tasked with reaching a consensus on principles of justice acceptable to all. Rawls mentions that this consensus can be reached "*only to the principles of political and social justice and not to all moral notions [since] for many philosophical and moral notions public agreement cannot be reached*" (Rawls 1980, p. 539). Further, he has no problem with accepting that a conception of justice need not be based on the "full truth of a society" but a part of it, since "*if the whole truth is to include the truths of religion and philosophy and of moral and political doctrine*" (Rawls 1980, p. 542), then it wouldn't be possible to reach such a consensus. In short, religious or mortal "truths" are to him part of each person's conception of a good life, thus pertaining to subjective considerations of value, and the only "truth" of a just society that we can find is the "political," reasonable truth. In this sense, justice as fairness, albeit Kantian in inspiration, is fundamentally different from Kant's ethical doctrine. The first reason for this is that the object is the basic political structure of society. While "*Kant's account of the Categorical Imperative applies to the personal maxims of sincere and conscientious individuals in everyday life [and thus] Kant proceeds from the particular [...] Justice as fairness moves in quite the reverse fashion: its construction starts from a unanimous collective agreement regulating the basic structure of society within which all personal and associational decisions are to be made in conformity with this prior undertaking.*" (Rawls 1980, pp. 552-53) We can thus say that Rawls's "*Kantian*" constructivism is not constructivism about moral norms but one about the possible "exoskeleton" of a pluralist society where all could agree to live according to fair principles, regardless of everyone's own personal conception of the good. It is meant to reach basic minimal consensus on general norms of socio-political functioning - principles of justice wide enough that they allow for different conceptions of the good to be advanced in society, hence labeled a theory preoccupied with the "practical task of political philosophy."

Committed to anti-realism about moral truths of the good, the Rawlsian conception of the person believes "citizens think of themselves as self-originating sources of valid

claims” who are also capable of “revising and changing (their own) conception (of the good and final ends) on reasonable and rational grounds” (see Rawls 1980, p. 543). As such, on this account, people are both free and endowed with the capacity to make their own judgments, while the procedure grants them a position of equal worth, or parity, from whence they must adopt institutional principles of social cooperation which they will then follow (see Rawls 1980, p. 546), due to being reasonable. From within this position, Rawls believes the first step will be to try and secure goods, hierarchized in terms of priority<sup>14</sup>. According to Rawls, the primary goods are the things which are generally thought of as necessary social conditions to enable human beings to realize their moral powers and pursue their final ends. In doing so, while the author accepts that there might be different interpretations on how this hierarchy will play out and the principles that can be formulated, in advancing his own interpretation of the matter, “*Rawls breaches the veil of ignorance to allow knowledge of the universal desirability of primary goods such as rights, liberties, and wealth into the original position.*” (Watkins & Fitzpatrick 2002, p. 350). His own conception of the person points their agreement in a specific direction, and his “constructivism” is built on an idealization – rather than abstraction - of the “person” and the procedure<sup>15</sup>. This is particularly true for the final piece of the “Rawlsian puzzle,” reflective equilibrium, is the process through which the actual justification and for the bindingness of the principles conceived and their endorsement is achieved. Onora O’Neill describes it as such:

“Reflective equilibrium is a process rather than the determinate outcome of a process. Judgements may be modified to achieve coherence with principles of justice that supposedly would be chosen in a hypothesized original position; provisional principles of justice and the hypothesized original position may be modified to achieve coherence with actual judgments. The most fundamental justification which Rawls offers for his principles is not that they are derivable from a hypothetical original position [...] It is that they receive the ‘mutual support of many considerations, of everything fitting together in one coherent view’.” (O’Neill 2006: 36)

In it, we go “back and forth,” placing the principles formulated in the original position against our “best considered judgments about justice.” If those principles formulated in the original position do not coincide with our basic intuitive judgments about justice, then they must not be the principles of justice we are seeking. Reflective equilibrium amounts

---

<sup>14</sup> Since the focus of this thesis is on moral constructivism, I will not further explore this question of political nature other than the brief remark that follows.

<sup>15</sup> See O’Neill (2010).

to a process of evaluation which “confirms” the principles of justice<sup>16</sup>. O’Neill makes the claim that, in the end, it is reflective equilibrium that is the true method of justification for Rawls and makes the agreement not-so-hypothetical as initially intended. The original position ends up being simply a device of representation subordinated to the justification process of reflective equilibrium, which raises the issue of scope about whose considered judgments are to count for the process of justification (see O’Neill 2015, p. 61). The bindingness of the principles stems from the fact that they are not only agreed upon within the original position but that they receive further endorsement from complying with such considered judgments of the parties. Now, while Rawls argues that:

“The parties in the original position are moved by their preference for primary goods, which preference in turn is rooted in their highest-order interests in developing and exercising their moral powers. Moreover, the agreement of the parties takes place subject to constraints that express reasonable conditions.” and thus, if “such a construction does yield the first principles of a conception of justice that matches more accurately than other views our considered convictions in general and wide reflective equilibrium, then constructivism would seem to provide a suitable basis for objectivity.” (Rawls 1980, p. 568)

Herein lies a problem with Rawls’s approach. We can argue that if there is no independent moral order which encompasses moral facts, then the parties do not “find” moral facts but instead create principles based on their subjective, or relative, endorsements of value. The claim that these can be “objective” can now possibly only be understood as “*minimally objective*”<sup>17</sup>. This is an issue which pertains to scope, as stressed above. If the principles of justice conceived in the original position are to be “tested” against the considered judgments of actual persons, whose judgment counts? Whose reflective equilibrium matters for norm-making? In his later works, particularly *Political Liberalism* (1993), there is a clear emphasis on political justification with a closed society in mind. In this later work, there is a distinctive conception of public reason where “*citizens count, and noncitizens do not [...] the societies in which such reasoning is seen as possible are now explicitly confined to **bounded, liberal and democratic** political societies, each conceived of as a complete and closed social system*” (O’Neill 2015, pp. 61-62). This is concerning since it indicates a shortening of the scope of applicability of the principles of justice, as well as an apparent slip into heteronomous thinking. While the parties described as autonomous must reason in fair terms, the basis for agreement regarding the fundamental principles of justice must now rely on a shared

---

<sup>16</sup> See Onora O’Neill (2015, p. 60).

<sup>17</sup> Again, resorting to Savoia’s (2015) terminology.

liberal democratic disposition, but invariably influences the reasoning and justification for the political principles elaborated. These considerations are “prejudged”, according to O’Neill, and non-citizens are excluded from consideration, possibly meaning that those who hold the kind of presupposed considerations about matters of justice, religion, political forms of organization opposed to those endorsed by Rawls’s conception of the democratic person must be excluded from the reflective equilibrium, as they do not share the basis for this understanding.

While the initial veil of ignorance tried to avoid heteronomous reasoning that Kant would have rejected, relying instead solely on an instrumental conception of rationality, O’Neill first stated that Rawls’s initial approach in the original position<sup>18</sup>, it seems that the veil is actually in some way breached to allow for certain preferences and political cultures to permeate into reflective equilibrium, influencing the construction of norms, hindering the desired autonomy and turning it into a heteronomy of sorts. By both lifting the veil to assume which would be the primary and secondary goods chosen<sup>19</sup>, as well as limiting the scope of this exercise so that the parties share certain presupposed values, “*Rawls’s idealizing assumptions are built into his account of the original position, hence determine what would be agreed on by persons in such a position.*” (O’Neill 2010, p. 33). From both failing to secure that the veil of ignorance prevents people from making heteronomous judgments and by “corrupting” the procedure itself with normatively charged claims and assumptions of his own, Rawls fails at distancing himself from heteronomy. This analysis shows how Rawls’s contention of moral constructivism can be depicted as *minimally objective* (due to being relativist), or simply *coherentist*<sup>20</sup>, charged with presupposed and unjustified values of liberalism and pluralism. On the words of Onora O’Neill: “*His account of the reasonable (as opposed to the merely, instrumentally rational) assigns weight to what happens to be shared among fellow citizens, and his version of public reason assumes the boundaries of some actual public or people. In consequence his account of ethical justification is not universal in scope, and unsurprisingly his conception of justice is not cosmopolitan.*” (O’Neill 2003, p. 323). For this reason, O’Neill put forth her own conception of moral constructivism, whose strategy

---

<sup>18</sup> See O’Neill’s *Constructions of Reason: Explorations of Kant’s Practical Philosophy* (1989).

<sup>19</sup> *Ibid.*

<sup>20</sup> See O’Neill 2003.

involves thinner starting points and a broadening of the scope of moral norms and can be found in the following section of this chapter.

## **2. Onora O’Neill’s “Modal” Constructivism**

The criticisms addressed to Rawls in the last section of this chapter issue are extremely relevant for the development of O’Neill’s contention of moral constructivism. Her several disagreements with Rawls fuel her different approach. Rawls’s method, by using the veil of ignorance in the original position and depending on reflective equilibrium, “eliminates the possibility of disagreement.” It dehumanizes the agents in a way that removes knowledge of what distinguishes them from one another, and “*seemingly obliterates the very context for any agreement or contract by eliminating distinctions between persons: it trivializes the notion of agreement by eliminating the possibility of disagreement.*” (O’Neill 2003, p. 319). O’Neill’s approach does not resort to a kind of procedure which tries to depersonalize agents or remove knowledge they may have of themselves and one another or take persons to be correspondent to an idealized account. Instead, it starts out from a minimal thin starting point of *facts* about the agents and constructs norms, which must respect certain criteria (of followability and greater scope), to be applied resorting to the judgment of agents endowed with the power to be aware of their predicaments and of others. The rejection of idealization in combination with expanding the scope of constructivism are central for O’Neill’s approach. Whether her position holds up to these standards or not, I will discuss at the end of the section.

### **1.1. The Problem of Scope**

Regarding scope, in the article *Changing Constructions* (2015) which I have already quoted, O’Neill addresses the issues and changes between *A Theory of Justice* (1971) and *Political Liberalism* (1993). Her analysis suggests that while on *A Theory of Justice* the devices served the purpose of establishing a hypothetical just social contract through usage of a procedure – hence the claim that it is actually contractarian – in *Political Liberalism*, the notion of hypothetical agreement is dropped in favor of a more “political constructivist” stance, due to the focus on emergent themes of globalization and open societies: “*the metaphors of the social contract were mostly replaced by the idea that principles of justice can be constructed by reasonable procedures*” (O’Neill 2015, p. 58). This shift results in a problem of scope which I already hinted at in the previous section and moves Rawls from seeking modest objectivity to minimal objectivity. The inclusion

of concepts such as *citizenship* are accounted for and matter for the construction of norms. To accommodate the change from a society conceived as a closed system to one that is open<sup>21</sup>, and in focusing on the construction of political norms, Rawls takes the *democratic* culture of certain societies as essential, embodying and accounting for *pluralism* of values embedded into the process of norm-making. This revamped approach also relies more on reasonableness rather than rationality which leads Rawls to the substantive conclusions he holds about the shared democratic culture and values. The reasonable allows for norm-making that accounts for the substantive values of us and others, concerned with public affairs and a shared common liberal democratic culture and political beliefs. Taking all of this into account, O’Neill identifies an impasse:

“Normative content cannot be derived from nothing, but if it is derived from the norms that are taken to be shared by fellow citizens in liberal democracies, then justification will be relativized to those norms. On the other hand, if it wholly abstracts from actually accepted norms and does not build on any determinate content, will it be possible to establish any principles of justice?” (O’Neill 2015, p. 66)

She holds that assumptions about the substantive shared values of citizens should be dropped to avoid restrictions and exclusions which lead to norm-bias. In a way, at least in *A Theory of Justice* the procedure shielded the original position from the substantive values taken for granted in *Political Liberalism*. In trying to reclaim true objectivity for constructivism, while accounting for true pluralism, O’Neill argues one should not presuppose these values but rather put forth norms which are *universal* in both form and scope; if these norms hold universally, they are lawlike and actually shared by all, despite the substantive values of each. A norm which is universal in form but not scope is not really lawlike or universal but rather relative to those whose substantive set of values accept it, like the specific institutional arrangements and forms of liberalism and democracy put forth by Rawls. Since the point is to try to create an ethical theory that tries to address individuals who are not decontextualized, she claims that:

“Combining universal *form* with universal *scope* may provide a basis for a more useful conception of public reason that is relevant to identifying proposals and principles that can be communicated and potentially justified to an unrestricted audience [because] the claim that scope is a secondary matter and that we can generally justify principles of justice and subsequently adjust their scope as an afterthought or corrective may be mistaken. *Scope often also matters for justification.* [...] There are many practical principles that can readily

---

<sup>21</sup> See O’Neill 2003 (p. 322) for a deeper understanding of this dynamic and shift and the concept of “restrictive” justice in Rawls.

be adopted by *some* agents but that cannot be adopted by *all* agents and so cannot be coherently proposed to or justified to an unrestricted audience [because they] would be expressed and reflected in action that would disable or prevent at least some others from like action. Principles of violence or coercion, of manipulation or deception, and many others cannot coherently be proposed to or justified to *all* others.” (O’Neill 2015, pp. 68-69).

In combining the criteria that the norms must both 1) hold for all and 2) take a lawlike form, O’Neill’s approach is closer to the Kantian standard. Its focus is not on finding norms which *can* be justified and accepted in the process of political construction between individual with shared values, but rather asking which norms *cannot* be universally adopted. She grants that the principles that ought to arise from such an approach will undoubtedly be quite broad and general, focused on the *protection of agency* for those accounted as affected by the action. In rejecting Rawls’s idealization of the person, and thus the kind of values they will stand for as not truly constructivist<sup>22</sup> in the desired Kantian sense, O’Neill seeks to broaden the scope by taking the long route of justification of very broad and wide principles of non-coercion, respect and protection of others is preferable since justification matters for settling both political and moral disputes, and it is through the criteria of justifiability (which as we saw must meet a criteria of universal scope) that such process must be done; we cannot simply presuppose a democratic and pluralistic culture to be the most adequate and without need of justification. Thus, O’Neill’s approach rests on a thinner starting point in the form of a modal question.

## **1.2. Practical Reason and Judgment**

As a constructivist, O’Neill embraces (or brackets) anti-realism by insisting that there is not an independent moral order from which we can uncover moral facts as such. This, however, does not stop her from trying to advance that objectivity within anti-realist moral constructivism is attainable through a Kantian form of constructivism which can establish criteria for identifying objective moral norms, by resorting to practical reasoning to guide action. In a constructivist account, then, we must build norms from the ground-up, maintaining the legitimacy of an anti-realist constructivist approach (or bracketing anti-realism) by resorting to the use of abstractions – not idealized conceptions – of the person. This abstraction is characterized by abstracting from anything but empirical truths as starting points, associated with facts about human life, capacities, and

---

<sup>22</sup> See Budde’s *Constructivism all the way down – Can O’Neill succeed where Rawls failed?* (2009).

vulnerabilities”<sup>23</sup>. According to Budde, this commitment to using only empirical facts about human life abstracts from metaphysical assumptions<sup>24</sup> which, due to the fact of plurality of agents, cannot be securely vindicated. Of course, these

“empirical (abstract) starting points have to be truly and only empirical facts without there being any inherent (moral) value attached to them (...) That is, an assessment of human capacities just has to look empirically at what capacities human beings have without attributing to (some of) them any intrinsic value, for example, by seeing them as something worthwhile to preserve, in short, without making any value judgment about them in the beginning. The second point is that O’Neill clearly treats these potential empirical truths or facts as non-constructed. (...) While O’Neill therefore contra Rawls assumes that the procedure is to be itself constructed, she seems to agree that constructivism has to start from some basic material which is not constructed.” (Budde 2009, p. 206).

The facts about human life prescribed above are simply pointed out due to empirical knowledge about them, without any inherent value attributed by the philosopher prior to the reflection about the norms who would be constructed. The characteristics chosen as starting points for the depiction of persons are not the product of a construction, but rather what the construction (of values and norms) will be grounded on; characteristics which function as empirical common denominators for all agents considered. In admitting such a common ground can be found on the empirical known capacities, capabilities, and vulnerabilities shared by all human agents, O’Neill succeeds in making this a very thin starting point of abstractions. The conclusion that follows is that any principles that would coerce, exploit, or hurt them on the basis of those characteristics would not be acceptable or adopted by *all*. Those for whom the norms or principles would be coercing could not, or would not, accept the norm as valid or good. Because the point is to reach norms that could be accepted by all, the empirical facts of finitude, limited capacity for action and vulnerability place constraints on which facts *could* be adopted. This is not due to a moral value ascribed to the necessary protection of human life or reason, but simply through the reasoning that they cannot be universalized or publicly shared and followed, and thus cannot be objective moral norms. Because moral realism has been bracketed and there is no independent order of moral values, according to O’Neill, moral principles can only have normative force either because reason can vindicate its own authority or if there was some other source of authority, but within a framework of anti-realism the second option cannot be the case. As such, the source of authority for the norms must be the internal

---

<sup>23</sup> See O’Neill’s *Towards Justice and Virtue: A Constructive Account of Practical Reasoning* (1996).

<sup>24</sup> See Budde 2009, p. 205.

authority of reason. But where does this internal authority come from? According to O'Neill, there is a “*‘fact’ of plurality and the ‘fact’ of the absence of a pre-established harmony or coordination of this plurality (and the supposed ‘fact that this plurality nevertheless ‘needs’ coordination) will play a vital role in establishing the procedure of constructivism, the principle of practical reason.*” (Budde 2009, p. 206). In recognizing these facts, O'Neill claims that agents must find a common authority to coordinate them, and this authority cannot be based on arbitrary commitments<sup>25</sup> (such as subjective whims or conceptions of the good), which leaves practical reasoning as the source of such authority. Because a plurality of agents needs to establish principles of coordination, they must reach “*agreement based on principles that meet their own criticism. In short, reason, and thus the practical principles based on it, can be justified on O'Neill's account, but only recursively by means of the standards that can be accepted or agreed upon by a plurality of rational agents engaging in public or political discourse.*” (Watkins & Fitzpatrick 2002, p. 354). It is a form of constructivism whose legitimacy draws from a shared conception of public reasoning. Now, because reasons are to be shared and endorsed publicly by rational agents, O'Neill's constructivism rests on the double modal claim that any reason for action must necessarily be held “capable of being endorsed and followed by all relevant others”<sup>26</sup>. This modal claim establishes the desired “thin notion of equality,” product not of substantive valuing but from “a construction of practical reason as inherently intersubjective” (Ronzoni 2010, p. 88).

The proposed view grants that from personhood abstraction, empirical knowledge of facts of life, and the agent's practical reasoning in the form of public shared reasoning purposefully leads agents to create norms that harmonize their interactions with each other. The requirements for being followable must rest on the noncontradiction with the conception of the person bred from abstraction, that is, the norms must take into account the possibilities for action according to the empirical truths identified at the beginning of the constructive process. In being followable and acceptable by all, they acquire validity. This resembles the Kantian requirement of universalization, the First Formula of the Categorical Imperative. Regarding scope, or the question of who is relevant for the criteria of followability, the construction of the norms must bear in mind the agents which are to be in any way involved or expected to be involved by the principles. We know our actions

---

<sup>25</sup> See Melissa Barry's *Constructivist Reasoning and Objectivity* (2013).

<sup>26</sup> *Ibid.*

will have an impact on *someone*, and thus whenever we act, or take ourselves as capable of acting and interacting with others. Furthermore, for O’Neill, activity presupposes plurality, connection, and finitude, insofar as we know that others, who are “independent sources of activity” and are “limited in their powers to be vulnerable to the effects of our activity,” might be affected and react to our actions<sup>27</sup>. Due to this,

“We *must*, by pain of incoherence, or unreasonableness, include them in the scope of our practical reasoning. [...] Agents cannot coherently, or reasonably, refuse to include in the scope of practical reasoning all others whom they take to be on the receiving end of their intellectual and non-intellectual activity. Thus, the scope of practical reasoning might not be *strongly* universal in including truly everyone who was, is and will be capable of inclusion in that scope. But it would at least be weakly universal (or ‘cosmopolitan’) in being fully universal within the scope of our other-regarding activity.” (Besch 2009, p. 4).

It seems clear that this form of constructivism encompasses a theory of action that fully understands the relationship of authorship between agents and actions that I mentioned before through the work of Bagnoli (2013). Agents are sources of deliberation and activity; they are vulnerable, finite, and they are connected to other agents which share those same characteristics despite the plurality of other subjective differences between them. In developing the constructivist position which takes only the essential features of human life into consideration, “*she holds that the only materials of construction that we should accept are qualities that we can assume a priori of human beings – hence, no contingent and empirically disputable considerations.*” (Ronzoni 2010, p. 85) I argue that the categories of finitude and vulnerability can indeed be acceptable materials devoid of empirically disputable considerations, to an extent. I contend it is undeniable that human beings are *finite* and vulnerable in different degrees, thus the permanence and universality of these qualities is not subject to contingent and empirically disputable considerations. What is, however, disputable is the extent to which each individually is contingently vulnerable. Nonetheless, the inescapable qualities of finitude and vulnerability seem to work as driving forces for moral deliberation in O’Neill’s theory:

“We know that human beings are vulnerable and interdependent, and we cannot even conceive of humans without thinking of those qualities; we do not, however, know *how* vulnerable and dependent on each other they are exactly. The assumption of human vulnerability is, of course, factual in kind; it is not, however, empirically problematic. For O’Neill makes that assumption on the basis that we cannot even think of humanity without presupposing some sort of finitude and vulnerability. *The account of agency that O’Neill’s theory relies on is therefore minimal: issues of justice arise when a plurality of imperfectly rational and imperfectly independent agents has to interact.* Imperfect rationality and

---

<sup>27</sup> See O’Neill (1996), Besch (2009, p. 7)

imperfect independence are the result of a process of abstraction, rather than idealization: *they are predicates that we can legitimately assume of all human beings, without denying that, in particular situations, individuals could have varying degrees of rationality and independence.*” (Ronzoni 2010, p. 87)

Before moving to the next section where criticisms addressed to O’Neill are reviewed, I want to point out that I fully agree with this statement, and it will be of paramount importance for my own proposal. Because all human beings are finite and therefore vulnerable, these are what we can consider inescapable conditions of our existence. As such, they are characteristics shared by all, albeit unequally. The degree of vulnerability, however, varies according to contingent factors; it varies according to each subject’s embodiment (which influences both physical and mental capacities), as well as particular contexts and predicaments. Following such an understanding, this approach clearly states that no principles which would infringe upon or hinder another agent’s capacity to act, such as principles of violence, coercion, perjury and deception, for instance, would ever be endorsed by all who share such characteristics, and thus they would fail to meet the followability criteria put forth by O’Neill. Of course, such a thin constructivist account can only vindicate objective principles too broad to guide action specific action, one can argue that the scope need not always be universal, but rather universal for those affected by the action, through O’Neill’s reply that this constructivist account formulates principles of action that require context-specific judgment from the agents:

“It is a serious mistake to believe that principles exhaust what needs to be said about practical reason: ‘principles are always to some extent abstract but they are not the whole of practical-reasoning. They must always be applied in ways that take account of actual context; and *they never determine their own application*’. [...] because application is a distinctive operation of practical reason which requires judgment.” (Ronzoni 2010, p. 94)

It seems then, that only some universally objective principles of non-coercion can be vindicated, while the theory accommodates room for reaching “modest objectivity” by adjusting the scope and application to the relevant agents and subjects implicated by the principle of action. Now that this has been explained, the next step is to address the faults pointed towards O’Neill’s contention of moral constructivism, which will be helpful to later explain how I depart from it.

### **1.3. “Substantive” Problems**

Essentially, the criticisms exposed in this section address concerns which have already been hinted at in the former section and are twofold. On one hand, O’Neill’s stand has been criticized on the grounds that 1) it does not actually build from an entirely constructivist standpoint, and instead it masks a goal or end-oriented starting point, and

2) it is not actually able to prove the applicability and of the principles put forward due to their broad nature, and thus their bindingness, and the critique ties in with the first one. On the first issue, recall that O'Neill's main commitment at the beginning of the formulation of her constructivism had to do with making sure no substantive claims of value were imbued into the starting point of constructivism. Instead, Budde claims that:

“Looking at the form of O'Neill's justificatory argument it seems clear that the principle of followability is ‘constitutive or instrumental to’ the end of coordination [...] coordination is not one of the basic conditions or facts from which we construct the principle of reason, but instead is the pre-given end or goal from which we deduce the principle of reason as one necessary condition to achieve that goal. [...] Because it follows an instrumental or teleological pattern of argumentation, it can only hold if we accept the aim or goal of (autonomous, non-violent) coordination as an objective and authoritative goal or value. [...] It seems, therefore, that the search for objectivity cannot evade the need for an (probably metaphysical and foundational) argument for the value of (peaceful and free) human existence as well as capacities and capabilities for autonomous action.” (Budde 2009, pp. 212-213, 218)

This concern is corroborated by Melissa Barry's analysis on the issue. Her view is that O'Neill grounds the bindingness of the norms and her approach on assuming that *moral* coordination is the best and only way to tackle the problem of coordination. She argues that this is not the case, as there are alternative options for coordination (for instance through the use of force). Thus, she claims that the principle of *followability* “subtly relies upon a similar appeal to the substantive value of rational agency” (see Barry 2013, p. 31, 33). The second point of criticism, which pertains to the vagueness of the principles of non-coercion. This mostly has to deal with the problem of varying degrees of vulnerability dependent on contingent factors, making the task of accounting for namely about how far the concept of vulnerable can be extended and what counts as coercion. The extent of capacities which we take to matter for action will vary and are dependent on the considerations of each agent, and thus it becomes unclear what can be acceptable or not under the wide scope of the “non-coercion” principles.

“If the requirement does not entail the demand for the preservation of all or all possible capacities for action (which would presume what the universality test wants to construct) then no practical principle apart from murder, or injury which necessarily leads to death, would extinguish the capacity to act *per se*. [...] Apart from death, which really destroys all capacities, therefore all capacity to act, a principle of injury that falls short of this cannot be rejected on the ground that universal adoption meant that some could not adopt it. Unless there is a worth or value ascribed to all or all possible capacities to act, injury does not prevent acting *per se* and not even acting on the principle of injury. [...] Rather than being the result of the construction I argue that O'Neill's argument works only if the normative commitment to the protection of our capacities to act is one of the presuppositions or starting points of the construction. Thus, I argue that what is substantially underlying, and

motivating O'Neill's argument is a commitment to the value and worth of human beings' particular capacities and capabilities, and a commitment to the ideal of free and equal human interaction." (Budde 2009, pp. 215-217).

Furthermore, by quoting O'Neill's *Towards Justice and Virtue*, Budde adds that the commitment to the value and worth of human beings is heightened since she describes how injury pertains not only to the capacity to act but also to the ways in which the connections and relationships between agents might be affected, possibly leading to an actual increase in vulnerabilities due to damage to the environment that surrounds them. Attending to how complex interactions and capabilities may be affected through every possible action, it becomes difficult to understand how the norms can be objective and applicable in practical reality and cements the claim that there is substantive valuing on the protection of autonomy which isn't really justified by the process of construction but rather imposed on it by the philosopher.

In what concerns the criticism about the method being goal-oriented, a lot can be said about it being an actual problem or concern. If O'Neill's moral constructivism attempts and succeeds at bracketing moral realism, then we have no independent criteria as to why we should comply to any kind of moral values other than trying to come up with our own reason to do so, it seems. In that regard, one may argue that being mutually interdependent, purely instrumental reasoning may lead agents to conclude that it is in their best interest to cooperate on the basis of reasoning which guarantees their survivability and autonomy. I concede that this is already positing value on the question as such, but it does help in some way to explain how morality comes about, if there is prevalence of this kind of reasoning (of course, this is far from the bold statement that it is objectively how agents would always face the issue). On this account, it seems that we must find a sort of common denominator, characteristic, which is essential to all of us and *makes* us compelled to act morally. A sense of self-preservation, in this sense, accompanied by our reasoning demanding that we be *coherent* in our considerations and judgements seems like a promising combination to do so, and what O'Neill tries to advance. Morality is goal-oriented if it is indeed a creation the way constructivism takes it to be. It must be created by agents out of something, either a desire or necessity. In this case, it seems like O'Neill simply identifies – taking vulnerability and finitude as facts – that when reasoning, one would on principle never accept harm toward themselves, and on pain of incoherence, like she said, would also agree to refrain from harming another if that would guarantee their autonomy.

Regarding the second problem of the application of principles, I believe it rests within what I mentioned before about the applicability of the principles always requiring the application of context-specific judgment<sup>28</sup>. Broad principles are principles, guidelines if you will, and the measurement of the actions impact on complex capabilities and interactions between agents must be reflected upon through our judgment.

Finally, while I share a lot of sympathies with the kind of constructivism endorsed by O'Neill in the way it defines a process of construction for those *already committed* to moral deliberation, I will try to avoid the moral realist criticisms that it faces due to positing substantive value of need for coordination. My approach will draw on the kind of reasoning that "*since agents already accord ethical standing to some vulnerable and connected independent sources of activity, they cannot reasonably refuse to accord that standing to all others whom they are committed to construe as such sources.*" (Besch 2009, p. 11). I accept and endorse that agents and the authorship of their actions, as well as recognizing their impact on others, motivate and shape the way in which they indulge in moral reasoning, but I deny that it is a matter of necessity of coordination. I will say, however, that the moment we realize that our actions have implications towards others, we are immediately drawn to question whether these consequences are rightfully justified for them or not. It induces moral reasoning but, I argue, does not necessarily provide ultimate substantive value for the agents. For now, it suffices to say that *in principle*, coercion toward themselves would be rejected by all agents, and on pain of incoherence within moral reasoning, they must then resort to rejecting it toward others as well. Of course, to justify this, we must prove that *both* reason *and* ontological vulnerability, or finitude, play a constitutive role in our existence (and for that matter, for all living beings), as *inescapable conditions* of our existence that prompt us to moral inquiry. I pretend to side with this statement in the closing chapter of this thesis. If I manage to succeed, I hope to provide a solid explanation as to how vulnerability and reason can ground objective moral reasoning while still accepting anti-realism about value<sup>29</sup>. For now, I shall move onto the next section, which will follow Korsgaard's Constitutivism, an approach that tries to justify the bindingness of moral norms on the grounds of reason alone, as well as the ultimate value of humanity, through the notion of our practical identities as human beings.

---

<sup>28</sup> See Ronzoni (2010, p. 90).

<sup>29</sup> This seemingly counter-intuitive proposal is the central claim of my work.

## 1. Korsgaard's Constructivist Constructivism

In the former sections I laid out the earliest attempts to develop a suitable form of Kantian Constructivism along with the primary difficulties which they have faced. In the case of Rawls, it seems like his own theory, which fluidly evolved over time, could not escape the grasps of relativism by relying only on a high degree of intersubjectivity. It was unable able to ground universally objective norms and instead settled for the agreement of people who already shared the substantive values of democracy and liberalism within the procedure of construction. O'Neill, on the other hand, has been targeted for dealing with too broad and general concepts, while at the same time taking for granted that agents share a need for coordination through reason. This makes it a "goal-oriented" form of constructivism which according to (Budde 2009, p. 218) must require a "(metaphysical and foundational) argument for the value of peaceful and free human existence as well as capacities and capabilities for autonomous action" in the search for objectivity. Indeed, the search for an objective vindication of the bindingness of morality through reason alone has been the toughest hurdle for Kantian Constructivists to overcome. In trying to account for objectivity through universal endorsement, they are charged with moral realism, and in trying to escape moral realism, they seem to fall into at most a high degree of intersubjective validity, which pertains only to modest or minimal objectivity, at most. I have sympathized with O'Neill's acknowledgment of finitude and vulnerability as a factual common denominator for all agents. Beyond this denominator, I have also accepted as necessary that we be *coherent* in our considerations and judgements, acknowledging authorship of our actions. This section will be focused on this precise issue: the role of practical reason and its bindingness for us as rational reflective agents which must be unified and coherent in their judgments.

Christine Korsgaard is a cornerstone of Moral Constructivism, specifically the Kantian variant of Constructivism. Her approach is labelled formalist because it tries to abstain from subjective substantive accounts of moral value and instead takes practical reason to ground formal principles (akin to the formulas of the categorical imperative) that are binding for rational reflective agents simply on the grounds that they unify truly autonomous rational action, which coincides with morality. Contrary to Rawls, where moral norms arise from the deliberation of representatives of rational and reasonable agents, or O'Neill, who relies on a necessity for self-preservation and coordination in virtue of our qualities as finite beings, Korsgaard focuses instead on our character as

*rational* agents, and how the relationship between our *Reason* and our *Identity* binds us to act morally, establishing parallels between Morality and Rationality that make them coincide by drawing influence from Aristotelian Virtue and Kantian Deontology. This is paired with a focus on a theory of agency and action, in the sense that the agent is *constituted* by their actions:

“The principles of practical reason, if they are to be normative, must be principles of the logic of practical deliberation. They must be formal principles.” (Korsgaard 2009, p. 67)

Korsgaard’s constitutivist approach begins from the understanding that agency exists within a spectrum, meaning that there are varying degrees of agency in the world<sup>30</sup>. For an action to occur, there must be in the first place an incentive for it. These incentives are perceived by agents and lead them to act in a particular way. For instance, a plant which turns to the sun is simply responding to the necessity for sunlight. The need for sustenance is an incentive for it to turn, but in this case, there isn’t rational autonomous agency *per se*, because the plant is not aware of its intentionality when moving; it simply responds to the need for sustenance. It simply turns. Non-human animals, on the other hand, can be attributed with intentionality, in the sense that their actions can *fail*. An animal’s perception of another animal as a predator makes it try to run away. It reacts to the stimulus and intentionally takes an action in response. But this form of agency still isn’t “complete.” It may fail, but the incentive in this scenario simply creates a *reaction*. The animal is directly responding to the incentive, so in this sense it is not *autonomous*; we cannot say that it *chose* to run; it simply felt the drive to. Moving along the spectrum, then, we can find and characterize human beings as rational reflective agents, marked by the capacity to be subject to an incentive, stop and reflect upon or question the legitimacy of the incentive and the action it will choose. At this point, incentives become *inclinations* (in the Kantian sense of the word) because there is space between the incentive and our action: the space for *reflection*. We can perceive something as dangerous, and still choose to get close to it. We are reflective and thus have capacity for autonomy in the sense that our actions are our own, they are not entirely predetermined by our perceptions of stimulus. We have a *choice*.

This capacity for reflection, however, comes with a problem. While animals and plants take the world to be as they perceive it and always respond in kind, we are haunted by the

---

<sup>30</sup> See Korsgaard’s *Morality and the Distinctiveness of Human Action* (2006).

capacity for choice. There is a gap between our instinctive reaction and the action we endorse, a division on what kind of grounding we give our actions. But if we are divided, agency is threatened. If our actions seemingly flow arbitrarily between mere reactions to incentives or actual reflective choices, we are not unified agents, and we cannot have our own practical identity. In order to be truly a unified agent, I must always act according to norms I can endorse and justify. If for a certain Circumstance-C I choose Action A, it must be because I believe that for any Circumstance-C, action-A is the correct response. This is what it means to be autonomous. Otherwise, I am simply a slave to momentary impulses and fail to be an autonomous unified agent. Thus, in accordance with the Kantian tradition, we must *will* our action universally and not particularly. Our agency must be autonomously unified, or heteronomously divided, in Korsgaard's view. An agent must be both *autonomous* (the movements must be their own) and *efficacious* (the movements must be in accordance with bringing about the desired end). For these two capacities, Korsgaard argues that the categorical imperative is the constitutive principle of autonomous action, while the hypothetical imperative is the constitutive principle of efficacious action. She takes the act of deliberation to be an attempt to reunite with ourselves, to unify our agency. Normative standards are "the glue" that holds agency together, and if we are to be united with ourselves, we must choose principles which would apply for us in the same exact situations across time and space. Further, in willing universally in space and time for ourselves, we are also in a way willing universally for all rational agents, and therefore, we are already accounting for others since the law ranges over all rational beings. For this reason, laws and norms have inherent public nature. On her account, then, we create ourselves continuously through our actions. We create our own practical identities throughout our lives, and it is up to us to keep them unified. Finally, if we constitute ourselves well (coherently and consistently), we are good at being persons, and therefore we will be good persons. The moral law is the law of self-constitution. The moral law is the rational law, for Korsgaard.

### **3.1 Sources of Normativity, Agency, and Self-Constitution**

Korsgaard's work is fundamental for the purposes of this thesis because it begins by trying to prove *why* human beings – rational, reflective individuals – cannot escape the *need for norms*. The necessity and bindingness of norms for human action, even those which are not necessarily moral in kind, is the central feature of her book *The Sources of Normativity* (1996). It is from the study of the rational need for normativity, combined

with the association between rationality and morality, as well as the argument for a unified practical identity, that Korsgaard tries to vindicate her kind of moral constructivism. Korsgaard's approach, however, is difficult to place within the anti-realist/realist debate. Since it is a constructivist account of morality, she (like Rawls and O'Neill) necessarily subscribes to a form of anti-realism that takes *value* to be a mind-dependent property<sup>31</sup>. Despite admitting that the natural order of the world is teleological<sup>32</sup> and this is how animals naturally and instinctively interpret the world around them (including us), which some strong moral realists use as a way to ground moral mind-independent facts, Korsgaard endorses only a kind of "*procedural moral realism*," in opposition to substantive moral realism. The difference, she claims, is that:

"Procedural moral realism is the view that there are answers to moral questions; that is, that there are right and wrong ways to answer them. Substantive moral realism is the view that there are answers to moral questions because there are moral facts or truths, which those questions ask about." (Korsgaard 1996, p. 35)

It might seem confusing, but in short, Korsgaard is both a value anti-realist and a procedural normative and moral realist. It seems that, according to her, these can both be defined in contrast with the substantive moral realist position that takes value to be metaphysically or naturally fixed upon the world. Anti-realism rejects this notion, and so does procedural realism, which is the view that simply states there are truthful or untruthful norms. To be a normative realist is essentially to believe that when we are faced with practical questions, there are correct and incorrect answers, and when applied to morality, to deny procedural moral realism is simply to be a nihilist about ethics. The value of *good* and *bad* does not exist in the world or metaphysically as an independent property which we as agents must track (despite admitting there is natural teleological organization of the world, it does not assert value as an independent fixed property):

"The denial of procedural normative realism says that there is no ought, should, must, or reason at all. But procedural realism does not require the existence of intrinsically normative entities, either for morality or for any other kind of normative claim. It is consistent with the view that moral conclusions are the dictates of practical reason [...] As long as there is some correct or best procedure for answering moral questions, there is some way of applying the concepts of the right and the good. As long as there is some way of applying the concepts of the right and the good, we will have moral and more generally normative truth." (Korsgaard 1996, p. 35)

We, as rational and reflective agents, have the capacity to reflect and posit value on objects or actions, but at the same time there are standards with which we do so, and it is

---

<sup>31</sup> In Korsgaard 2009, p. 122 she claims value is relational and what it is related to is our nature.

<sup>32</sup> See Korsgaard 2009, p. 114.

in those standards that we may find the objectivity on which Korsgaard intends to ground morality. There are practical problems we face, and from the application of our reasoning to these practical problems we can come up with correct or incorrect solutions. Through normative standards, we can identify moral standards.

“Values are not discovered by intuition to be ‘out there’ in the world. Good maxims are intrinsically normative entities, but they are also the products of our own legislative wills. In that sense, values are created by human beings. [...] The form of realism I am endorsing here is procedural rather than substantive realism: values are constructed by a procedure, the procedure of making laws for ourselves.” (Korsgaard 1996, p. 112)

Having established that there are normative standards for correctness and that our reflective nature allows us to assess the value of the maxims we create for ourselves, it must also be explained how these standards are necessarily binding for us. This pertains to our reflective rational nature, which sets us apart other beings in that there is space for questioning the motives and incentives for our actions, like I mentioned before. We have a certain kind of freedom from our impulses that animals, for instance, do not. To make this clear, let us take a look at how Korsgaard describes the structure of an action:

“Both animal actions and human actions involve the interaction of two factors, an incentive and a principle. The incentive is a motivationally loaded representation of an object. It presents the object as desirable or aversive in some specific way [...] The principle determines, or we may say describes, what the animal does, or tries to do, in the face of the incentive” (Korsgaard 2009, p. 109)

This description perfectly encapsulates the spectrum of animal agency and intentionality I introduced in the beginning of this section. The difference between animals and humans is that while an animal feels the incentive of fearing a predator and displays the intentionality of running away (the principle of the action), this response is automatic in the face of the incentive. On the other hand, however, human beings have access to a kind of reflective “space” between the incentive and the formulation of the principle, because we can question the incentives themselves. Self-consciousness turns incentives into *inclinations*<sup>33</sup> and “*we choose the principles of our own causality*” (Korsgaard 2009, p. 110). This fundamental difference explains why we are, in a Kantian sense, beings charged with autonomy and freedom. We have a special kind of reflective self-consciousness which gives us the capacity to *make* our own laws, while for other animals, their instincts *are* their laws.<sup>34</sup> For animals, their instincts are their reasons, their *grounds* for action. There is no room for questioning. The world is structured as it appears

---

<sup>33</sup> See Korsgaard 2009: p. 120.

<sup>34</sup> See Korsgaard 2009: pp. 109, 115.

to them, in the form of several incentives on which they act. Humans, on the other hand, *must* make sense of the world, because inclinations serve only as *potential grounds* for action. Presented with choices, we must convince ourselves, give ourselves *reasons* as to why we must act in this or that particular way. In sum, we have the:

“Unique human capacity to take responsibility for ourselves, to give shape and form to our own identities or characters, and to make laws for our own conduct. [...] Morality so regarded is one manifestation of the human capacity for what I am calling “normative self-government” [...] our capacity to assess the potential grounds of our beliefs and actions, to ask whether they constitute good reasons, and to regulate our beliefs and actions accordingly.” (Korsgaard 2010, pp. 7-8)

Furthermore, to have this capacity, to be free and autonomous, is simultaneously to be able to recognize ourselves as *causes* in the world. We have agency, and we are aware of our agency, and therefore we *know* that we are responsible for the consequences our actions have on the world. That is why we question the legitimacy of the inclinations which prompt us to act. We know we are responsible, and therefore our actions, in the form of laws, must make sense to us. We must be able to *endorse* them because we are their *authors*, and as such, we are *accountable* for them. Thus, in acting, we are simultaneously creating our own practical identities, according to Korsgaard, and the necessity for justification, the necessity for normativity, is inescapable:

“When you determine yourself to be the cause of the movements which constitute your action, you must identify yourself with the principle of choice on which you act. [...] You regard the choice as yours, as the product of your own activity, because you regard the principle of choice as expressive, or representative, of yourself – of your own causality.” (Korsgaard 2009, p. 75)

This, of course, also attaches a problem which no other animal currently has: the problem of self-constitution.

### **3.2 Practical Identity, Self-Constitution, and the Moral Standing of Others**

In the previous section I elaborated on Korsgaard’s metaethical stand regarding procedural normative and moral realism and how it connects with our unique capacity for self-reflection, culminating in the capacity (and necessity) to act based on reasons, thus creating norms for ourselves. This next step will elaborate on why in doing so we *must* will on universal rather than particular grounds. As I mentioned, for Korsgaard, acting is simultaneously a process of creating and establishing our practical identity. We, as agents, are the product of the actions which we choose to *endorse*, and that is why we strive for justification and establishing their legitimacy. At the same time, Korsgaard argues that in doing so, we must invariably reject “particularistic willing.” What does this mean?

According to Korsgaard<sup>35</sup>, principles can be categorized into 1) general, when we think they apply to a wide range of similar cases, 2) absolutely universal, when we think they apply to absolutely every case of a certain sort, but all the cases must be exactly of that sort, 3) provisionally universal, when we think they apply to every case of a certain sort, unless there is some good reason why not. Conversely,

“Particularistic willing would be a matter of willing a maxim for exactly this occasion without taking it to have any other implications of any kind for any other occasion. You will a maxim thinking that you can use it just this once and then so to speak discard it; you don’t even need a reason to change your mind” (Korsgaard 2009, p. 75)

Given that we are self-reflective beings, autonomous and free, we have what Korsgaard calls, a capacity for “self-determination,” and this, says Korsgaard, requires that we identify and endorse the principle of choice on which we act. But particularistic willing is always dependent on the contingencies of our predicaments, it is based on hypothetical maxims, eliminating the distinction between the agents and the incentives on which they act. If we act based on a principle concerned only with the present situation, without thinking it should also guide action in similar situations, the action becomes lawless, since it is simply responding to the impulses we are presented with at the time. At the same time, if we don’t find ourselves capable of endorsing the principle on which we just acted in future occasions, then we are *failing* a reason which holds, a reason which we can resonate with later. In short, we are failing to establish our practical identity because we lose coherence in our judgments; if we don’t will with universal normative force, we are not committing ourselves to anything at all. Our maxim must, therefore, be willed as a universal law for every rational being and that has normative force for all rational beings, or we lose our autonomy and indulge in heteronomy<sup>36</sup>.

The picture presented thus far is the following: We are beings capable of reflection. This means we are able to question the grounds on which we act. In reflecting, we can recognize standards of validity for our considered judgments on the incentives for our actions and choose which actions (understood as having inclinations and a principle) we want to endorse. In doing so, we are establishing our practical identities as autonomous beings capable of self-determination. We recognize ourselves as authors of our actions, and therefore as agents, and we must strive for universal validity (even if provisional) to fulfill our autonomy and establish a coherent practical identity. If my reasoning today

---

<sup>35</sup> See Korsgaard 2009: 73-74.

<sup>36</sup> See Korsgaard 2009: 76, 79-80.

does not have the capacity to hold for me in the future, I will not be a unified agent. Normative principles “*bring integrity and therefore unity – and therefore, really, existence – to the acting self.*” (Korsgaard 1996, p. 229).

Again, the constitutive standards for action are *efficacy* and *autonomy*, which correspond to Kant’s two imperatives, hypothetical and categorical, respectively. Because when we act we are rendering a change in the world, we must be conscious that our actions have consequences, and part of our agency rests on our efficacy in bringing about our intended action. Efficacy is achieved through hypothetical imperatives with the structure “if I will an end, I ought to will the means to that end” (if I want to quench my thirst, I must drink something). At the same time, the categorical imperative, which requires the universalizability of the action, assures autonomy from our natural desires, ensuring that our principle for action is a norm which is objectively and universally endorsable. In universalizing the norm, making sure that my future self will also be able to endorse it, I am at the same time making sure that any other rational reflective being like me, can endorse it. But this does not exhaust what Korsgaard has to say on the matter. Our reasons to act morally in a rational sense come from our practical identities. These are the identities we acquire and hold throughout our lives, like being a mother or a professor. We have *reasons* for acting on norms related to those identities, as do other persons. But out of all of our practical identities are contingent except for one:

*“Your contingent practical identities are normative for you only insofar as they are endorsable from the point of view of your human identity [...] you come to see them as various realizations of **human possibility** and **human value**, and to see your own life that way: as one possible embodiment of the human”* (Korsgaard 2009, p. 212).

The practical identity of being a human agent as a valuing creature is what gives value to our other practical identities. We can endorse hypothetical imperatives whose reasoning comes from a contingent practical identity we value, only because we value that identity. This addition Korsgaard’s justification of Kantian constructivism comes in light of the criticism that acting only on universalizable norms would deprive us of private reasons and private identities. In Korsgaard’s view, such a criticism does not apply, because when we have our own projects, we posit value on them and invite others to share those same values. But to be able to value our own projects we must value our own humanity, and in doing so, necessarily value the humanity of others. This brings her moral constructivism reasoning to the last stop. If our projects are only valuable because we value them, and value is attributed by valuing creatures, we must endorse this value

among our peers in public fashion. In a fashion of regression, since for Korsgaard value is posited in the world by our own rational will, we must regard ourselves and our rational will as valuable in order to regard our ends as valuable, and when we do so, we value humanity itself as the characteristic which makes us valuable. In this sense, the practical identity of being human acts as a “source practical identity,” the one that grounds the value of all others in an optic of regression. This understanding necessarily attributes others equal moral and rational standing when deliberating about values, ends and actions or principles of action. In short, if we are to act, we must endorse reasons for our actions. If those reasons are based on our private practical identities, we need only be efficacious through the use of hypothetical imperatives. But if we are to value our private practical identities we must value our practical identity as human agents, and also recognize the value of this quality in other human agents. Thus, when positing moral value, they have equal worth, and our moral reasoning must be autonomous from our desired ends and take the form of universal willing by means of the categorical imperative. In other words:

“In order to function practically, we must regard ourselves as unconditionally valuable qua rational agents. [...] If we must regard ourselves in this way, simply because of our possession of rational nature, then we are equally committed to so regarding other rational beings. [...] On this interpretation of the argument, the value of rational agents comes into being as a practical commitment on the part of the valuers who are also rational agents.” (Watkins & Fitzpatrick 2002, p. 359).

### **3.3 Issues with Constitutivism and Further Considerations**

Watkins and Fitzpatrick (2002) suggest that Korsgaard’s argumentation in favor of the value of humanity is lackluster, because it seems to be counterintuitive to our predisposed moral considerations. It seems to take the wrong route because it establishes the baseline for valuing other agents with the agent itself in first place. The route of justification of the value and moral standing of others is directly dependent on the acknowledgement that our own ends and reasons are valuable *because we as humans* are valuable. They argue that this does not seem right when we think about arguments against slavery, for instance, that rest on the so-called dignity of other beings as a value of their own which be preserved and seems intrinsic to them:

“It seems far truer to our moral experience just to say, from both the normative and the meta-ethical perspectives, that the wrongness consists directly in the fact that the action egregiously fails to respect the dignity and value of the victim as a rational being whose nature directly makes such treatment by any moral agent capable of recognizing the value in question inappropriate.” (Watkins & Fitzpatrick 2002, p. 362)

They propose a more realist approach to the value of humanity where it is accepted, even by Kantian standards, as something intrinsically valuable due to their rational agency. Something “indemonstrable” but that we can acknowledge and thus accept, not based in God or any other metaphysical entity but rather that simply exists without the need to be constructed. While this may seem intuitive, I have my qualms about accepting the kind of argumentation offered, as does Sharon Street. The argument for slavery raises an important debate. Korsgaard’s argumentation relies on firstly identifying our commitment to our own value (and therefore to value *our* humanity) and then acknowledging others deserve the same degree of value, due to their humanity. One could argue, *contra* Williams and Fitzpatrick, that there is nothing wrong with Korsgaard’s logic, since through her proposal, we simply acknowledge that we don’t have the right to hurt others or undermine their moral standing in comparison to ourselves. We simply cannot put ourselves above them, and that seems reasonable, and still a good enough argument against slavery. There are however two points to be made in regard to this. First, there is the problem that one must first value themselves and be able to extend this to others. Second, this applies logic to us, human agents as equals and peers of the same “species,” but what if there was another kind of being who deemed themselves superior to others, for instance due to being capable of higher forms of cognition and therefore more deserving of moral status? They value their own judgments above ours because their intellect is superior and take their considerations to hold superior rational value. If these beings deemed us unworthy of moral status due to our rational inferiority, like we often deem animals undeserving of similar moral status, it seems that reason itself and the constitutivist criteria fall short of being able to ground objective morality and moral norms. Something else must be added, and it does not seem like it can be the necessity of constituting oneself as agent.

There needs to be something else entirely that propels reason to reflect upon the moral status of things, from a constructivist point of view. While reasoning gives us means to act intentionally and reflectively, granting us moral status as agents, there must be something which limits the ways in which we can apply instrumental reasoning towards others and stimulates moral questioning by granting moral status as subjects to others. I will argue that the facts of finitude and ontological vulnerability can indeed serve this purpose. The question is whether this brings constructivism closer to a kind of weak moral realism or not. I will argue that these notions are in fact compatible with a constructivist

approach that nonetheless can identify objective moral norms without giving up anti-realism about value. These *facts* of existence are inescapable by agents and the actual driving force for morality, and perhaps Bagnoli was right when saying that practical reason acts as a constraint which produces justification for morality but does not have a foundational role itself<sup>37</sup>; rather, I argue that perhaps this foundational role pertains to finitude and ontological vulnerability. Before moving to this issue, however, it is time to deal with Street's qualms with Kantian Constructivism.

## Chapter II

### Humean Constructivism

#### 1. Asserting Anti-Realism Against Kantian Constructivism

I have thus far described the general Kantian framework Moral Constructivism where different approaches tend to look for the justification of objective moral norms grounded in human agency and value, marked by rationality, reasonableness, and coherence. In the works of Rawls, O'Neill and Korsgaard, norms need to be grounded on reasons which ought to meet a criterion of *publicity*<sup>38</sup> - that is, they need to be *shareable* and *acceptable* when shared – if we are to consider them objective. When such reasons are publicly accepted, or would be publicly accepted by all, they claim to have objectivity. The difficulty, however, in trying to find norms which fit the universal acceptability criteria and vindicate them in a purely formal sense, has promoted criticisms of the kind that they either rely on idealized versions of human agency and cognition, dismissing the importance of human individuation and the “selves” constructed throughout our along with our needs and wants in the particular contexts of our lives. In summary, Kantian constructivism's crusade for objectivity completely dismisses contingency and each agent's substantive set of values. Sharon Street's rejection of Kantian moral constructivism rests precisely on such reasons, since human deliberation cannot shed what makes each agent unique. This led her to put forward her a Humean moral constructivist approach. According to Street, Kantian constructivists try to derive

---

<sup>37</sup> See Bagnoli 2014.

<sup>38</sup> Expression advanced by Carla Bagnoli's *Kant in Metaethics: The Paradox of Autonomy, Solved by Publicity* (2017) which offers a great overview of the significance of publicity for Kant, which I take to also matter for the Kantian contentions of constructivism mentioned above.

morality from pure reason or “empty” formalist principles and argue for its normative bindingness on all rational agents<sup>39</sup>. Alternatively, she claims that there can be no bindingness of this sort since each agent’s evaluative viewpoint is different from one another. Despite this, her alternative Humean approach shares some notions with the Kantian framework of moral constructivism<sup>40</sup>. In both cases moral norms are derived from the reflective thinking of rational agents from within a “practical standpoint.” Normativity is produced through practical reasoning to tackle practical problems they face. The key difference seems to be the kind of answers to be expected from those agents, depending on each author’s depiction of what the practical standpoint should consist of:

“The debate between Kantian and Humean metaethical constructivists – concerns whether moral values (or for that matter any other particular substantive values) are entailed from within the practical point of view as such, such that no matter what one’s specific starting set of values, it turns out to be entailed by those values, in combination with the non-normative facts, that one has reason to be moral.” (Street 2010, p. 379)

While Kantian constructivists believe the practical standpoint itself provides substantive content for the construction of the norms through application of pure practical reason impartially which leads agents to objective moral norms independent of their will and beliefs, a Humean constructivist like Sharon Street admits that the substantive content must be brought by the agents themselves into the standpoint, and thus does not take the bindingness of morality for granted. According to the Humean constructivist, an empty practical standpoint without the subjective substantive set of values brought by each agent as postulated by the Kantian constructivist should lead nowhere. “Pure practical reason” cannot bind all agents into adopting specific moral values, not even a commitment to the value of humanity as postulated by Korsgaard. Street’s qualm is that this argumentation presupposes we are all independently bound to having to value ourselves as valuing agents when this is not the case. Resorting to the example of a social reflective insect who is endowed with all aspects of reasoning and valuing as a human but *does not value itself non-instrumentally*<sup>41</sup>, Street tries to convey that it is possible to *value something* without having to *value ourselves*. Such an insect may only value the life of its queen, and thus value itself only insofar as it can benefit the queen. Nothing here requires that the insect

---

<sup>39</sup> See Street 2016, p. 164.

<sup>40</sup> Detailed descriptions of the ways in which they interlap can be found in both Street’s *What is Constructivism in Ethics and Metaethics?* (2010) and *Coming to terms with contingency: Humean constructivism about practical reason* (2012, p.41-42).

<sup>41</sup> See Street 2012.

holds intrinsic value for its own life outside of this instrumental dimension. Such a criticism nullifies Korsgaard's modern justification of the Kantian Formula of Humanity regarding the value of humanity as a source practical identity with intrinsic value.

From this rebuttal, the Humean constructivist acknowledges that at some point in our regressive search for normative validity, in never finding one objective and universal source of value, we are left to invariably reach a point where we must choose a set of values to endorse, and Street argues we would therefore have no choice but to resort to and endorse our own contingent substantive sets of values. In the case of the social reflective insect, it does not have to endorse its own value, only the queen's. In similar fashion, Street argues, we humans would probably endorse our own value or the value of others, but nothing dictates this beyond the own valuing attitude of the agent. Thus, the conclusion is that for Street there is no independent standard of normative value; rather, each of us endorses specific sets of values from within the standpoint, endorsing reasons to choose those values in detriment of another. This is what it truly means for her to be an anti-realist about value and to accept contingency. In other words, Street accuses Kantian constructivism to indulge in the very thing constructivism should not: moral realism<sup>42</sup>. By trying to argue for the objective value of Humanity (or in a broader sense, the value of the valuing creatures themselves) as the common denominator, Kantian constructivism depends on something "valuable" independently of the valuing creature standpoint, thus indulging in moral realism, the very thing moral constructivism should reject. For Street, if the social insect as a valuing creature does not value itself, then it should not be forced to accept its own value as more than merely instrumental. After all, value is something *created* and *attributed by the valuing creature in the practical standpoint*; if the valuing creature does not attribute or want to attribute value to something, then that thing simply is of no value to it.

## **2. How To (Try To) Vindicate Objectivity Through Constructivism**

This is not to say one should give up on trying to find objective moral norms, as Street clearly states that "*the objectivity of ethics would be vindicated ... if we could somehow show that there is a universal problem, faced by every individual agent, to which morality is the universal and best (or only) solution*" (Street 2016, p. 166). The issue, however, lies in proving that all agents who enter the practical standpoint with their own set of

---

<sup>42</sup> See Street 2012, p. 56.

substantive subjective values will share such a universal problem, and further, the answer to said problem. On this matter, her best attempt at an answer identifies the universal problem as “*the problem of a mismatch between the world as it is and the world as that agent would like it to be,*” or what she calls “*the problem of attachment and loss*” (Street 2016, p. 166). This means that *if* from the practical standpoint every agent could agree that there is a problem for which morality is the answer, we could vindicate ethical objectivity. Of course, for her, this is a challenge that seems too demanding and hard to meet. Still, and despite the criticisms addressed to Kantian constructivism, she grants that constructivist attempts at vindicating some form of moral objectivity have merit of their own, and gives an account of their general structural arrangements and strategies:

“(A) According to metaethical constructivism, the objectivity of ethics would be vindicated, without metaphysical and epistemological mystery, if we could somehow show that there is a universal problem, faced by every individual agent, to which morality is the universal and best (or only) solution.

(B) A plausible candidate for a universal problem faced by every individual agent is the problem of a mismatch between the world as it is and the world as that agent would like it to be.

(C) It’s at least *possible* that there is a universal, best solution to this problem.

(D) It’s also at least *possible* that the solution mentioned in (C) somehow necessarily involves taking up an ethical perspective in the world.” (Street 2016, 166)

Accordingly, a constructivist then takes three steps in their approach, according to Street:

“Step 1: Identify some (more or less widely shared) evaluative point of view that is of interest. Identify the point of view by saying what is constitutively involved in occupying it.

Step 2: Identify some (more or less widely shared) problem that faces anyone who occupies the evaluative point of view identified in step 1. (...)

Step 3: Identify the solution (or solutions) to the problem identified in step 2. This will require substantive argument to the effect that some particular solution S is the best, or only, way to solve the problem identified in step 2.” (Street 2016, 168-169)

She then proceeds to describe how this structure fits within the Rawlsian and Korsgaardian approaches. While Rawls’s step 1 – evaluative point of view of interest – pertains to a restricted set of agents who already share evaluative concerns and dispositions (as well as a similar context), Korsgaard tries to argue for something which holds for all agents independently of their evaluative viewpoint. This leads to the labeling of Rawls’s constructivism as “restricted” and Korsgaard’s as “maximally ambitious”, with the latter defined as that of “(*Constructivist Objectivity*) *There are some reasons that*

*every agent has no matter what the particular substantive content of his or her starting set of evaluative attitudes*” (Street 2016, p. 173). Even though Street disagrees with the strength of Korsgaard’s proposal, she believes the general strategy for vindicating the objectivity of ethics is helpful and can be summarized as follows:

“To vindicate the objectivity of ethics, try to show that morality is somehow built into the structure of agency itself, such that if you qualify as an agent at all, then necessarily the problem (and therefore also the solution) is yours (by your own, evaluative lights) [... further] no matter what the particulars of a given agent’s evaluative point of view, he or she faces the problem of attachment and loss” (Street 2016, p. 174).

The shift from one approach to the other is that essentially now we are no longer trying to vindicate one absolute value of the agent within the standpoint but rather to recognize that the problem is invariably present for any and every agent, despite their contingent substantive set of values. If this is proven to be true, then every evaluative standpoint faces said problem. In her own proposal, this universal problem would pertain to the problem of attachment and loss, a gap between how the world is and how we as valuers would want it to be. Still on this regard, one can imagine invulnerable valuers who would not face the problem of loss:

“There are at least two types of invulnerable valuers we might imagine. The first type is *valuers who are invulnerable in virtue of being omnipotent*. I will set this type aside because there are no such valuers. Every particular conscious point of view on the world, and *a fortiori* every valuer, depends for its existence on a particular organization of matter that will not persist indefinitely; it is therefore subject to accident, death, and more generally circumstances beyond its control. I take this to be a deep fact about the nature of the universe, as opposed to something that merely happens to be true of the forms of conscious life we are familiar with.” (Street 2016, p. 182)

The second type of invulnerable valuer would be a someone who is invulnerable in virtue of the nature of their ends; the things they value themselves are invulnerable, for instance if they valued that  $2+2=4$ , whatever they value would not cease to be true. But on this case, while they are invulnerable valuers, they aren’t much of an agent, since for them:

“There is nothing at all that needs doing. The world is normatively perfect just as it is. Thus, to the extent that an agent (as opposed to a mere valuer) is a being who is subject to reasons to act – a being for whom there are normative reasons to do things – we may have secured the conclusion that all agents are necessarily subject to the problem of attachment and loss.” (Street 2016, p. 183).

Street uses these two types of valuers to prove that invulnerable agents “*are not part of the audience that a vindication of ethics needs to address*” (Street 2016, p. 183). The problem of attachment and loss itself, then, is the form of a problem that arises for an agent derived from his/her most urgent substantive problems. In other words, “*whatever*

*force his or her problems have from his or her point of view -that is the force of the problem of attachment and loss for that agent”* (Street 2016, p. 183). At this point, if we have rejected that the audience for a vindication of ethics does not encompass either valuers who are omnipotent (infinite) neither invulnerable valuers, we are no longer trying to claim that morality is *binding* on all rational agents due to their rational capacities, and instead we must try to prove that the universal problem in the evaluative must be universal for all *finite vulnerable* valuers, and recognize that morality is grounded only on the struggle to tackle the universal problem for all other kinds of agents. In the case of Sharon Street, that would be to close the gap between expectation and reality for rational agents, but I want to move away from her suggestion<sup>43</sup> now, for I believe to have mapped out all the relevant aspects for what I believe to be the problem.

Despite not focusing on the specific problem of vulnerability, instead choosing the problem of attachment and loss, like other constructivists Street ends up resorting to the qualities of “finite” and “vulnerable” to describe necessary aspects of beings for whom the problem of morality arises. In describing why infinite or invulnerable valuers are “removed” from the equation, I find it interesting that she didn’t account for the two hypothetical variables which became “constants” for the universal problem which morality is to address. If infinite and invulnerable valuers would never be “the audience” that ethics needs to address, shouldn’t morality be the answer to the universal problem of finitude and vulnerability? This, in turn, shall be my contention for this work, but before moving to the final chapter, I want to raise an important objection to Street’s rebuttal of Kantian constructivism’s inability to provide substantial claims from within empty formalism. This reply to Street’s claim that Kantian constructivism either cannot justify normative reasons or relies on a mind-independent normative reason to do it is made by Sem de Maagt (2019), and it rests on a “constructivist justification of the principle of instrumental rationality”. Recall how in Chapter II I mentioned how Street does not believe that Kantian constructivism’s empty formalism can create or vindicate substantive claims. This was mentioned in quite a remarkable way by Street herself when claiming that “Humean versions, in contrast, deny that the rabbit of substantive reasons can be pulled out of a formalist hat: to get substance out, we need to put substance in.” (Street

---

<sup>43</sup> I find it curious that her approach rests on adopting a standpoint that transcends finitude and identifies the problem of ‘self with the ‘others,’ which indeed resembles in a way the Kantian line of argumentation, which Street so vehemently criticized before. Nonetheless, that goes beyond the scope of my work. For more on this, see Street 2016, p. 187

2010, p. 370). As it turns out, according to de Maagt, Humean constructivism endorses that contingent instrumental reasoning can provide substantive reasons for action, or claims. The way in which instrumental reasoning works, in short, is that it endorses the means to take an end following the structure of “If I want X, I must Y.” With this in mind, de Maagt argues that there are actually two types of instrumental reasoning: 1) contingent instrumental reasoning, and 2) categorical instrumental reasoning. Contingent instrumental reasoning is explained in exactly the same way as Korsgaard’s hypothetical imperative, and it works by applying the aforementioned structure to satisfy a contingent end that an agent might have. Resorting to the example I gave in Section 3.1 of Chapter I: “if I want to quench my thirst, I must drink something.” Since Humean Constructivism accepts contingent substantive set of values, then it must accept that within the evaluative standpoint agents may resort to contingent instrumental reasoning based on their ends to ascribe moral value. Conversely, the wit of de Maagt’s argument against Street’s claim, rests on there also being categorical instrumental reasoning, which gives agents categorical substantive claims even from an “empty” evaluative standpoint where they may stand:

“There might also be instrumental reasons involved in adopting *any* end whatsoever. For instance, being alive is a necessary means for engaging in any action whatsoever [...] a necessary means for acting, whatever one’s contingent ends, is that one is alive. Even someone who commits suicide has to be alive in order to achieve his or her end (cf. Gerwith 1978, 136-137). In this sense, one has reason to want to be alive, independent of one’s contingent ends.” (de Maagt 2019, p. 285)

Essentially, de Maagt points to the *necessity* of some means to even be able to act in any kind of way, towards any end whatsoever. These categorical instrumental means are not dependent on any contingent ends, and for this reason are capable of generating categorical substantive claims that endorse the value of the capacity for action. Because Street’s Humean Constructivist also rests on (contingent) instrumental reasoning, de Maagt claims she has to accept Kantian constructivism as a coherent project, dropping the “emptiness objection.” While the principle of instrumental reasoning itself is a formal principle, de Maagt says, “*the material content of this principle is provided by the necessary means of action [...] it has material content once applied to the first-person perspective of an agent.*” (de Maagt 2019, p. 288). If de Maagt’s categorical instrumental reasoning argument holds (as I believe it does), he suggests that it can be complemented with either 1) “a further specification of what it means to be an agent and a consideration of relevant, general anthropological insights concerning the abilities and vulnerabilities

of human agents” or 2) “to put forward an additional argument for the claim that one has moral obligations towards (all) other persons. Now that Street’s objection is out of the way, it’s time to move toward the final chapter, where I hope to address both paths.

## Chapter III

### Practical Reason, Finitude and Vulnerability

In this chapter I will articulate the conclusions that follow from previous chapters, in trying to posit a moral constructivist approach which can be both objective and anti-realist about value. The first section will address the aforementioned categories of “finite” and “vulnerable.” Understanding their implications of agency and reasoning will lead me to argue that they sustain a universal problem for the evaluative standpoint of any agent to which ethics pertains to. This, in conjunction with de Maagt’s argument in favor of categorical instrumental reasoning, will help me assert that there is an objective standard of validity for moral reasoning in the form of a principle of non-coercion. Finally, I will address how this can still be considered an anti-realist contention of constructivism.

#### 1. Finitude and Ontological Vulnerability

In the previous chapter I explained why, following Sharon Street’s contention of how moral constructivism could try to vindicate ethical objectivity, the universal problem that is inherent to any evaluative point must be that of *finite* and *vulnerable* agents. Neither infinite nor invulnerable valuers would ever enter an evaluative standpoint from where a problem would arise for them, on top of their existence being taken as a factual impossibility, according to Street. Thus, it is now important to evaluate the ways in which this finitude and vulnerability could vindicate some sort of objective normativity. Carla Bagnoli brings an insightful contribution to this problem and debate in her article “*Vulnerability and the Incompleteness of Practical Reason*” (2016). I mentioned that for Street, the problem of attachment and loss draws its force from the actual problems an agent faces at the time and the substantive content of the practical standpoint comes from the context they are in. It holds up to nothing more than a mismatch between their expectation or desire and the world as it presents itself to them. The higher the mismatch, the higher our expectation or desire for a certain outcome that is not met, the higher the *importance* (understood as *value*) that we attribute to the problem. But Street herself says

that problems of this kind, which could be used to justify morality, could not apply to *invulnerable* valuers, because the two kinds described by her wouldn't ever feel the need to conceptualize action to change anything. Either they are invulnerable due to their omnipotence which makes the events meaningless, or the things they value are invulnerable to change and thus their expectations and desires cannot be motive for action.

Street argues that the problem of attachment and loss is not substantive itself, but rather the *form* of problems which manifest in our daily lives. But on the contrary, isn't our unavoidable condition as finite vulnerable beings – in every sense of the word – itself a substantive matter? In line with de Maagt's argument about substantive categorical reasoning, I believe so. As far as categorical instrumental reasoning is concerned, agents have a categorical substantive claim to the preservation of their ability to *act*. I believe this claim subtly implies that every agent who is finite and rational, understands that its capacity to act rests on such finitude and vulnerability. If on Street the problem of attachment and loss was only formal, and its value and contents depended on the contingent practical problems we faced, so can a parallel with the problem of finitude and vulnerability be drawn. We are limited beings, both in time and space, as well as in intellectual and physical capabilities. These limitations are nothing more than the product of our finitude, and they are exposed to detriment by our own vulnerabilities. On top of the maximum way our substantive claim to action can be violated (death), we can also recognize that the different contingent manifestations of our vulnerability will shape the intensity of the problem. The universal problem of finitude and vulnerability is heightened in proportion with the apparent way in which it may hinder our capacity to act. Shouldn't this consideration itself shape our understanding of the problem of morality and how we approach our interactions with each other and the world? I believe it does. As Bagnoli puts it, "*vulnerability as an ontological category provides the normative standard for identifying distinctive ways in which we function or fail as agents.*" (Bagnoli 2016, 13-14). Bagnoli stresses the importance of ontological vulnerability for normativity because vulnerability itself defines the capacities and limitations of the kind of agents we are dealing with, and specifically what kind of norms they need and choose to opt for. Vulnerability shapes the way in which agents perceive themselves, the world and consequently it shapes the way they interact and relate with one another in terms of expectations and actions, and Bagnoli categorizes vulnerability in terms of constitutive –

the broad sense in which our agency can be affected – and circumstantial – the particular ways in which said vulnerabilities are exposed and endangered given specific contexts and our embodiment (I will address to these as contingent). Thus, vulnerabilities, in one way or another, invariably weigh and pressure agents to act in a certain way. Agency as the capacity to act is ontologically limited, threatened and conditioned by vulnerability in a temporal, social and emotional way. Further,

“When understood as a constitutive feature of human agency, vulnerability is primarily characterized as an aspect of embodiment – a dimension that humans share with other animals. [...] First, it pairs vulnerability to situatedness, hence revealing how human agency is exposed in contextual contingencies. [...] Second, when tied to embodiment, vulnerability makes the temporal structure of human agency apparent. Humans are finite [...] agency acquires meaning and significance only within the framework of mortality. [...] Third, emotional vulnerability allows humans to develop a complex network of dispositions, skills and capacities by which to respond to the predicament of life” (Bagnoli 2016, p. 16)

I argue that underlying the problem of attachment and loss is precisely the notion of finitude and vulnerability, which constitute the true core problem for agents. If we think about it, the only reason why we face a problem of attachment and loss in the first place is because we are *limited* in both our ability to change the world to match our expectations as well as *vulnerable* in many different contingent ways. This is also true for the problem of our valuing attitudes, who are similarly vulnerable, unlike for the invulnerable valuer described by Street. In other words, it is not factual that the world *should* be the way we want it to be, it is only factual that we want it to be that way. Ontological vulnerability and finitude are *facts* of life and existence as such. Vulnerability takes precedence to the attitude of valuing itself because it applies independently of our attitude of valuing. It is an inescapable embodied characteristic that furthermore has direct influence on our capacity to value<sup>44</sup> and to act<sup>45</sup>. I believe this is enough to convincingly claim that indeed the universal problem within the evaluative standpoint is that of finitude and vulnerability.

A point I want to briefly refer to pertains to how the acceptance of vulnerability as a human fact can affect contentions of Kantian constructivism, due to the notion of autonomy being seemingly counter-intuitive to it. This, fortunately, is something which I can do with relative ease due to the several occasions where the issue has been addressed, both for Kantian ethical theory, including constructivism accounts (see O’Neill 1989,

---

<sup>44</sup> Suffering a brain injury, for instance, may affect our cognitive capacities. It is a case of embodied vulnerability being attacked which affects our capacity for value deliberation.

<sup>45</sup> A permanent physical injury, for instance, can hinder our capacity to act in determinate desired ways.

Formosa 2014, Bagnoli 2014, 2016, 2020)<sup>46</sup>. Essentially, vulnerability and autonomy need not be incompatible but rather complementary. Vulnerability shapes practical reasoning due to its susceptibility to temporal, embodied and social constraints on the individual. Vulnerability also implies dependence on others, to an extent, which according to Bagnoli:

“The completion of practical reason does not aim to correct or erase vulnerability, but it exploits vulnerability as a resource to build up autonomy, which is a requisite for moral and rational agency.” (Bagnoli 2016, p. 24)

I side with Bagnoli when it comes to the defense that Kantian constructivism should embrace rather than reject the notion of vulnerability when it comes to the characterization of autonomous agency. Moral constructivism is a project which seeks moral norms from within the practical standpoint of reflective rational thinking. However, we are not perfect rational beings. Even if we were to hone our rational capacities to the fullest and develop perfect reasoning, we would still be limited by the capacity with which we can act. This explains why the principles achieved through pure reasoning won't always be applicable as we'd like them to be, which makes us confront difficult moral predicaments where all choices seem bad. Accepting finitude and ontological vulnerability as undeniable facts of existence can complement pure practical reasoning and function as the universal common denominator which grounds objective morality from within the evaluative practical standpoint. Understanding that each agent has a categorical substantive claim to be able to act, which is always limited and threatened by their finitude and ontological vulnerability, essentially means that, even in a Humean constructivist characterization of the practical evaluative standpoint, there is universal common ground between all agents. Even if there is no procedure to eliminate their individual substantive set of values, the substantive claim to a principle of non-exploitation of their finitude and vulnerability (which, in my view, could also translate to a principle of preservation of agential autonomy in the broad sense of the word) will always be present and act as universal common ground and possibly a starting point for deliberation. Thus, finitude grounds morality, while practical reason builds its exoskeleton. Through the constraints of finitude and reason, morality is born and shaped. This claim is corroborated by Bagnoli:

---

<sup>46</sup> Joel Anderson's *Autonomy and Vulnerability Entwined* (2014) is also elucidative for this matter.

“For finite rational beings, morality necessarily takes the form of bonds and constraints. Norms prove necessary because finite rational agents suffer from limitations. [...] Infinite rational beings are invulnerable but also, and for the same reason, *unconcerned* with the normativity of laws. The question of how to transgress and how to be bound by such laws does not arise for them. Conversely, the *normativity* of the laws of reason is crucial for finite rational beings. As I see it, Kantian constructivism [...] begins with the acknowledgment of constitutive vulnerability as the original feature that gives practical reasoning a point: it is ‘the being rather than simply the rationality of finite rational beings that is limited.’” (Bagnoli 2016, p. 23)

Moral norms are constructed by agents from within a practical standpoint, and in this standpoint, they use practical reasoning to do so. This means their reasoning must obey practical reasoning constrictions, but we must not forget that they are also real agents situated in the real world, where inescapable contingencies apply, namely their own vulnerabilities and those of other beings and things. Finitude is inescapable. Vulnerability is inescapable, and thus must be accounted for. I do, however, disagree with Bagnoli on one aspect, which ties in directly with 2) “to put forward an additional argument for the claim that one has moral obligations towards (all) other persons. It has to do with how I try to vindicate anti-realism about value despite the commitment to an objective moral standard for the construction of norms, through the notions of moral agency and what I call moral “subjectness.”

## **2. Moral Agency and Moral Subjectness**

There are some important implications from what I have already sketched throughout this chapter. First, in line with Street’s anti-realism, I argue that morality is not *necessary*. Nonetheless, in Bagnoli’s last quoted statement, I included a segment where she states that “*infinite rational beings are invulnerable but also, and for the same reason, unconcerned with the normativity of laws,*” and it is from this claim that I want to depart.

I contend that is not granted that every single being in this universe will resonate with the concept of morality or moral norms. Indeed, through the logic described for the vindication of moral norms through the recognition of finitude and ontological vulnerability, it seems that *invulnerable infinite* (rational) beings may not find the need to abide by such a principle of non-coercion, simply out of their capacity to be above the notion of vulnerability. Through this logic, because the problem of vulnerability does not apply to them, they would never be motivated to pursue moral reasoning. They cannot be hurt, affected, or proven wrong, they cannot be damaged or “wronged” in any way or capacity by other beings and therefore they may see no need for morality. In a sense,

Bagnoli's reasoning is that they would never abide by moral norms because they will never be *moral subjects*. However, I want to draw a line between what defines moral agency and what defines moral "subjectness" (I will refer to this term to avoid confusion with the notion of subjectivity). I believe the distinction between moral agency and moral subjectness is necessary if we are to truly understand whether morality is *binding* or not. Throughout the study of all variants of moral constructivism in this work, it seems clear that there's no question: "*Finite rational beings are vulnerable because they are finite, but they are agents insofar as they are endowed with reason.*" (Bagnoli 2016, p. 22) Indeed, the capacity for moral action comes from reason, from the reflective nature the being who, as a result of questioning the legitimacy of the reason or principle for their action, becomes an agent. But there is more than one rational way in which a being can endorse a reason, and I argue, if one ultimately subscribes to anti-realism about value (which I intend to), the reason that comes out on top of the deliberative process and truly motivates action is strictly dependent on the valuing attitude of the being. This may seem confusing, so I'll provide an example that hasn't yet been mentioned but is used by Street to show the difference between the two brands of constructivism I have been working on: Caligula, a being who values the activity of torturing others for fun, despite being fully aware of their pain. As much as it pains me to admit, my view on Caligula's rational consistency changed throughout this work. While I first sided with the contention (commonly associated with Kantian constructivism) that Caligula is rationally inconsistent, I am now going to argue that they, in fact, aren't. This seems problematic, but I am also ready to assert their immorality, so bear with me. The claim that Caligula is rationally inconsistent pertains to the logic that, were the positions reversed, they wouldn't want to be tortured, and thus, by pain of incoherence, should refrain from torturing others. This is instinctively and somewhat true, I agree, but Street clearly states that they are aware of non-normative facts, namely the effect of the torture on the victims, or – in this framework – the victims' inescapable finitude, vulnerability, and their substantive claims against being tortured. Caligula is also aware of their own finitude, vulnerability, and that they would also try to protect themselves from torture by making a substantive claim to non-coercion were the positions reversed. So far, there is nothing irrational, or unaware about normative reasons in Caligula's understanding of the situation. The problem is, they *don't value the claims of others more than they value their claim to torturing because it's fun*. Sure, Kantian constructivists will argue that torturing because it is fun isn't a good reason for action, since it doesn't respect the value of human

life and also because it's heteronomous reasoning instead of autonomous. And in their framework, it is, but it does not make this irrational. Caligula has a reason to act, which they value, and the ones being tortured have a reason to not want to be tortured, which Caligula doesn't value. Even if Caligula is aware of every other person's claim to be safeguarded from violent coercion, Caligula simply doesn't value others or their reasons above their (Caligula's) own. I claim that this is a "problem" of value-attribution and the kind of reasoning that prevails above others. Caligula's predominant valuing of their own fun makes their reasoning selfish and instrumental, but rational, nonetheless. Conversely, it is valuing how our actions affect others that prompts moral reasoning.

The Kantian constructivist's mistake is the mistake of thinking that moral requirements must bind us independently of the *particular evaluative nature* with which we find ourselves – and in particular, independently of whether we already have moral concerns as a deep part of our nature. But the right view, according to the Humean constructivist, is that moral requirements do not bind us irrespective of our particular evaluative nature. In particular, if one lacks moral concerns altogether, then morality does not bind one. But if one is a moral agent, as opposed to just an agent, then part of what that involves is taking oneself to be bound categorically (in certain cases) with respect to what one feels like doing, what one finds pleasant and attractive, and so forth." (Street 2012, p. 56)

Let me try to clarify further by returning to the example of the infinite invulnerable rational being and the claim that it will never concern itself with morality on the grounds that it won't be ever affected by anything. My claim is that, despite not being affected by anything, the notion of authorship of its actions along with its rational capacity *can* prompt a disposition for moral reasoning and action. Through the exercise of their rationality and self-reflectiveness, if they decide to think about whether their actions are correct or wrong, they can indeed become *moral agents*, because they are *concerned with other moral subjects*. This is because *reason* attributes moral status as an agent (an entity with capacity for moral action), while *finitude* and *vulnerability* attribute moral status as subjects (entities which are on the implicated receiving end of moral action, passive in this sense). Thus, despite being infinite invulnerable beings (therefore not moral subjects) but rational and aware that their actions have consequences for other beings who are finite, these beings can question whether their actions are legitimate, and if they value the moral and normative standing of others, they will indeed feel the bindingness of morality, due valuing the categorical substantive claim of finite moral subjects to not have their finitude, or vulnerability, exploited.

Human beings, so far, are the only beings that we know of who – due to their finite and rational nature – can *both reflect* on this issue *and* at the same time *be wronged*, thus,

holding the moral status of both *agents and subjects*. On the opposite ends of the spectrum, we can find *infinite invulnerable rational beings*, who are *moral agents but not moral subjects*, and *animals*, who so far have not explicitly been able to communicate a capacity for conscious reflective moral reasoning, but are finite and vulnerable, thus holding the status of *moral subjects but not moral agents*. In short, I argue that on one hand, reflective reasoning grounds moral deliberation and agency, but finitude and vulnerability ground moral “subjectness.” It is through the weaving of these characteristics or factors, that morality is born. Thus, one may ask: well, but how can a moral norm be binding to a moral agent who is not a moral subject? And the answer lies within this delicate conjunction. An infinite invulnerable rational being cannot be *wronged* but it *can wrong others*, and as soon as it realizes this, morality’s seeds begin sprouting in its cognitive capacity, giving it *reason* to indulge in moral reasoning. But this is not strictly binding, since of course, it can choose to ignore the value of others or their normative stand in relation to his own. In Bagnoli’s *Practical Knowledge, Equal Standing, and Proper Reliance on Others* (2020), she argues in favor of a criteria of normative parity, which resembles the Korsgaardian notion that we cannot escape being reasoned with. When one being gives another a reason to do or not do something, the latter must either give themselves a reason which trumps the former or accept it. In my view, this is not incompatible with the anti-realism about value I’m advocating for. In the case of Caligula, for instance, the others’ claims reach the recipient, but they give themselves a reason antagonistic to the first ones (the reason being that they value torturing more than they value the normative standing of others). In doing so, they are not eliminating the objective moral criteria we have elaborated through the substantive claim born out of finitude. They are simply not valuing the moral status of others as subjects.

This is because moral objectivity pertains only to those who are indulging in moral reasoning, since moral reasoning implies a valuing of the other’s moral status and thus moral normative claims. If one does not value such a thing, one is simply reasoning; reasoning in a way that is not moral reasoning. For instance, Caligula can hope to justify to their victims that they are torturing them because it’s fun; in doing so, a reason is being given to justify the torture. Of course, unless the victims enjoy torturing, they won’t value Caligula’s enjoyment, and won’t accept their reasoning. What’s at stake here is the mutual non-agreement on the value of each other’s reasoning since one is selfish reasoning and the other is moral. As much as Caligula tries to justify their actions to others, they will

never hold as *morally justified*, because *there is a standard and principle for moral constructivism*. However, Caligula does not care about the opinions of others, and may feel justified in their actions simply through instrumental selfish reasoning. The victims are vulnerable, finite, they have substantive claim to moral status, and if their vulnerability is exploited, they are wronged. In moral reasoning, there is nothing that supplants this. But this is not binding on any living creature that has free-will on an anti-realist account of value, since every creature who has free-will can choose to not value morality, dismissing moral bindingness, even though it can never justify that breaching of another creature's moral status and integrity is not wrong for the other creature, and therefore morally wrong.

Now, another interesting question arises. What if the victims enjoy being tortured? Let's suppose both Caligula and the victim endorse and agree upon the value of engaging in bodily mutilation on opposing ends. In this sense, could the action of torturing be morally acceptable? Let's look at it and at the case of a person who wants to commit suicide simultaneously. On the account sketched thus far, it is stipulated that every finite vulnerable being has an intrinsic, inherent substantive claim (not intrinsic value, but intrinsic claim to be valued) to not having their embodied and intellectual autonomy violated. This substantive claim is entailed within *every* agent's substantive set of values as *just another* substantive value among other (contingent) ones. Hence, on this setting, within the evaluative standpoint, the being has claims to both this categorical substantive value other the contingent ones, but in being committed to anti-realism, we accept that the being may endorse a contingent value of their own above the intrinsic claim to autonomy, and this too is a manifestation of his autonomy in the stipulated sense. The only thing the universality of the categorical substantive claim guarantees is that *another* agent should, in moral principle, refrain from infringing upon another's autonomy. It acts as a passive (in the sense that it continuously holds without having to be called upon) protective principle which, in situations of potential violation, can be evoked by the subject to protect themselves normatively. This in no way seems incompatible with the subject's surrendering of such a claim. If the subject themselves actively choose to dismiss this claim to pursue a contingent reason which they value more, this too is an expression of autonomous free-will, in the anti-realist sense. To assume the agent would not have the power to give up this value would be to incur in moral realism about objective value of the person, which is not the case. Values are created by the person, and as such

may also be dismissed by them. I hope these extreme examples don't confuse the commitment to an anti-realist account of value with the relinquishment of moral objectivity. To be clear, the universal access to a categorical substantive reason to not have our finitude and vulnerability exploited means that this holds objective *moral* value in principle. But this value is posited by both the facts about finitude and vulnerability as well as the value-commitment to moral reasoning. This kind of objectivity pertains only to the domain of morality, where a prior commitment to the value of morality has been endorsed. The principle of non-coercion is objectively valuable *in* the moral domain, but not in a merely rational domain of reasoning. This is because what's moral is also rational (due to being born out of practical reason), but what's rational isn't necessarily moral.

Finally, having sorted the question of value, I want to elaborate on how the process of moral construction can be influenced by the principle identified. In cases of moral predicaments such as the age-old Kantian problem of the killer who knocks on the door to kill the friends hiding in the agent's house. How can constructivism of this kind solve the problem? Assuming commitment to moral reasoning, the agent who opens the door realizes the killer wants to exploit the physical, and consequently moral integrity of their friend. By lying to him, the agent is exploiting the killer's vulnerability to being lied to, therefore *wronging* him, but in telling him the truth about the friends' location, the agent would also be *wronging* the friends, by aiding the killer to end their existence. I will argue that this kind of constructivism offers the tools to deal with the situation in a seemingly acceptable way. Because we are dealing with moral constructivism and not Kantian deontology, despite there being a principle of non-exploitation of vulnerability, we are not charged with absolute duty not to kill or lie. In this sense, we must weigh our options and their consequences in terms of vulnerability. Since without finitude, without consequences, there would be no morality, weighing the consequences against the objective principle of morality seems acceptable. We *must* weigh within our practical and evaluative standpoint, those which we know to be the immediate consequences of our actions and decide on what course of action lessens the degree of vulnerability exploited. Following this criterion, death completely terminates an agent's existence, it annihilates permanently the capacity to act, and it is the maximum degree of violation of the categorical substantive claim to autonomy (although in wanting to kill another person, the killer has already given up a commitment to moral reasoning, the agent at the door still attributes value to moral reasoning and thus must abide by the principle). In contrast,

a momentary lie that simply stops the killer from pursuing its contingent (and extremely immoral) end at that point, nevertheless preserves the killer's life, respecting and prolonging in time his right to the categorical moral claim to enjoy autonomy as the capacity for further action<sup>47</sup>.

I understand how this kind of reasoning shares O'Neill's theoretical problem about how the conceptualization of vulnerability, for instance, influences the choice of norms endorsed. Ontological vulnerability, as Bagnoli (2016) pointed out, can manifest in a manifold of contingent vulnerabilities, either in the form of embodiment fragilities or the problem of attachment and loss, for instance. How, then can we consistently account for them? Since the approach is anti-realist about value, it shouldn't try to objectively hierarchize norms, instead accounting for flexibility in the considered judgments. Despite this, if I must take a stand, I am willing to concede that in a reasoning similar to the case of the killer, embodied vulnerability will probably outweigh the other more often than not, for the reason given above. Embodied harm can lead to a potentially perpetual incapacity at bringing about a desired action into the world, which potentially outweighs the psychological damage of not having an end or need being met. Still, this isn't to be taken for granted, since some people might value ideals above their limbs, for instance. From this point some might be inclined to suggest that herein lies a sort of utilitarianism of vulnerability, where instead of maximizing utility or happiness, we aim at minimizing exploitations of vulnerability, or autonomy breaches. Recall what I mentioned about O'Neill's problem of *scope* and *justification* in Chapter II. I suggested that her approach was closer to the Kantian standard than Rawls's because her focus was on finding norms couldn't be accepted by all, hence rejecting them, rather than norms that could. This would result in a moral theory focused on the *protection of agency*.

My understanding, and proposal, is that vulnerability in itself provides substantive content to the practical standpoint in the form of a restriction. When agents enter the practical standpoint where they reflect about what is the right thing to do, they do so because they know their actions have consequences, and their actions have consequences only because they impact the finitude of other beings and things, because they are vulnerable. If shooting a gun against someone's head did *nothing*, it wouldn't be *wrong*. It simply wouldn't matter. But it matters because it does *something*, and in this particular

---

<sup>47</sup> I can see how some would draw similarities between this kind of reasoning and a "utilitarianism of vulnerability," but due to shortage of space, I will hopefully address this claim in another work.

case, that something is ending someone's life, the maximal exploitation of vulnerability possible. It ends that person's agency. The notions of vulnerability and finitude breed morality.

But this is not the case only with morality. When we think about the right thing to do for ourselves, it is also in a way a question that exists only because we are finite. When as teenagers we are faced with the choice of a career, we often think "what should I do?/what should I choose?" and while this question is manifold and may have a number of different formulations and motivations, such as "what would make me happier?/what would be the best for my family?/what would improve the world?", all of these stem from the notion that our time is limited. In the first case, we know our time is limited and we will one day die, therefore we try to find the *best, perfect* option to feel like we are not "wasting it." On the second case, we know a steady income is necessary to help our loved ones, because they are *vulnerable*. On the third case, we extend the range to the notion that we will one day die and would like to believe the world is a better place thanks to us. All of these have to do with *finitude*. Finitude and vulnerability are the conditions that haunt us, that press us to make the *right* choices. If we were immortal, invulnerable, and everything around us too, such questions wouldn't matter. We would have all the time in the world to pursue a lawyer career, and then swap to carpenter, or learn to play and make a career in volleyball. Finitude creates right and wrong. Therefore, finitude and vulnerability create in themselves substantive content for the practical standpoint. Their moral status is inescapable as soon as once thinks of *the right thing to do*. And this, in my view, defeats Street's stance that there are only formal constraints within the practical standpoint and no substantive ones. There is the substantive content that it is wrong to exploit finitude and vulnerability. This is normative.

In my view, I don't think this is the same as withdrawing normativity or bindingness from the otherwise "objective" norms formulated around the protection of agency. Understanding and accepting that there are vulnerabilities that when exploited are more harmful and hinder agency more than others, - that autonomy (understood as not having vulnerabilities exploited) should be maximized, or rather that the exploitation of vulnerabilities should be minimized – is not saying that some vulnerabilities are okay to exploit. Like O'Neill said, "*the defeasibility of normative principles is a strength, not a weakness, because it provides the basis for an account of practical judgement. (...) Those who see principles as defeasible still take them to be important, and insist that even where*

*justified principles cannot be enacted, they must be taken seriously.*” (O’Neill 2018, 188)

The example of the killer in front of your house, or the trolley problem, are postulated as thought experiments meant to challenge one’s moral stance. It looks as though there is no right answer, no acceptable outcome that *feels* genuinely good. But these thought experiments are precisely the kind that challenge comprehensive frameworks for moral thinking. It is conceivable that such a scenario could play out in real life, and this is what one would call being dealt “a really bad hand” where you just can’t “win.” But you still have to *act* (or not, but opting to stand still is an action in itself). So, if we endorse the kind of moral reasoning I have proposed, what would play out?

We are agents who accept that there is a moral duty *not* to harm, exploit vulnerability or otherwise endanger the already fragile finitude of another being. Doing so would be necessarily wrong according to the framework presented. We take this to be a *normative principle*. It does not aim at being a comprehensive moral theory that tells you what the *right* action is. Instead, it tells you what is ultimately wrong, *in principle*. It doesn’t tell us the *right* thing to do because we aren’t all-knowing. It tells us what the absolutely *wrong* thing in principle is, so we can avoid it. And watching someone die will always be wrong, because it will mean we failed to protect their autonomy, their finitude, their vulnerability, when *that* would have been the right thing to do, as that is the foundation of morality. An agent who enters the practical standpoint with restrictive moral constructivism of my kind in mind will *know* that *in principle* there is an *objective duty to avoid any harm or foul to a person’s vulnerability*. This is relevant for the agent. But it is also fallible in specific scenarios such as the one selected. There is no outcome which avoids this objective substantive principle, and thus the agent is thrown into a position where it will depend on each particular agent’s set of subjective beliefs and inclinations to make a choice who will never *be* or *feel* entirely right or wrong, because *we too are finite and vulnerable, and not all-powerful*. Were we infallible, all-powerful, and morally inclined, we would simply stop the train and save everyone. *That* would be the right thing to do, the morally correct, the one action who could preserve every subject involved in the scenario. But we are not. And accepting this limitation means to accept that there is an objective principle of not harming other subjects inherent to moral questions, which we should *always* strive to abide for, but that there will be morally gray situations where we are thrown into an impossible choice. Moral character is evaluated whenever we do *everything in our power* to prevent harm. When someone harms another simply for

personal enjoyment or desire, we can be sure that it is *wrong*. When we want something but to get it we would have to exploit another subject's vulnerability, and we refrain from doing it, we know it is *right*. And when we are incapable to avoid this because it is not within our power to stop it, the principle still throws us aback, it still *matters* to us, but we cannot conform to it, and therefore *right* and *wrong* are ideas which haunt us without escape, even though the *right* option is unavailable to us. Alternatively, one could argue that an objective duty not to cause harm would translate into an objective duty to reduce harm. In this sense, then, the best course of action would be to pull the lever, reducing the number of casualties to one instead of five. This is not to say that it would be the *right* decision, but the lesser wrong one, as it causes *less harm*. Nonetheless, it still pertains to a morally gray area where the moral character of the agent and the action are trialed by the simple concern with the objective duty not to harm and either option couldn't be considered entirely *wrong*. If the agent's motivation is still in line with the objective duty not to harm, but out of his/her limitations they cannot comply with it and opt for one or other (perceived lesser wrong) action, it does not mean they are of poor moral character or did the wrong thing. There will always be casualties, and in both cases the agent would face accusations of doing the wrong thing, simply because there are no clear moral criteria that apply here outside of "*avoid causing harm*," where harm was unavoidable.

Finally, how does one vindicate morality toward a skeptic? Simply put, one does not, for it is *not* binding. To claim something is binding is to claim there is no other option but to abide by it, and constructivism, as far as it is anti-realist, rests upon the considerations of each agent within the practical standpoint. The truth, it seems to me, is that no single moral theory can ever hope to challenge an agent's skepticism, and that is a flaw not of moral theories but of skepticism itself. Shouting at a wall won't move it. A person with eyes closed cannot see a movie, and someone who covers their ears cannot hear a song. In the same way, skepticists close themselves to the realm of morality, hence it holds no sway over them. They just do not value moral reasoning because they do not believe in its bindingness, which is dependent on the act of valuing. And choosing whether to value something or not is an act of free-will, even if Kantians claim that one who does not act morally is not truly autonomous. Morality, as far as it is considered, may indeed hold objective moral norms, but it can only do so within its own realm which pertains to those agents who are truly engaged in moral reasoning and reflectivity. In short, morality is not *absolute* for all beings, but there are *absolute* moral principles who holds for those that

care about morality. In the end, the moral status of others is permanent, but your moral agency becomes an act of judgment, which defines you. This concludes my research on vindicating a form of anti-realist objective constructivism.

## **Conclusion**

Unsurprisingly, it is no easy task to try and vindicate moral objectivity, particularly within a constructivist understanding of moral reasoning. The lesson drawn from Rawls is that one must refrain from the idealization of the agents relevant for the construction of norms, under penalty of incurring into relativist (understood in the minimally objective sense), objective only for a restrictive group of agents who share specific characteristics and a common set of considered judgments and values. O'Neill stresses the importance of scope for ethics, and under the form of the modal claim makes clear that norms should be followable by others, meaning that they should in principle be able to be endorsed, establishing a restrictive requirement to dismiss norms which *cannot* be followed by all, in Kantian fashion. Further, the application of norms must be oriented by our judgments about who is implicated and matters for the action. However, O'Neill's theory is accused of being goal-oriented because it assumes morality serves the purpose of coordination intended by all agents and is the best practical option to achieve that end. Conversely, Korsgaard tries to justify the bindingness of moral norms in the ultimate value of humanity, deeming moral action the only kind of unifying action for agency. The categorical imperative is taken to be the supreme principle of reason itself, meaning if one does not act morally, one is not unified as an agent in past, present and future, because one cannot give reasons who are universally consistent. Finally, Street's criticisms place this supposed value of humanity under trial as something which is also presupposed by Korsgaard and not valued intrinsically, as shown through the example of the social bug who does not value itself other than instrumentally for the purposes of serving the Queen. Through this example, the argument that a valuing agent must value itself for its values to matter falls flat and fails. Street then claims that Humean Constructivism is the true anti-realist constructivist project, and that one must not seek intrinsic value in anything and rather accept that norms are constructed through rational standards of correctness using each agent's substantive set of values. Still, in a later work, she acknowledges that universal norms should be ideally sought out, by identifying a universal problem for which only morality can be the answer but denies the feasibility of such a task (even her own proposal of the problem of attachment and loss). I address her claim that one cannot

come up with substantive reasons through “empty formalism” resorting to de Maagt’s proposal that categorical instrumental reasoning can indeed ground a substantial objective universal claim to life and autonomy positing that this claim, accompanied by the inescapable feature of finitude and vulnerability for all living beings, can in fact ground a morally objective principle of non-exploitation of these features.

Resorting to Bagnoli’s work, I propose that the universal problem sought by Street can be the problem of finitude and vulnerability. Understanding both the inescapability of authority for one’s actions and therefore responsibility for their consequences for others, as well as the universality of the feature of finitude calls practical reason to question itself, to make agents question whether their actions are correct for others, seeking justification. The problem of finitude is universal in kind for all living beings, and at the same time, every being needs, to pursue any kind of end, autonomy to do so, and therefore has a substantive claim to be safeguarded from the problem of vulnerability. I argue that this universal problem and this universal substantive claim grant every being the status of moral subject, while the recognition of our authorship and the consequences of our actions and the capacity for practical reasoning grant us the status of moral agents. If every being were infinite and invulnerable, our actions would be inconsequential for others and therefore we would only be concerned with purely instrumental reasoning to achieve our ends. Conversely, because every being is finite and vulnerable, and our actions are always consequential for others, we are called to question whether they are correct or wrong in the moral sense; this is the seed of moral reasoning. In short, while purely selfish instrumental reasoning cares not for the finitude or vulnerability of others (only whether it is effective and successful in satisfying our ends), moral reasoning does, because vulnerability is the condition through which we question whether our action is right or wrong for others.

Because the problem of vulnerability is universal, and having vulnerability exploited damages autonomy, we must realize that exploiting vulnerabilities is invariably wrong for the one that is being exploited or damaged (in the broad sense of the word), hence *wronged*. Thus, when we wrong someone by damaging their finitude through the exploitation of their vulnerability, we are dismissing their substantive claim to autonomy. Furthermore, in agreement with Korsgaard and Bagnoli, I endorse that as rational agents we cannot escape normative parity, meaning that when faced with a reason not to do something, we must either refrain from doing it, or do it only if we can justify our action

with other reasons (because that is what it means to act rationally). Consequently, because every being has a substantive reason to want to live autonomously, the only way through which we can act in a way that goes against that claim is by justifying our actions with some other reason. And here is the problem: because we are in the domain of moral reasoning and we have identified the problem of vulnerability as universal and the substantive claim to autonomy also as universal, there is an objective reason *not* to damage another subject's autonomy by exploiting their vulnerability. In conclusion, because we cannot conceive of any other *moral reason* which is also universal and objective as the moral reason to not have vulnerability exploited, we cannot *ever* break this principle *through moral reasoning*. As such, there is an objective, universal standard, and guiding principle for moral action. Moral objectivity is thus vindicated.

We are now left only with the problems of the bindingness of morality and the anti-realist character of the kind of constructivism expressed here. Both these problems are intertwined, I argue, since moral reasoning is binding only insofar as each individual values acting right for others from when it comes to this realization. Because moral reasoning is the kind of practical reasoning which is committed to doing the right thing by others, if one does not value acting right by others, I take it that morality is simply not binding, despite there being an objective criterion for moral action, and skeptics represent such a problem for moral theorists because everyone wants to assert absolute bindingness of moral value. I take skeptics to simply be, in Street's fashion, agents whose substantive set of values dismisses the value in acting right by others. Thus, because constructivism is an anti-realist theory, they are not *bound* to value morality. Purely selfish instrumental reasoners are still rational agents, but they value only themselves, and we cannot force them to value others, simply put, because value is dependent on each agent's valuing. We cannot say such an agent is irrational, only that they are immoral. And the truth is, they can be okay with being labelled immoral simply because morality pertains to our relationship with others, and they may not care for how others label their actions. They may not value others' reasons, and thus act immorally (but they cannot escape having their action labeled as immoral because they cannot coherently justify their actions as moral if they dismiss the other's substantive moral reason to not be wronged). One who recognizes how he harms others, like Caligula, may resort to keep harming them, dismissing their pain, and suffering, but he cannot escape his action's authorship, his responsibility for it and its consequences, and he cannot reasonably justify his actions

other than with his own set of substantive values and subjective reasoning, which will always be selfishly instrumental, and never objectively moral in kind. In this sense, I conclude, objective moral norms are possible for *moral* constructivism, even if this one is “truly” anti-realist about value, thus making moral norms binding only for those who value other’s reasons and moral status. I also take my position to be safe from the goal-oriented criticism pointed to O’Neill. After all, I am not assuming that morality rests on a need for coordination. Rather, it is born out of the acknowledgement of our authorship of actions and their consequences for others, something possible only due to the fact of finitude or ontological vulnerability. This ultimately rests in how much each agent values such an understanding, thus being anti-realist and not goal oriented.

I hope to have shown both the merits and shortcomings of the predominant works of these great recent and contemporary scholars, drawn to the vindication of objectivity through the maximal intersubjective agreement possible: universal agreement. Considering my analysis of their projects, I sought to identify a universally common denominator, objective due to being a mind-independent fact which necessarily constrains our reasoning and capacity to act. If both my argument regarding the problem of finitude and ontological vulnerability as a universal problem, as well as the categorical substantive claim to not have the capacity to act hindered hold, I believe to have found it. In being both a mind-independent fact which constrains our capacity to reason and act, as well as universal in kind, I argue it holds objective moral status. If nothing else, because it is a universal problem and it sustains a universal substantive claim, we can argue that it is present in every agent’s substantive set of values within the practical standpoint, even from the viewpoint of a Humean constructivist. As such, it seems to hold weight for both Kantian and Humean Constructivism alike. It shapes and impacts the norms which will be constructed, providing an objective standard for moral action which nonetheless can be accommodated to the contingent ways in which vulnerability manifests for each agent. When constructing moral norms dialectically, subjective substantive reasons can hold on their own insofar as they don’t go against the universal (and thus objective) substantive claim to not have vulnerability exploited (unless this claim is surrendered by the subjects themselves). The criterion takes the form of a principle, but the norms constructed are flexible and pertain to the contingent manifestations of vulnerability of each agent.

Finally, if my argument about how practical reason grounds moral agency and finitude or ontological vulnerability ground moral subjectness holds, I believe this could be a promising attempt at trying to vindicate a non-anthropocentric account of morality.

## Bibliography

Anderson, Joel (2013). Autonomy and vulnerability entwined. In Catriona Mackenzie, Wendy Rogers & Susan Dodds (eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy*. Oup Usa. pp. 134. doi.org/10.1093/acprof:oso/9780199316649.003.0006

Bagnoli, Carla (2013). Constructivism about Practical Knowledge. In *Constructivism in Ethics*. Cambridge University Press. pp. 153-182. doi:10.1017/CBO9781139094221.009

Bagnoli, Carla (2014). Starting Points: Kantian Constructivism Reassessed. *Ratio Juris* 27 (3), pp. 311-329. <https://doi.org/10.1111/raju.12047>

Bagnoli, Carla (2016). Vulnerability and the Incompleteness of Practical Reason. In Christine Strahele (ed.), *Vulnerability, Autonomy and Applied Ethics*. London: Routledge. pp. 13-32.

Bagnoli, Carla (2017). Kant in Metaethics: The Paradox of Moral Autonomy, Solved by Publicity. In M. Altman (ed.), *Kant Handbook*. Palgrave. pp. 355-377.

Bagnoli, Carla (2020). Practical Knowledge, Equal Standing, and Proper Reliance on Others. *Theoria* 86 (6):821-842. <https://doi.org/10.1111/theo.12246>

Barry, Melissa (2013). Constructivist Practical Reasoning and Objectivity. In David Archard, Monique Deveaux, Neil Manson & Daniel Weinstock (eds.), *Reading Onora O'Neill*. Routledge. pp. 17-36

Besch, Thomas (2009): Kantian Constructivism, the Issue of Scope, and Perfectionism: O'Neill on Ethical Standing, in *European Journal of Philosophy* 19 (1), pp. 1-20. <https://doi.org/10.1111/j.1468-0378.2009.00376.x>

Bratu, Christine & Dittmeyer, Moritz (2016). Constitutivism About Practical Principles: Its Claims, Goals, Task and Failure. *Philosophia* 44 (4):1129-1143. <https://doi.org/10.1007/s11406-016-9748-1>

Budde, Kerstin (2009): Constructivism all the way down – Can O'Neill succeed where Rawls failed? in *Contemporary Political Theory* 8 (2), pp. 199-223. <https://doi.org/10.1057/cpt.2008.41>

de Maagt, Sem (2019). Why Humean constructivists should become Kantian constructivists. *Philosophical Explorations* 22 (3), pp. 280-293. <https://doi.org/10.1080/13869795.2019.1617341>.

Engstrom, Stephen (2013): Constructivism and Practical Knowledge, in *Constructivism in Ethics*, Cambridge, pp. 133-152. doi:10.1017/CBO9781139094221.008

Formosa, Paul (2011): Is Kant a Moral Constructivist or a Moral Realist?, in: *European Journal of Philosophy* 21 (2), pp. 170-196. <https://doi.org/10.1111/j.1468-0378.2010.00438.x>

Formosa, Paul (2014). The role of vulnerability in Kantian ethics. In Catriona Mackenzie, Wendy Rogers & Susan Dodds (eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy*. New York: Oxford University Press. pp. 88-109.

Harsanyi, J. (1975): Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory., in *American Political Science Review*, 69(2), pp. 594-606. doi:10.2307/195909000

Korsgaard, Christine (1996): *The Sources of Normativity*, Onora O'Neill (ed.), Cambridge.

Korsgaard, Christine (2006). Morality and the distinctiveness of human action. In Stephen Macedo & Josiah Ober (eds.), *Primates and Philosophers*. Princeton University Press. <https://doi.org/10.1515/9781400830336-008>

Korsgaard, Christine M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.

Korsgaard, Christine (2010): Reflections on The Evolution of Morality, in: *The Amherst Lecture in Philosophy* 5, pp. 1-29. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:5141952>

Lafont, Cristina (2004): Moral Objectivity and Reasonable Agreement: Can Realism Be Reconciled With Kantian Constructivism?, in *Ratio Juris* Vol. 17 No. 1, pp. 27-51. DOI: 10.1111/j.0952-1917.2004.00253.x

Laitinen, Arto (2016): Hegelian constructivism in ethical theory?, in Italo Testa & Luigi Ruggiu (eds.) *"I that is We and We that is I": Perspectives on Contemporary Hegel: Social Ontology, Recognition, Naturalism, and the Critique of Kantian Constructivism*, pp. 127-146.

O'Neill, Onora (1989): *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge.

O'Neill, O. (1996). *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511621239

O'Neill, Onora (2003): Constructivism vs. Contractualism, in: *Ratio* 16, pp. 319-331. <https://doi.org/10.1046/j.1467-9329.2003.00226.x>

O'Neill, Onora (2010): 2. The Method of A Theory of Justice: (§§ 7, 9, 23–26, 63–64, 87), in: Höffe, O. ed. *John Rawls – Eine Theorie der Gerechtigkeit*. Berlin: Akademie Verlag, pp. 27-43. <https://doi.org/10.1524/9783050050225.27>

O'Neill, Onora (2015): Changing Constructions, in: *Constructing Authorities: Reason, Politics and Interpretation in Kant's Philosophy*, Cambridge, pp. 57-71.

O'Neill, Onora (2018). *From Principles to Practice: Normativity and Judgement in Ethics and Politics*. Cambridge University Press. <https://doi.org/10.1017/9781316286708>

Rawls, John (1971). *A Theory of Justice: Original Edition*. Belknap Press.

Rawls, John (1980/1999): Kantian Constructivism in Moral Theory, in: Samuel Freedman (ed.): *Collected Papers*, Cambridge, MA, pp. 303-358.

Rawls, John (1993). *Political Liberalism*. Columbia University Press.

Rees, Dafydd Huw (2020). Habermasian Constructivism: An Alternative to the Constitutivist Argument. *Kantian Review* 25 (4), pp. 675-698. doi:10.1017/S1369415420000382

Ronzoni, Miriam (2010): Constructivism and Practical Reason: On Intersubjectivity, Abstraction, and Judgment, in *Journal of Moral Philosophy* 7 (1), pp. 74-104. <https://doi.org/10.1163/174046809X12544019606102>

Savoia, Paolo (2015): Objectivity: Philosophical Aspects, in *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition), pp. 96-102. <https://dx.doi.org/10.1016/B978-0-08-097086-8.63054-4>

Street, Sharon (2010): What is Constructivism in Ethics and Metaethics?, in: *Philosophy Compass* 5 (5), pp. 363-384. <https://doi.org/10.1111/j.1747-9991.2009.00280.x>

Street, Sharon (2012): Coming to terms with contingency: Humean constructivism about practical reason. In Jimmy Lenman & Yonatan Shemmer (eds.), *Constructivism in Practical Philosophy*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199609833.003.0003>

Street, Sharon (2016). I - Constructivism in Ethics and the Problem of Attachment and Loss, in: *Aristotelian Society Supplementary Volume* 90 (1), pp. 161-189. <https://doi.org/10.1093/arisup/akw002>

Watkins, Eric & Fitzpatrick, William (2002): O'Neill and Korsgaard on the Construction of Normativity, in *Journal of Value Inquiry* 36 (2-3), pp. 349-367. <https://doi.org/10.1023/A:1016164822219>