How to ‘Orient’ a Theory of Justice: Rawls and the Ideal/Non-Ideal Distinction

By

Keith Searing

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts

Dalhousie University
Halifax, Nova Scotia
December 5, 2019

To my father,
Jim Searing.

TABLE OF CONTENTS

LIST OF FIGURES:

ABSTRACT:

This thesis deals with two broad questions: (1) “What makes a theory of justice ideal as opposed to non-ideal?”, and (2) “what are the consequences of taking one or the other approach?”. In the first part of the thesis, I discuss ways of drawing a distinction between ideal and non-ideal theories of justice, drawing from contemporary debates in political philosophy. I then propose that we ought to view the distinction as being about the overall purposes of a theory of justice. Theories that try to give an account of justice, I argue, are ideal theories and those that give an account of injustice are non-ideal. This leads to the conclusion that justice is a negative concept, indicating the lack of injustice in the world. Where theories of justice have traditionally gone wrong in trying to give an account of justice from the wrong direction, by beginning with an account of a perfect society, I argue that non-ideal theories have a better chance of responding to real injustices. Finally, I show that there are negative consequences from taking the ideal theory approach. In particular, I show that the ideal method- as exemplified by John Rawls- can exclude discussion of important injustices because of its methodological starting point.

LIST OF ABBREVIATIONS USED

DP	Difference Principle
EL	Principle of Equal Liberties
FEO	Principle of Fair Equality of Opportunity
TJ	*A Theory of Justice*

ACKOWLEDGEMENTS

For all of the helpful comments and careful support I received throughout the writing process, I would like to thank my supervisor Greg Scherkoske. I would also like to thank my committee members, Chike Jeffers and Stephanie Kapusta, who were invaluable resources and influences during my studies at Dalhousie. Finally, a special thanks goes out to my mother, Patricia Hickling, and to my brother and sister, Allison McKim and Alex Searing.

CHAPTER 1: INTRODUCTION

The distinction between ideal and non-ideal theory, in its present form, was first introduced by John Rawls in *A Theory of Justice.* For Rawls, ideal theory is the first part of a complete theory of justice. Its primary function is to articulate principles of justice for a well-ordered society. Call them 'principles of perfect justice'. The principles specify a particular distribution of social goods and positions that would be chosen as the best among possible alternatives by rational agents in an 'original position' of initial equality.

The original position is a thought experiment that models an impartial deliberative process in which citizens weigh and decide on principles of justice for a (soon-to-be) well-ordered society. Citizens choose from behind a 'veil of ignorance', which means that they don't know certain fact about themselves. In particular, citizens do not know whether they would be a member of the least well-off group in the society that emerges. The veil of ignorance restricts the possible types of reasons that citizens might offer in such a deliberative process. Rawls argues that under these conditions, citizens would choose principles of justice that promote a fair distribution of goods and positions.

In short, the original position works as a methodological justification for the principles of justice in Rawls' ideal theory. A principle of justice is reasonable insofar as it would be chosen in a situation of initial equality similar to the original position. Rawls states that in the original position citizens would choose the two principles of "justice as fairness", along with a priority rule, which governs trade-offs between the principles. The first principle, or the principle of equal liberties (EL), states:

(a) citizens ought to share in the most extensive set of basic liberties possible[1]

[1] In *Justice as Fairness: A Restatement,* Rawls lists four categories of basic liberties, which include claims rights, freedoms, and entitlements: "freedom of thought and liberty of conscience; political liberties (for

(b) each citizen's set of basic liberties ought to be compatible with the greatest possible scheme of liberties for all citizens

The second principle of justice has two parts: a principle of fair equality of opportunity and the difference principle, which is a principle of distributive justice. The second principle states that social and economic inequalities ought to be arranged so that they are both:

(a) to the greatest benefit of the least advantaged, as determined by their share of primary social goods (the difference principle or DP);

(b) attached to offices and positions open to all under conditions of fair equality of opportunity (FEO, or the Fair Equality of Opportunity principle)

(Rawls, 1999, 42-46; 1971, 95-97).

The priority rule states that equal citizenship takes priority over other positions, such as those created from distributions in income and wealth. So, there can be no 'trade-offs' between having a greater share of goods in exchange for having a lesser share of basic rights as a citizen. Rawls writes: "each person holds two relevant positions: that of equal citizenship and that defined by his place in the distribution of income and wealth" (Rawls, 1971, 96). The priority rule requires that the position of equal citizenship cannot be undermined for the sake of a higher position in income and wealth. In other words, the first principles of justice (EL) takes priority over the second (FEO and DP). Furthermore, fair equality of opportunity must be in place first before the difference principles can be applied. That would mean that inequalities of opportunity resulting from the "natural

example, the right to vote and to participate in politics) and freedom of association, as well as the rights and liberties specified by the liberty and integrity of the person; and finally, the rights and liberties covered by the rule of law" (Rawls, 1999, 44).

lottery" at birth (such as one's neighborhood, family status, access to public goods, natural talents, and so on) should be corrected for first. This would be done by policies that provide "careers open to talents", as well as policies that redistribute wealth generated by economic exchanges and "natural assets" (Rawls, 1971, 73-74). Hence, the principles are lexically ordered according to the importance of the type of inequality addressed by each.

One crucial question for the discussion ahead is: why did Rawls only take there to be two relevant social positions? One possible answer is that Rawls' conception of the human subject, which views each person as having two basic "moral powers"- reasonableness and rationality- led him to believe that any ideal society must respect these fundamental moral powers.

Reasonableness corresponds to a citizen's 'sense of justice', which motivates them to act in accordance with principles of right and wrong. It is our capacity to act on principles that allows us to engage in social cooperation. If we could not align our different action-guiding principles towards a single goal, such as cooperation, society would not be possible. Rationality, on the other hand, means that each citizen will have their own conception of the good, which implies preferring a greater share of wealth, power, and prestige, since these are means to achieving one's own conception of a good life. These two 'moral powers', Rawls argues, would guide the deliberators in the original position and would form the "basis of equality among citizens as persons" (Rawls, 1971, § 77; 1999, 20). Citizens who have these moral powers are thought to have the requisite capabilities to engage in social cooperation and to view themselves as being 'free and equal' in relation to each other. Thus, the conception of the human subjects as a free and equal- having two basic “moral powers” of reasonableness and rationality- is sufficient, according to Rawls, for the purposes of ideal theory, which is to specify a particular social ideal.

The overall purpose of ideal theory, according to Rawls, is to “set up an aim to guide the course of social reform” (Rawls, 1971, 245). This means that ideal theory ought to be able to orient our own search for a better world. In particular, it ought to give us a sense of where we are going in our pursuit of justice.

1.1: STRICT COMPLIANCE AND THE IDEA OF A WELL-ORDERED SOCIETY

The main reason for calling ideal theory *ideal,* according to Rawls, is that it involves two key (idealized) assumptions: “strict compliance” and the idea of a “well-ordered society”. The assumption of strict compliance stipulates that people will act in accordance with the principles of justice once they are agreed upon in the ‘original position’ (Rawls, 1971, 454). The idea of a “well-ordered society” stipulates that all citizens, viewed as free and equal persons, recognize and share a particular conception of justice, whether it be a libertarian ‘natural right’ doctrine, a utilitarian ‘maximin’ rule, or ‘justice as fairness’, which they use to settle disputes (Rawls, 1971, 4-5; 453-462).

Since no actual human societies have achieved this ideal of behavior, in Rawls’ own words, these assumptions amount to a “considerable idealization” (Rawls, 1999, 9). Some have responded that his view of an ideal society therefore represents an impossible goal that could never be achieved (Weins, 2017). But ideal theory is supposed to be consistent with what Rawls calls the ‘circumstances of justice’, which include realistic, yet reasonably favorable, conditions of social cooperation. So, a well-ordered society that exhibits justice as fairness as its particular conception of justice would be possible because it would share the same basic conditions as some actual societies.

The circumstances of justice, which Rawls draws from Hume's account, are the basic necessary conditions of social life that make cooperation possible. They include co-existence, similar physical and mental powers, moderate scarcity, competing "plans of life", and differing conceptions of the good (Rawls, 1971, 127). Moderate scarcity leads people to into schemes of cooperation because social cooperation makes it easier to gather and share resources. Persons' individual plans of life, however, lead to conflicts in their claims to various social and natural resources. Finally, a reasonable amount of pluralism among people's comprehensive moral and religious doctrines leads to the impossibility that each person in society will be able to agree on the same comprehensive doctrine or conception of the good. The circumstances of justice, then, are thought to reflect the conditions of developed societies that exhibit reasonable pluralism. In his later work, Rawls is more explicit in his intent to offer a theory of justice for a liberal democratic society: "Justice as fairness is a political conception of justice for the special case of the basic structure of a modern democratic society" (Rawls, 1999, 14). This change seems to suggest that Rawls thought it would not be possible to achieve justice as fairness in societies with less favorable historical conditions or with rigid moral doctrines.

1.2: THE MOVE TO NON-IDEAL THEORY

If ideal theory assumes strict-compliance and a well-ordered society, then one kind of *non*-ideal theory would involve adjusting the assumption of strict compliance to admit various degrees of non-compliance or partial compliance. Rawls' own brief forays into non-ideal theory of this kind include just war theory and civil disobedience (Rawls, 1971, 248, 365). But it would also make sense to adjust the assumption of a well-ordered society to admit a variety of competing conceptions of justice. Rawls' dealings with non-liberal societies in *The Law of Peoples* might plausibly count as non-ideal theory of the second type (Rawls, *Law of Peoples*, § 9-10). The first type tells us what to do when others are not complying with the rules or when we shouldn't comply for reasons of justice. The second type tells us what to do when there aren't any fair rules to comply with at all because no one can agree on basic principles. Examples might be found in the interaction between states in the global context, or in the domestic context, when the rules are rigged in favor of some specifiable group, as in the case of institutional oppression.

Rawls did not think it was possible to apply the principles of justice arrived at in ideal theory directly to non-ideal circumstances: "The principles and their lexical order were not acknowledged with these situations in mind and so it is possible that they no longer hold" (Rawls, 1971, 245). He proposes, however, a certain priority relationship between the two types of theories:

> "The intuitive idea is to split the theory of justice into two parts…Nonideal theory, the second part, is worked out after an ideal conception of justice has been chosen; only then do the parties ask which principles to adopt under less happy conditions…The lexical ranking of the principles [of ideal theory] specifies which elements of the ideal are relatively more urgent, and the priority rules this ordering suggests are to applied to nonideal cases as well. Thus, as far as circumstances permit, we have a natural duty to remove any injustices, beginning with the most grievous as identified by the extent of the deviations from perfect justice" (Rawls, 1971, 245-246).

Here Rawls claims that his principles could provide a roadmap, so to speak, for developing a compatible non-ideal theory that tries to understand how to get from normal circumstances (i.e. the status quo) closer to the ideal of perfect justice. Such a theory would assume partial compliance or unfavorable circumstances in order to arrive at principles that guide action in non-ideal cases. But this would not be possible without a prior conception of ideal justice because without it there would be no way to identify deviations in the first place. Rawls writes: "The reason for beginning with ideal theory is that it provides…the only basis for the systematic grasp of these more pressing problems [of non-ideal theory]" (Rawls, 1971, 9).

The relationship between ideal and non-ideal theory in Rawls' view, then, is primarily one of logical priority. The work of ideal theory comes first and is necessary for the non-ideal project to take place. But the priority relationship is also ethical in the sense that, just as there can be no 'trade-offs' between equal citizenship and higher or lower social positions, there can be no 'trade-offs' between the fundamental conception of justice as fairness in ideal theory and the 'transitional' normative principles that guide action in non-ideal circumstances. In other words, a fully developed non-ideal theory, whatever it may actually look like in Rawls' vision of a complete theory of justice, cannot propose a radical curtailment of justice with the goal of achieving a perfect society in mind.

1.2.1: DIVIDING THE EPISTEMIC LABOUR: SOME CHALLENGES

An interesting question, however, is whether the loose framework Rawls presents for non-ideal theory really works in practice. Writing on civil disobedience, a partial compliance issue, and hence, part of non-ideal theory, Rawls states that it is a public, nonviolent, political act contrary to law (Rawls, 1971, 364). Being public and political means that it cannot be justified on the basis of "group or self-interest" or by "personal morality" or "religious doctrine" (Rawls, 1971, 365). Hence, civil disobedience must be justified (if it is to be justified at all) on the basis of "the commonly shared conception of justice that underlies the political order" (Rawls, 1971, 365). But if there is neither strict compliance to rules nor a shared conception of justice within society, how can non-ideal theory, as conceived by Rawls, tell us how civil disobedience is justified on political grounds?

The problem here seems to be that Rawls' non-ideal framework can only accommodate degrees of partial compliance or degrees of well to non-well-orderedness but not both at once. That seems to be because of his conception of society as "a fair system of cooperation from one generation to the next" (Rawls, 1999, 5). Once we begin to ask a truly non-ideal question like "what if society is neither fair nor fully cooperative?", it appears that we are no longer dealing with 'societies' at all, in Rawls' view, but rather with some form of coercion or oppression which cannot be properly labeled social.

If we allow degrees of non-compliance to the rules, which are understood as principles of justice and not merely as the laws of a society, then unless there is some underlying agreement as to what the principles of justice that underpin the laws are, my act of civil disobedience will appear just to some and unjust to others. The reason we must incorporate some degree of non-compliance into non-ideal theory is that, in normal circumstances, not everybody is going to comply with the rules, even if they are just. On the other hand, if we also accept that our society is not-well-ordered, then issues of partial or non-compliance disappear altogether. In other words, strict versus partial compliance problems are dependent on there being a 'sense of justice' that is agreed upon, public, and mutually recognized. Otherwise, there are only coercive laws under which people do not always have recourse, by means of appealing to 'justice', to adjudicate their claims.

But there is a broader problem in the picture Rawls presents with the relationship between ideals, facts, and normative principles. Once we consider certain facts about social life and social beings; their psychological traits, the probabilities that they will follow the rules, or the extent to which they disagree fundamentally on certain issues, non-ideal theory seems to be faced with rather intractable problems. First, it is unclear what the relationship between ideal and non-ideal theory actually is in circumstances when people don't comply with the rules and when there is pervasive disagreement over individual conceptions of justice (i.e. circumstances like ours). Remember that ideal theory is meant to guide social reform by setting up a goal to orient our path towards achieving a more just world. But in order to recommend reforms, which is the function of non-ideal theory, we must start with the status quo as given. We must accept, then, that there is no agreed upon sense of justice that motivates people to act in accordance with principles of justice, let alone to act in accordance the ordinary coercive laws. Hence, the path towards the ideal society is underdetermined by the goal itself because there could be many ways to get there, and so the function of ideal theory as a guide for social reform is not very clear at first glance.

Second, non-ideal theory is faced with rather difficult epistemological problems such as the amount and complexity of facts to be considered, their relative weight, and the degree to which they contribute to changes in other parts of the theory. It is necessary to consider empirical facts when doing non-ideal theory because once we get rid of the assumption of strict-compliance, which stipulates that people will be motivated to act in accordance with the principles of justice, whatever they may be, we will then have to ask what, in fact, motivates people. For example, if we consider people's psychological propensity for envy or selfishness, whether they will be motivated to cheat the rules once the rules are agreed upon, their will to dominate or disadvantage others, or the extent to which they value social goods comparatively instead of absolutely, the normative principles of ideal theory will no longer tell us what to do because they overlook these facts. Other types of facts that would be necessary to consider from the point of view of a non-ideal theory of justice might be sociological facts (how likely people are to cooperate), historical facts (whether there is a history of injustice), or geographic facts (whether societies live in relative isolation or not).

1.3: CONCLUSION

In sum, Rawls claims that ideal theory is the first part of a theory of justice because it is necessary to orient our search for a better, more just, society. It serves an evaluative function by drawing out the conditions under which a society could be considered 'just' by the lights of the theory. It serves a normative function, as well, by recommending principles of justice for liberal democratic societies that would be chosen in an ideal deliberative process. Once a conception of the 'perfectly just' society is reached and agreed upon, the non-ideal project begins. The relationship between ideal and non-ideal theory is primarily one of logical priority. Without a conception of a 'perfectly just' society, according to Rawls, our search for a 'more just' society, here and now, would be pointless.

CHAPTER 2: DRAWING THE DISTINCTION

2.1: INTRODUCTION

In recent years, Rawls' distinction between ideal and non-ideal theories has been a topic of heated debate. Some, such as Robeyns (2008), Volacu (2018), and Valentini (2011), agree with Rawls' basic account of the distinction, but try to work out a more robust conception of the relationship between ideal and non-ideal theory. Others, such as Mills, (2005), Sen (2008), and Weins (2017), disagree with supposed priority of ideal theory over non-ideal theory, proposing instead that the two projects are "analytically disjoined".

There seems to be some agreement, however, that the ideal/non-ideal theory distinction, as it appears in Rawls' account, can be drawn by reference to two fundamental questions. Ideal theory asks: "what should we (as a society) do when circumstances are ideal and everyone cooperates?", while non-ideal theory asks: "what should we do in circumstances when others are not likely to cooperate?" (Valentini, 2012, 655; Mills, 2009, 162-163). For instance, Mills (2009) claims that there is an ambiguity in Rawls' conception of an 'ideally just' society as meaning "a society without *any* previous history of injustice" and 'ideally just' as meaning "a society with an unjust history that has now been completely corrected" (Mills, 2009, 162). If ideal theory is concerned with normative principles for an "ideally just" society in the first sense, as Mills argues, then it is unclear whether the principles arrived at can apply to actual societies, in particular those that have a complex history of racial injustice. This could plausibly be construed as a call for a theory- or theories- that accounts for various forms of oppression or non-cooperation, and hence, a focus on non-ideal theory in terms of non-well-ordered societies.

Laura Valentini (2012), on the other hand, states that once we ask the non-ideal question, “what should we do in circumstances when others are not likely to cooperate?”, the answers we reach will differ depending on context (Valentini, 2012, 655-656). For example, the answer to the question “how much of my own wealth should I give away to alleviate poverty?” will vary considerably depending on one’s own socio-economic status, as well as the likelihood that others in a similar situation would give an equal amount. Ideal theory’s supposed failure to generate action-guiding principles for real problems has led some political philosophers to seriously doubt the usefulness or necessity of Rawls’ approach: “From the perspective of these critics, contemporary political philosophy should shift its focus from full compliance to partial compliance” (Valentini, 2012, 655). Valentini, however, thinks that this criticism, which claims that ideal theory is insufficiently action-guiding for non-ideal circumstances, results from a misunderstanding of Rawls’s project and the overall purpose of ideal theory, which is to give us a relative ranking of political ideals and not to tell us exactly what to do in non-ideal circumstances (Valentini, 2011, 306-308).

2.2: AGAINST IDEALIZATION

The debate over the focus on full compliance vs the focus on partial compliance is sometimes taken to be a disagreement about fact-sensitivity versus idealization in the ‘inputs’, or initial assumptions, of a normative theory. Idealization is meant to imply either (1) falsification in the theory’s characterization of societies and citizens or (2) ‘fact-insensitivity’ in its affirmation of normative principles. In some cases, idealization is contrasted with abstraction (O’Neill, 1987; Mills, 2005). A ‘mere abstraction’ is a form of bracketing off certain characteristics of a phenomenon, like social cooperation, in order to arrive at a simplified model. Idealization, however, is about ‘adding’ new, falsified, information to descriptive models (O’Neill, 1987; Schwartzman, 2006). Idealization as fact-insensitivity in normative principles, on the other hand, is a matter of justification. An ideal, in this second sense, is taken to be ‘fact-insensitive’ if it is not ‘grounded’ in facts, as a matter of justification or entailment, but is accepted on grounds independent from facts.

Critics of idealization in the first sense, such as Mills (2005) and O’Neill (1987), call for ‘abstraction without idealization’. Simply put, this means better, less parochial, descriptive models of social phenomena. For instance, Mills distinguishes between descriptive modeling and idealized modeling. A descriptive model involves some level of abstraction, but only to simplify the phenomenon in question: “one will make simplifying assumptions, based on what one takes the most important features of *P* [-a phenomenon-] to be” (Mills, 2005, 167). An idealized model, on the other hand, will be an outstanding example of what the phenomenon “should” be like. Idealized modeling is a kind of “extrapolation, in the limit, of the behavior of *P”* (Mills, 2005, 167). The problem with focusing on idealized models is that they obscure real injustices, such as systemic inequalities along racial, gendered, or other lines. Moreover, idealized modeling can be susceptible to parochialism and ideology because the privileged in society will always more closely resemble the idealized model of a citizen than members of oppressed or marginalized groups (Schwartzman, 2006, 571).

For example, Mills claims that Rawls’ conceptual framing of societies as “cooperative venture[s] for mutual advantage” existing in isolation from each other represents a kind of theoretical “white-washing” of history up until the present. This prevents an understanding of societies as fundamentally non-consensual ventures, which might be advantageous (particularly for members of oppressed groups) when thinking about justice. Mills writes: “the *non-naming* of [colonialism] in current Western political philosophical discourse in a sense names it out of existence, deprives us of the cognitive resources to analyze it, or even, (legitimately) to *talk* about it” (Mills, 2015, 10).

Further, idealized models will involve ‘idealized social ontologies’, which universalize the experiences of the privileged elite at the expense of marginalized groups, thus becoming ideological. Mills argues that if *P* (a phenomenon) is human behavior, then in order to get from *P* behaves in *X* kind of way, to *P* should behave in *Y* kind of way, we will have to refer to capacities that human beings do not actually have in the way described by the descriptive model. Mills lists things such as a person’s “degree of rationality, self-knowledge, [and] ability to make interpersonal cardinal utility comparisons” as examples (Mills, 2005, 168). These capacities, however, are not representative of everyday people. For example, making impartial judgments about overall well-being seems to be complicated by what Rawls calls ‘special psychologies’ like status anxiety[2], spite, envy, the will to dominate, and so on, which seem at least to be present in some people.

The advantage of non-ideal theory, construed as a form of non-parochial descriptive modeling, is that it does not ignore important details about the experiences of marginalized groups, which can contribute to a broader understanding of injustice. The result of having more accurate and less ideologically informed models of injustice, then, is a more pluralistic conception of justice, which does not leave out important factual details. It will

[2] For a discussion of ‘status anxiety’, a phenomena in which people value relative gains over absolute gains, and its relation to comparative evaluations of social goods like income and wealth, see: Robert H. Frank and Cass R. Sunstein, “Cost-Benefit Analysis and Relative Position” in *The University of Chicago Law Review*, Vol. 68, No. 2 (Spring, 2001), pp. 323-374

also result in a more empowering concept of justice, particularly for people who are the victims of injustice, which can be used to generate political action in cases where an ideologically informed one will fall short.

2.3: IN DEFENSE OF IDEALS

Whereas the previous discussion of idealization was concerned with how and for what purpose certain facts are incorporated into a normative theory, there are others who take the predicate "ideal" in ideal theory to mean that its normative principles are fact insensitive. For instance, G. A. Cohen (2008) and David Estlund (2014), consider the demands of 'perfect justice', whatever they may be, to be relatively fundamental to normative theories. This debate is often carried out under the heading of ideal/non-ideal theory with different theorists weighing in on both sides.[3] Those who favor a fact-based approach at the expense of an idealized account are referred to as 'utophobes', while those who favor a focus on ideals like 'prime justice' are considered to be 'factophobes'. Interestingly enough, Rawls, who is thought of as the exemplar of ideal theorizing, comes out more in the middle of this debate than as being in favor of idealization *tout court*.

For instance, Cohen criticizes Rawls for not being utopian enough. He thinks there are normatively fundamental principles of justice in the form "we ought to do *X* if we can", which apply universally, and do not answer to facts about possibilities or likelihoods that they will be carried out. Rawls, on the other hand, only tries to construct a theory of justice for liberal democratic states. One reason is the demands of justice as fairness do not seem

[3] Utophobes, such as Mills (2005), Weins (2017; 2015), and Gaus (2017) are skeptical of not only the feasibility of achieving a utopian or 'perfectly just' scheme but also of our ability to know what a perfectly just scheme might look like and thus make appropriate comparisons between our own world and an envisioned utopia. Factophobes, on the other hand, such as Cohen (2008), Estlund (2014), and Valentini (2011) argue that without a theory of political ideals (or an ideal theory) we would be unable to make appropriately critical judgments about our society.

to hold for societies that do not meet a certain threshold level of material or social goods in order for questions of distributive justice to kick in. Another reason Rawls thought that justice as fairness would have to answer to "the facts" can be found in Section 80 of *Theory* on "The Problem of Envy":

> "The second part [of the theory of justice] asks whether the well-ordered society corresponding to the conception adopted will actually generate feelings of envy and patterns of psychological attitudes that will undermine the arrangements it counts to be just" (Rawls, 1971, 531).

In other words, strict compliance theory rules out the possibility for envy in a well-ordered society. That is because strict compliance means that citizens will be inclined to act in accordance with principles of justice. Thus, Rawls assumes that citizens who live in a well-ordered society will not feel envy about someone else having a greater share in primary goods, so long as the inequality in question is still to the advantage of the least well off (Rawls, 1971, 8; 530). But part of the job of non-ideal theory is to determine, if a well-ordered society with 'justice as fairness' as their shared conception of justice ever came to be, whether it actually would or would not create feelings of envy. If it did, according to Rawls, that would be a reason to change the principles of ideal theory. Thus, the process of reflective equilibrium, in which the principles of perfect justice are reached by means of weighing principles against moral intuitions, is not fixed once a particular conception of justice is reached. Wide reflective equilibrium continues on after the project of non-ideal theory begins, and this can affect the content of ideal theory.

For Cohen, however, principles of justice, understood as fundamental normative principles, do not 'respond' to factual considerations at all: "a principle can reflect or respond to a fact only because it is also a response to a principle that is not a response to a fact" (Cohen, 2003, 20). A principle that is not a 'response to a fact' is, in Cohen's words, a "fundamental normative principle" or, in another locution, it expresses "the normative ultimate" (Cohen, 2008, 253). Cohen's idea of "the normative ultimate" in general is somewhat vague. That is because, for one thing, it seems to imply a regress of first principles. That might not be a problem in itself but because each has less specificity, and hence is less intelligible, as the regress continues the principles become more and more vague and general. Further, the relation of Cohens view to the problems of partial compliance and idealization mentioned earlier is elusive.

Cohen's argument is that if there are normative principles that respond to facts, then there must be principles that do not, which ground the fact-sensitive principles and explain why they are true (Cohen, 2008, 232-234). For example, suppose that someone thinks we ought to keep our promises because "only when promises are kept can promisees successfully pursue their projects" (Cohen, 2008, 234). In this case, someone gives a factual reason for affirming a certain normative principle: "we ought to keep our promises". Cohen argues that if the factual reason is to have any bearing at all on the principle, then it must be because of another principle "we ought to help others pursue their projects", which is not grounded in the facts about whether promises help achieve this end (Cohen, 2008, 234-235). If that is true, then there must be fundamental normative principles that do not 'answer' to any facts at all. These principles are fundamental because all other normative principles are grounded in them by means of justification or entailment. They are "ultimate" principles because they come in the form "we ought to do *X* if we can" and do not change simply because we won't or might not carry them out.

To say that principles of justice do not respond or answer to facts, in Cohens words, is to say that we don't "include matters of fact as grounds for affirming them" (Cohen, 2003, 213). That is to say that justice is a fact-independent value, similar to a Platonic Form, which is affirmed on a-priori grounds (or perhaps some other form of justification). Keeping epistemic matters aside for a moment, I would like to focus on how the concept of fact-insensitivity relates to ideal theory as it appears in Rawls.

Some have argued that Cohen and Rawls are merely engaged in a verbal dispute (Valentini, 2012, 657; see also: Williams, 2008) where Cohen's "rules of regulation", which do answer to facts, are not sufficiently different from Rawls' principles of justice. The relation between fact-insensitive fundamental principles of justice and the "rules of regulation" of a society, in Cohen's view, is much like Rawls' conception of the relation between the moral intuitions and the principles of justice, as conceived in the process of reflective equilibrium (Rawls, 1971, 48-51). If that were true, and Cohen's dispute with Rawls is simply misunderstood, then Cohen might plausibly be construed as being engaged in ideal theorizing with strict compliance somehow worked into his conception of the ideal human subject. But if the dispute is not merely verbal, then Cohen's project seems further removed from the concerns of an ideal theory of justice. That is because there is no obvious sense in which his thesis is meant to guide social reform.

Whether or not fact-insensitive principles are a proper part of normative theories, it does seem to be the case that when there is a change in the initial factual assumptions of a normative theory, there will be a corresponding change in the 'outputs', or action-guiding principles, of the theory. This is illustrated by the earlier example of charity and socio-economic status. Suppose we begin with a basic normative premise "we ought to alleviate poverty if we can”. This might be easy to accept on little grounds other than that it seems self-evidently true. But once we consider the ‘can’ in the statement, and what it implies, we must ask certain factual questions like ‘how much do I have to make in order for this duty to kick in?”, or ‘if I gave away X amount, would I be able to meet my own basic needs?”. It seems to be the case that when a person makes more in annual in income, then their duty to alleviate poverty increases. Further, for the people who do have some kind of duty to alleviate poverty, it is still unclear how much they ought to give once we consider how much others who also have a similar amount of wealth might also give.

If these observations are plausible, then ideal principles can only generate ideal solutions to idealized problems. That is because ideal principles only tell us that we should alleviate poverty if we can and not what to do in specific cases, how much to give, etc. The demands of ‘justice as fairness’, similarly, only tell us what type of distributive scheme might be chosen if it is possible and not exactly how to get there from where we are now. Simply not being appropriately action-guiding in particular circumstances or falling short of being an all-in-one algorithm for behavior, however, can’t be an adequate ground to claim that ideal principles aren’t helpful at all. In other words, the complaint against ideal theory can’t simply be that it is useless because it fails to be action-guiding for non-ideal circumstances. This conclusion only shows that the answers to particular questions will vary depending on what the facts are. It does seem to suggest, however, that so long as our focus *should be* on real world problems, our theories ought to be more closely tailored to empirical realities (Schmidtz, 2015, 773-775).

But there is a risk, pointed out by Estlund (2014) and Valentini (2012), of a non-ideal theory taking on a maximal level of fact-based assumptions and thus turning out to be too apologetic to the status quo. For instance, if we lower our priority for achieving ‘perfect justice’ and instead opt for something like ‘peace’ or ‘stability’, which appears more empirically likely based on what we know about societies, we risk undercutting our own ability to make ethical evaluations about the current political order. This observation leads to another dimension of the ideal/non-ideal theory debate, which is carried out along the axis of feasibility versus desirability (Valentini, 2012; Volacu, 2018; Gilabert, 2012; Estlund, 2014). The key questions in this debate are “does ‘ought to’ always imply ‘can’?” and “if ought does imply can, then what restrictions should this place on normative theories?”.

2.4: ON THE OUGHT/CAN RELATION

In general, there seems to be agreement across various perspectives in moral and political theory that an “ought” always implies a “can”. This claim, interpreted weakly, simply says that normative principles are defeated by impossibilities. For instance, it seems futile to adopt a rule that no one can follow because they are simply incapable of doing so. The futility of the rule, in turn, gives us a reason to reject it. So, it seems that to the extent that we accept this version of “ought implies can”, we must also accept that normative principles can be refuted by certain facts. But, for Cohen, the fact-sensitive nature of some normative principles does not affect the “normative ultimate”, or the fundamental normative principles, any more than considering whether people might simply refuse to do something affects whether they ought to do it. In other words, finding out that it would be impossible to carry out a certain rule because the agents involved are incapable would only be a reason to reject the rule itself and not the principle “we ought to do it if we can”, which remains unaffected by the factual circumstances.

Following Cohen, Estlund (2014) argues that even if principles of justice turn out to be highly unrealistic because it is not likely they will be followed by normal people, that does not mean that they are thereby false or useless (118). For Estlund, "ought" does indeed imply "can", but that does not mean it must also imply "highly likely" or even "reasonably likely". He considers the case of 'professor procrastinate' who "declines a request to referee a manuscript that he is duty bound to referee" (Estlund, 2014, 124) on the grounds that he simply will not make the deadline. On the one hand, moral intuitions might tell us that he is not culpable for declining the request because he knew that his involvement in the project would inconvenience others. But, on the other hand, he had no other reason than that he probably would not be able to make the deadline. This is supposed to illustrate the difference between 'concessive' normative theories, which accept certain facts as given and say what we ought to do in those cases, and 'non-concessive' normative theories, which tell us what to do if those facts did not hold. The non-concessive 'layer' of a normative theory, which might only say 'we ought to fulfil our special obligations", does not stand to be rejected simply because, as in the case of professor procrastinate, it might not always tell us exactly what to do.

Political realists, on the other hand, take the opposite position with respect to the ought/can relation: "From a realist perspective, the achievement of perfect justice may be imaginable, but it is not feasible. It is therefore naïve, and ineffective, to hold existing societies to account on the basis of such demanding moral standards" (Valentini, 2012, 659). Realists, instead, typically accept that ideals play some role in evaluating the status quo in a society, but argue that questions of institutional design, legitimacy, and feasibility, ought to be handled primarily by the social scientists who have accurate knowledge of political systems (Guerrero, 2017, 149-150).

Some, however, like Gilabert (2011, 2012, 2017), take a more modest approach to understanding the demands of feasibility. He defines feasibility broadly as an agent's ability to produce an outcome in a particular circumstance. Typically, feasibility is applied as a binary concept. Either an agent is able to bring about an outcome or they are not. But the concept can also be used to imply relative likelihoods or capabilities. In that case, what we mean is that an agent has the ability to bring about a certain outcome to *n* degree of probability. Gilabert proposes a 'dynamic approach' to the relationship between justice and feasibility, in which there "dynamic duties to reshape feasibility constraints over time" (Gilabert, 2017, 124).

Some feasibility constraints are 'hard' in that they involve binary judgments about an agent's overall ability to achieve an outcome. Others are 'soft' constraints because they refer to particular circumstances where the outcome won't be reached, or which make the outcome less likely. In the latter cases, feasibility is a scalar judgement about an agent's relative capability to bring about an outcome given certain circumstances. A scalar feasibility judgment indicates a continuum or degrees of probability, whereas a binary judgment indicates two distinct possibilities.

To illustrate the case of 'hard' feasibility constraints, it is often argued against the principle of average utility (roughly, that we should try to maximize the overall well-being of the largest number of people, given whatever conception of well-being we choose) that it requires an impossible level of benevolence from everyday people. This criticism has been leveled against utilitarianism in many forms. But the main thrust of the objection is that the principle of average utility is simply not possible for actual human beings to act on. ‘Soft’ feasibility constraints, however, when applied to individuals, would refer to conditions that limit the agency of people to bring about desired ends in certain circumstances, such as poverty, disability, or the lack of opportunities and positions which contribute to a person’s autonomy.

But in a political context, feasibility takes on a collective dimension, which needs to be separated from the agent-focused sense in which I have just been using it. Collective feasibility takes as given the powers and abilities of the agents involved to bring about certain ends. In the collective sense, feasibility refers to relative ‘distances’ between one social scheme and another. For example, imagine three simplified distributive schemes: (a) a highly deregulated competitive economy, (b) an economy with ‘welfare-state’ redistribution, or (c) an economy with universal income and more substantive redistributive measures. A luck egalitarian, who views distributions where some are worse off by no fault of their own as unfair, would choose (c) as the best possible scheme. But if, for whatever reason, it turns out to be impossible to bring about (c), then this would mean a luck egalitarian would likely choose (b) as the (second) best possible distributive scheme. In this example, the ‘distance’ to (c) from our own social world (x) is conceived as ‘too far’ to be feasible to get from (x) to (c) directly. Gilabert proposes, however, that in situations like this one, luck egalitarians would have a dynamic duty to make (c) more likely. In practice, that would mean that they have duty to bring about (b) in the short term, but also to make the world such a place that (c) might one day come about.

Gilabert’s notion of a ‘dynamic duties’ implies a kind of hierarchy of normative principles, with a set of principles that mark what we ought to do when we consider the feasible options and a set of principles that mark what we ought to do regardless of the options that are available. Estlund's distinction, between 'concessive' and 'non-concessive' normative principles, also tracks such a supposed hierarchy of principles. At the top of the hierarchy of normative principles, according to this view, are the "global prime" requirements of justice and morality, which do not answer to facts about feasibility.

Weins (2017), however, takes issue with Estlund's idea of a "global prime" requirement of justice. His argument begins with a particular interpretation of the 'ought/can' relation, which he calls the 'uncontroversial thesis': "a set of directive principles is justified relative to a set of salient possibilities" (Weins, 2017, 154). He then proposes that the best way to make sense of the uncontroversial thesis is to introduce an 'optimization model' of normative theories in which a normative theory is defined by a set of evaluative principles, which serve to rank options, and a set of feasibility constraints, which serve to exclude options. 'Optimization', then, refers to the process of recommending the best possible option that is included by the theory's specifications.

Taking 'professor procrastinate' as an example, Wiens explains that what makes the non-concessive requirement (he ought to accept the request to referee) have the kind of primacy that it does is that it specifies the best option among a (locally) maximal set of possibilities. In order for a 'global prime' requirement of justice to hold the same kind of relation to a set of salient possibilities, however, one must make sense of a "globally maximally encompassing set of possibilities" (Weins, 2017, 165). But the only way to do that, Weins argues, is to imagine a society of angels because these would be the only being capable of carrying out these principles. Weins concludes that normative principles must always be specified relative to a set of feasible possibilities. 'Global prime' requirements, therefore, are not normative because there are no possible worlds, or possible agents, to which they apply. He leaves open, however, the possibility that ‘prime justice’, or some other ideal theory requirement, might serve an evaluative function, which could be included in a properly action-guiding (non-ideal) theory.

For a possibility to be ‘salient’ or ‘feasible’, as we have seen, it must be either:

(i) within an agent’s power to bring about under normal circumstances, or

(ii) within the ‘neighborhood’ of possible worlds (a, b, c…) similar enough to our own, such that there are enough characteristics in possible world (a), (b), or (c), in common with our own world (x) to make comparisons.

A ‘feasibility constraint’, on the other hand, is an assumption about the way the actual word is that limits what can count as a possible world within our own ‘neighborhood’. For example, given what we know about human psychologies, the demands of the principle of average utility, which implies that a perfectly benevolent utility maximizing person could exist, seem to be unfeasible. Another example might be that given humanity’s inevitable temptation towards corruption, trusting in the benevolent motivations of a socialist dictator to responsibly transfer control over the means of production to the proletariat/working classes seems similarly misguided and unfeasible. Thus, it seems that any normative system is incomplete without a set of feasibility constraints that specify what counts as our “neighborhood” of possible worlds, given what we know about our actual world.

2.4.1: OTHER RELEVANT DISTINCTIONS

The preceding discussion brings to light an important distinction that has been left unclarified: the distinction between evaluative principles and normative principles.

Evaluative principles offer a (critical) perspective that reaches beyond the status quo and ranks possible social arrangements based on some conception of the good. Normative principles, on the other hand, are supposed to be action-guiding, which means that they are always tied to feasibility constraints.

Although these two types of ethical principles are distinct, they are not necessarily always mutually exclusive. In other words, it is possible to derive a normative principle from an evaluative principle and *vice-versa.* If I start with the premise: "it is good to be reasonable", then it would follow that since being prejudiced and partial in my judgments is not reasonable, I ought to be fair and impartial in my decisions and my judgments. Thus, it appears that one can derive a normative principle, "I ought to be impartial and fair in my judgments and my decisions" from an evaluative principle, which merely says "it is good to be reasonable". Or, alternatively, I could come to believe that I ought not to be prejudiced independently of any particular conception of the good. Perhaps through a long and arduous experience with ethical trial and error I become virtuous. I might then formulate in my mind the principle "I ought not to be prejudiced". I may even go so far as to attempt to explain how my newly formed ethical principle could hold true in different cases. "It's because prejudice effects my ability to reason to a correct conclusion", I would say. It seems, then, that normative and evaluative principles are deeply connected, but they function as two distinct parts of an ethical perspective.

In Rawlsian language, an evaluative principle of justice would be stated in the form: "a distributive scheme is just if and only if inequalities are to the advantage of the least well off as identified by their indexes of primary goods", whereas a normative principle would be stated in this way: "we ought not to allow inequalities unless they are to the advantage of the least well off…”. Here it appears that the difference between evaluative and normative principles is merely formal. In other words, whether I choose an evaluative over a normative principle might just depend on how I want to articulate it for a particular purpose (to be critical of the status quo, for example). But one important difference between the two types of principles is that a normative principle must be specified in relation to some set of feasible possibilities whereas an evaluative principle might not be. That is because normative principles should tell us how we ought to behave. Since “ought always implies can” (whether this principle is meant in a ‘strong’ sense, as political realists often intend, or whether it is meant in a ‘weak’ sense, as in Gilabert’s or Cohen’s view), action-guiding principles must refer to salient or feasible options. Evaluative principles, however, may have a more ‘open-ended’ character, which allows me to imagine an absolute best option, even though I may not be able to achieve it.

This distinction leads into yet another; between a global best and a local best. Weins’ “optimization model” suggests that the function of a normative theory is to rank options based on ‘inputs’, like evaluative principles and feasibility constraints. The purpose (or output) of a normative theory, then, would be to show what salient options are ‘best’, ‘second best’, and so on. This process results in a comparative ranking of the options that are considered to be within our own ‘neighborhood’ of possible worlds. Feasibility constraints are necessary for normative principles in particular because the “ought implies can” principle requires knowledge of possibilities. In the case of human behavior within political systems, this must involve an examination of empirical data concerning (at least) facts about human capabilities, psychological motivations, as well as sociological and historical information about the parties involved.

When one imagines a set of feasible alternatives that are ranked comparatively, as in (a) is better than (b), which is better than (c), the question remains open, however, as to whether there are other alternatives available. The global optimal is the ‘absolute best’ option among a set with no known limit. Presumably the set of possible worlds becomes much larger as one begins to consider more speculative options that are not within our neighborhood. Thus, there must be a difference between a ‘local’ optimal, which is the best out of a set of known feasible alternatives, and a ‘global’ optimal, which is the best out of all possible alternatives. The concept of a global optimal captures our sense of the ‘very best’ or ‘absolute best’ as opposed to simply the best out of a limited set. Normative principles deal primarily with choosing between known alternatives within a certain ‘neighborhood’ of known feasible paths. Evaluative principles, on the other hand, can help us to understand or imagine the best overall option.

How are these distinctions related to the ideal/non-ideal distinction and the debates surrounding that distinction? Amartya Sen has argued that the identification of a ‘perfectly just’ institutional arrangement is neither necessary nor sufficient for advancing justice in the real world. Put simply: knowledge of the best does not entail knowledge of the better. Under this interpretation, ideal theory is concerned with the specification of the very best institutional arrangement and non-ideal theory is concerned exclusively with pair-wise comparisons between any two institutional schemes. Importantly, these institutional schemes are conceived of as existing within our own ‘neighborhood’ of feasible alternatives. The comparative approach recommends piecemeal reforms intended to advance justice or eliminate manifest injustice instead of proposing a radical shift towards prefect justice.

2.5: THE COMPARATIVE APPROACH

Sen claims that the two approaches to thinking about justice are "analytically disjoined". By saying that the they are analytically disjoined, he means that they are two separate methods of inquiry, which begin from different starting points: "Importance must be attached to the starting point, in particular the selection of some questions to be answered (for example, 'how would justice be advanced?'), rather than others (for example, 'what would be perfectly just institutions?')" (Sen, 2009, 9). This way of thinking about a theory of justice involves a much broader outlook than the previous views discussed above. It includes at least two broad criteria:

(i) *Starting point* (i.e. what types of questions are asked/not asked; ideal theories are thought to begin with answering transcendental questions about what types of institutions are fully just, whereas non-ideal theories begin with the status quo as given and attempt to answer comparative questions concerning institutional changes that might advance justice)[4];

(ii) *Orientation* (i.e. the focus and overall purpose of a theory of justice; ideal theories focus on 'transcendental justice', while non-ideal theories focus on 'comparative justice'. Alternatively, ideal theories are focused on creating fully just institutions, whereas non-ideal theories focus on removing injustice).[5]

[4] For Rawls, an ideal theory of justice begins with this question: "what would a just democratic society be like under reasonably favorable but still possible historical conditions allowed by the laws and tendencies of the social world?" (Rawls, 1999. *Justice as Fairness: A Restatement*. Harvard University Press, Cambridge. p.4). This question is far more limited than the question "what institutions are fully just?". Sen, however, thinks that the aim of Rawls' project is to uncover something like a transcendental ideal of perfect justice.

[5] Another way of thinking about the orientation of a theory of justice is in terms of short-term versus long-term goals, but this may not capture entirely what is meant by 'orientation'. A broader understanding is

At present, I will discuss Sen's characterization of the two approaches to justice, his argument for the comparative approach's uniqueness, its relative advantages over the transcendental approach, and then I will go on to conclude this chapter by summing up the various ways of drawing the ideal/non-ideal distinction that have been discussed so far.

The focus on identifying perfectly just institutions, which Sen calls "transcendental institutionalism", has historical antecedents in the work of Enlightenment thinkers like Hobbes, Locke, Rousseau and Kant. In more recent years, philosophers like John Rawls, Ronald Dworkin, and Robert Nozick have defended various versions of ideal theory, offering widely divergent conceptions of perfectly just institutions. Typically, transcendental institutionalism is concerned with identifying the very best institutional arrangement, as opposed to suggesting particular behavioral norms that would apply to the workings of actual societies. Some within this tradition, like Rawls and Kant however, have rather far-reaching theories, which also include insights into "the norms of right behavior in political and moral contexts" (Sen, 2009, 8). In a footnote to page 8 of *The Idea of Justice*, Sen gives an example of the reach of Rawls' theory: "In suggesting for what he calls a 'reflective equilibrium', Rawls builds into his social analysis the necessity to subject one's values and priorities to critical scrutiny" (Sen, 2009) (also consider Rawls' defense of civil disobedience in *TJ*[6] § 80, p. 530-532). Sen suggests here that Rawls' theory could be applies to evaluate political institutions but also for the purpose of evaluating one's own beliefs and moral values.

necessary to capture the overall purpose of a theory of justice as well as its function, which would include whether it recommends short-term or long-term goals.

[6] *Theory of Justice*

Typically, however, transcendental institutionalism has been concerned with identifying institutions that could be considered fully just. Most thinkers within this tradition have viewed a just society as a consensual agreement, or contract, between free and equal persons. The usefulness and influence of social contract theory in the Early Modern period, as well as in recent years, has been far and widespread.

The focus on comparisons between different "social realizations", on the other hand, where 'realizations' are conceived in terms of individual or collective well-being, the behavior of (actual/not-ideal) institutions, or the removal of "manifest injustice", can also be seen being produced roughly concurrently in the work of Adam Smith, the Marquis de Condorcet, Jeremy Bentham, Mary Wollstonecraft, Karl Marx, and John Stuart Mill (Sen, 2009, 5-8)[7]. This tradition, even though its thinkers differ widely, has been primarily concerned with comparisons between "societies that already existed or could feasibly emerge" rather than "transcendental searches" for a perfectly just society (Sen, 2009, 7).

For Sen, a picture of the very best society is neither necessary nor sufficient for making comparisons between feasible alternatives or for suggesting justice-enhancing policies. In this view, transcendental institutionalism is not logically prior to the comparative approach because, by parity of argument: "the fact that a person regards the Mona Lisa as the best picture in the world, does not reveal how she would rank a Gauguin against a Van Gogh" (Sen, 2006, 221). Hence: "it is not at all obvious why in making the judgment that some social arrangement 'x' is better than an alternative arrangement 'y', we have to invoke the identification that some quite different alternative 'z' is the 'best' or the 'right' social arrangement" (Sen, 2006, 222). For example, Sen states that when

[7] Sen's own work can also be included here (see: Amartya Sen, 2009. *The Idea of Justice.* Cambridge Massachusetts: Harvard University Press.).

comparing the relative height of two mountains, we do not need to invoke the tallest mountain on earth, Everest, to make our case. Now, Sen's focus on metaphors might seem altogether too simplistic at this point. So, it will be necessary to see what further support he provides for the major premise in his argument, which states that identifying the very best institutional scheme is neither necessary nor sufficient for the purposes of the comparative approach, which is to compare any two social schemes. If that is true, then one is not logically prior to the other or, in other words, they are "analytically disjoined".

One way that a transcendental best might be sufficient for doing comparative work would be through comparisons between relative "distances" from the transcendental picture of a perfectly just society. This seems to be the view that Rawls endorses when he states: "as far as circumstances permit, we have a natural duty to remove any injustices, beginning with the most grievous as identified by the extent of the deviations from perfect justice" (Rawls, 1971, 245-246). But Sen argues that the "characterization of spotless justice does not entail any delineation whatever of how diverse departures from spotlessness can be compared and ranked" (Sen, 2006, 220). In other words, when we are able to identify the 'best' option among a set of possibilities, this identification itself does not entail any specific ranking of all the other options in terms of 'second-best' or 'third-best'. If, on the other hand, transcendental justice is not taken to specify the "best" possible political arrangement but rather the absolute "right" political arrangement, in which every other possibility is either "not right" or "wrong" as opposed to "second best" or "third best", then it would not be the business of a transcendental theory to result in comparisons at all. In that case, Sen argues, all a transcendental theory of justice can do is to recommend a revolutionary shift to the "right" way of doings things.

Whichever way we specify the demands of justice, the transcendental approach is not going to be sufficient for making concrete comparisons between better and worse existing institutions, according to Sen's argument. But is a transcendental theory of justice somehow necessary to have in the background before one can go about making comparisons? There are two senses of 'necessary' that could apply here.

The first sense of "necessary" is that without knowledge of transcendental ideals, comparisons between better and worse societies would be meaningless or even impossible. That is to say that transcendental ideals are needed for the whole theory of justice to work. The assumption here is that there must be some coherence or unity to a theory of justice or else it might risk being reduced to a form of intuitionism. Moreover, there must be some goal by which to orient the second half of a theory of justice, as in Rawls' view, or else the policies recommended might not put us on a path towards a just society.

Sen thinks, however, that incompleteness in our judgments about social justice is not a fault of the comparative approach but an indication that the transcendental approach is too strict. It is often the case that we just can't decide one way or another on questions of social justice. Sen argues that even if our personal preferences and partiality were lifted away by a "veil of ignorance", we may still retain certain differentiating qualities that make us disagree about the priorities of social justice. One example of an intractable disagreement concerning social justice is "weighing the claims of need over entitlement to the fruits of one's labour" (Sen, 2006, 224). Sen imagines a scenario where we are supposed to choose which of three children is to get a flute about which they are quarrelling:

> "Child A is the only one of the three who knows how to play the flute (the others do not deny this); child B is the only one without any toys of his own (the other two concede that they are much richer and well supplied with

> engaging amenities); child C has worked hard to make the flute all on his own (the others confirm this)" (Sen, 2006, 225).

A utilitarian, an egalitarian, and a libertarian, Sen argues, will all have different, yet equally compelling, arguments to give for why one child should get the flute and not another. I view this scenario as similar to what would happen when representatives in the original position are deciding on different principles of justice that are concerned with just distributions (i.e. the second principle of justice). Sen thinks that a transcendental approach such as Rawls' must assume that a mutual agreement will arise from the original position on such matters. But the reality is far less clear. A comparative approach, however, need not be worried about the lack of a definitive answer to questions of social justice or the lack of a complete ranking of societal arrangements.

Do comparisons get made, then, on the basis of mere intuition? No. In Sen's view, we ought to measure the 'effective freedoms' or 'capabilities' of individuals under different societal arrangements, which include people's abilities to live the kinds of lives they want to live, given the amount of material wealth or 'primary goods' that they may have. There are good reasons, Sen claims, for "not confusing means with ends, and for not seeing income and opulence and good in themselves" (Send, 2009, 226). The capabilities of individuals, as opposed to other measures, is a more 'direct' indicator of the well-being of individuals, according to this view. The capabilities approach, as Sen calls it, provides a metric that can be used to compare the lives of individuals in different social schemes. It does so without merely looking at people's measures of 'primary goods' like wealth and social capital, which neatly generalize into overall distributive schemes and 'indifference curves', but which do not give accurate portrayals of well-being. It is not necessary, then, to consider how much a particular social arrangement measures up to ‘perfect justice’ when the capabilities approach works even better for making pair-wise comparisons.

The second sense of “necessary” is a weaker claim that states: “if comparative assessments can be systematically made, then that discipline must also be able to identify the very best” (Sen, 2006, 222). Or conversely, if transcendental questions cannot be answered, then neither can comparative ones. But, Sen responds, only with a well-ordered ranking do comparisons result in identifying a “best” option. An example would be “a complete and transitive ordering over a finite set” (Sen, 2006, 223). Presumably, however, we can and do still make comparisons over non well-ordered rankings, such as what would be the best policy to pursue in a given context. Judgments such as these cannot be well-ordered because we are often unsure about how many options really are ‘open’ to us. Sen thinks that the inference which states “if transcendental questions can’t be answered, then neither can comparative questions” is a non sequitur. In other words, he thinks that comparisons are possible without some presupposed or latent transcendental standard. He admits, however, that it is possible to have a ‘conglomerate theory’: “it is, of course, possible to have a theory that does both…but neither of the two types of judgments follow from the other” (Sen, 2009, 16).

A theory that focuses exclusively on transcendental questions, as in Rawls’ ideal theory, can lead to “problematic exclusions”. Sen lists at least six exclusions within Rawls’ framework: (a) ignoring comparative justice, (b) ignoring ‘social realizations’, (c) ignoring global justice by excluding “the possibility of adverse effects on people beyond the borders of each country from the actions and choices of this country” (Sen, 2009, 90), (d) being at risk of ideological or “parochial” values, (e) ignoring the infeasibility of a “unique transcendental agreement" on principles of justice (*ibid,* 10), and (f) ignoring the commonplace reality of unreasonable behavior "through forceful use of the sweeping assumption of compliance with a specific kind of 'reasonable' behavior by all" (*ibid*, 90).[8]

By calling these exclusions "problematic", I assume that Sen means a good theory of justice ought to include these things. If, for example, it turned out that a transcendental theory of justice was at risk of being ideological- in the pejorative Marxist sense of a belief system that obscures and thereby perpetuates systems of oppression- because it excluded discussion of important injustices, then it would be problematic. In chapter 4, I will spend some time discussing whether or not it makes sense to claim that ideal theory is inherently ideological, or whether Rawls' theory in particular is ideological, and what that might mean for the practice of political philosophy.

Some relative advantages of the comparative approach to thinking about justice, on the other hand, include: (g) not keeping us engrossed "in an imagined and implausible world of unbeatable magnificence", (h) recognition of the inevitable plurality of principles relating to justice, (i) the permissibility of partial resolutions, and (j) a wider diversity of inputs and interpretations (Sen, 2009, 106-109). Sen articulates these advantages in terms of social choice theory, which is a largely mathematical discipline that explores the functional relation between people's priorities or values and their choices in particular circumstances.

[8] Sen is careful not to conflate the limitations of a transcendental focus with the limitations of an ideal theory using full-compliance as an initial assumption: "The last item on this list of omissions and commissions has received some attention in the standard literature, in a somewhat stylized form, through the recognition of the need for theories that deal with 'non-ideal' conditions The other items, however, are not helpfully understood in terms of the distinction between 'ideal' and 'non-ideal' theories, and must not be brushed under the same carpet" (Sen, 2009*,* Footnote p. 90).

In his view, social choice theory provides a better framework for thinking about justice than traditional contract theory because, in short, it is more closely tied to the workings of actual democratic societies. For example, social choice theory was originally designed by French Mathematicians after the French Revolution as a way to analyze the voting patterns of a democratic community. As Sen explains in his Nobel Laureate lecture "The Possibility of Social Choice", the primary motivation for these theorists was to develop "a framework for rational and democratic decisions for a group, paying adequate attention to the preferences and interests of all its members" (Sen, 1999, 350). More recently, thanks to pioneering work by Kenneth Arrow (1951), social choice theory has been applied to welfare economics as a way of ranking and weighing various decisions against individual priorities or preferences. The resulting analytic system can be applied to issues of social justice and welfare economics, in Sen's view, because it provides a rigorous mathematical basis for making comparisons between different social systems.

In traditional contract theory, by contrast, the focus is typically on explaining social organization as something that we would or ought to rationally consent to. The remoteness of this rational decision, however, which takes place in a mythical state of nature, or in an idealized 'original position' of initial equality, from normal everyday decision-making procedures, is evident by the fact that no historical societies ever made such agreements. Even though the hypothetical nature of Rawls' contract theory is meant to shield it from criticisms of remoteness like this one, Sen's approach is meant to be more closely tailored to the empirical realities of decision making, and hence, superior to Rawls' hypothetical contract because it tracks actual democratic decision making processes.

The major advantage of the comparative approach and of social choice theory, according to Sen, is that a policy which would remove 'manifest injustice', but that might not put us directly on the path to ideal justice, would be advocated for by the comparative approach while being rejected by the transcendental approach (assuming that it would put us on a path *away from* the ideal of perfect justice). Such a policy might seem attractive, however, for various reasons. These might include the obvious benefits that would follow from removing the injustice, the lack of a clearly defined ideal that we can all agree upon, as well as general skepticism that we might ever reach a 'perfectly just' society, along with problems of feasibility and 'trade-offs' between long term and short term goals. If we agree with Sen's position, then it would require us to rethink our general approach to theorizing about social justice. In particular, it would require us to focus on the removal of 'manifest injustice' as a priority over eventually achieving an 'ideally just' society.

2.6: CONCLUSION

In this chapter I have identified roughly five sets of criteria that are thought to distinguish between two types of theories (ideal and non-ideal):

(a) strict-compliance versus partial-compliance;

(b) idealization versus abstraction

(c) fact-insensitivity versus fact-sensitivity;

(d) desirability versus feasibility (aspirational versus concessive principles)

(e) orientation and starting point (i.e. focus on 'transcendental justice' or on 'comparative justice').

An interesting question to ask at this point in our discussion is "what kind of distinction is being made or proposed by each set of criteria?". Three types of distinctions come to mind: binary, scalar, and categorical.

A binary distinction involves either the presence of a property or its absence. For example, a light switch is either on or off. A scalar distinction, by contrast, indicates a distribution of a certain property along a continuum or scale. For instance, 'redness' indicates how much of the property (red) an object has. We may say that something is either red or not, but this does not adequately distinguish between 'border-line' cases like magenta, which could be either purple or red. Thus, 'redness', a scalar property, is necessary to distinguish just 'how much' red an object has as opposed to some other color. Finally, a categorical distinction does not involve either binary or scalar judgments, but rather it indicates two entirely distinct concepts or classes of things. For example, water and hydrogen peroxide, though made up of the same basic components, hydrogen and oxygen, are categorically distinct types of molecules because of their structures. Another example of a categorical distinction would be the distinction between facts and values, or descriptive judgments and prescriptive judgments. A description involves an 'is' statement, whereas a prescriptive judgment involves an 'ought' statement, neither of which can be reduced to the other. A categorical distinction implies that the definitive property of that category (i.e. chemical structure or 'is' statements versus 'ought' statements) cannot be used to describe or understand an object in another distinct category.

What is of interest, then, is whether the predicate or property 'ideal' is best understood in terms of 'on/off' judgments, a continuum, or rather as describing a distinct class of objects, say 'ideals'. Clearly, something being ideal is not an 'on/off' type of judgment because it makes sense to talk of a particular situation, normative principle, or theory of justice as being more or less ideal in a way that it does not make sense to describe a light switch as being either more or less on or off (excluding, for obvious reasons, lights with a dim-switch).

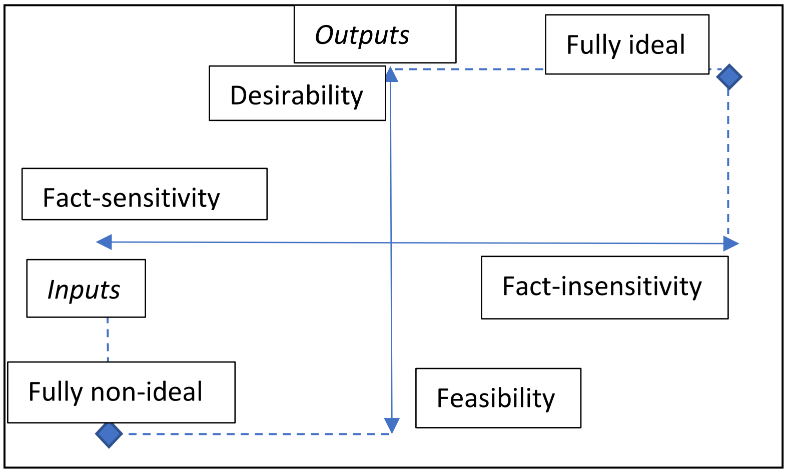
But my guess is that finding the answer to the question of what type of distinction is being made is going to be more difficult than it initially appears. That is because there are just about as many ways of drawing the distinction between ideal and non-ideal as there are theorists who care to talk about it. There have been some attempts, however, to reduce the various criteria down into a single-dimensional or a dual-axis framework. For instance, Ingrid Robeyns (2008) accepts the transcendental/comparative distinction as basically correct, while adding a 'transitional' component to non-ideal theory, thus reducing criteria (a) and (e) into a single axis. She states:

> "In cases in which we are not in a fully just society, we need theory to guide us for two important tasks: first, to be able to make comparisons between different social states and evaluate which one is more just than the other; and, second, to guide our actions in order to move closer towards the ideals of society. The latter is sometimes called the theory of transition." (Robeyns, 2008, 346).

She also separates issues of institutional design and policy making from non-ideal theory, which is a purely theoretical endeavor. In doing so, she restricts some criteria that can count for making the distinction. For example, feasibility problems drop out of view from her framework because feasibility only applies to the workings of institutions and the enacting of policies. Finally, she explicitly excludes Cohen's "pure" theory from the debate, arguing that it is too remote from issues of social justice to really count as a theory of justice: "this kind of theory does not start from a concern with addressing social ills and contributing to justice-enhancing practice, but rather with an interest in knowledge for the sake of knowledge" (Robeyns, 2008, 343). Hence, Robeyns reduces the debate over ideal and non-ideal theories into a single-axis debate between theories of transcendental justice, which provide a picture of "paradise island", and theories of comparative justice and transitional justice, which tell us how to get to "paradise island" (Robeyns, 2008, 343-344).

Alexandru Volacu, on the other hand, proposes to reduce the distinction between ideal and non-ideal theories into a dual-axis framework, between desirability and feasibility on one axis and fact-sensitivity versus in-sensitivity on another, which results in a scalar distribution along which theorists can be placed:

FIGURE 2.1[9]:



[9] This diagram was taken from Volacu, Alexandru. (2018). "Bridging Ideal and Non-Ideal Theory". *Political Studies*. Vol. 66, No. 4, p. 893.

Others, however, like Valentini (2017) and Hamlin and Stemplowska (2012), think that the best way to understand the respective debates is to keep them separate and to recognize the differences in the arguments proposed for each sets of criteria. In the next chapter, I will discuss these views and argue for what I think are the best set of criteria for making the distinction between ideal and non-ideal theory. I will propose that orientation and starting point offer the best criteria that can distinguish between, in my view, two categorically distinct types of theories of justice.

CHAPTER 3: THE ORIENTATION THESIS

3.1: INTRODUCTION

In this chapter I argue that a theory's starting point and orientation offer the best set of criteria to distinguish between ideal and non-ideal theories. I call this *the orientation thesis.* I offer two reasons for this view: (*i*) these criteria (starting point and orientation) offer a categorical distinction and (*ii*) they properly differentiate 'paradigm' examples of ideal and non-ideal theorists, such as Rawls and his critics.

There are two separate questions at stake here. The first is about which criteria are best for making the distinction, which is tied to the question about what type of distinction is being made in the first place. The second question is about the relationship, if there is any, between the two types of theories. An answer to this second question will depend on which criteria are found to be best because, on various readings of the distinction, like Amartya Sen's[10], there is no necessary connection between the two types of theories.

In this chapter, I will focus primarily on the first question. But I will do so only as a necessary step to making what I think is the somewhat more interesting claim that injustice is a phenomenon that can be studied independently from ideals like justice. Put simply, I will argue that we do not need transcendental standards to make simple comparisons between better or worse states of injustice. Injustice, I would contend, is a phenomenon that sticks out to a moral agent with normal cognitive powers. Although there are cases, as we will see in the next chapter, of 'gaps' in the interpretive resources of a

[10] It is important to note that Sen's own characterization of the two approaches to justice is explicitly not conflated with the distinction between ideal and non-ideal theories (see: Sen, 2009*,* Footnote p. 90). But, for the purposes of this thesis, I will be using at least parts of his view to characterize the orientation or purpose of a theory of justice. Other theorists, like Ingrid Robeyns (2008), have also found his approach useful for similar reasons.

community that can effectively disappear certain forms of injustice, this is not a normal state of affairs. Normally, when we see black folks in the US and elsewhere being hauled off to jail for drug charges at massively disproportionate rates than white folks, we can recognize this as an injustice. It is only when there is rampant, systemic, and institutionalized injustice in the world that our theories will fail to make the right identifications. When we find ourselves in a world where there is pervasive injustice, I would argue that mere intuition is far more effective at tracking real injustices.

Hence, I do not think, as many have presumed, there must be a different standard of justice (like Sen's comparative justice and the capabilities approach, for example) that allows us to make comparisons between different social worlds. In my view, comparisons between different states of injustice are possible without a metric of justice because injustice is something which can be recognized independently of any standard.

My overall view, which I will discuss further at the end of this chapter and in chapter 5, relies on there being a categorical distinction between ideal theories on the one hand, as theories concerned with the study of justice, and non-ideal theories on the other hand, being concerned, broadly speaking, with the study of injustice. It also relies on what I am calling *the orientation thesis,* which states simply that theories of justice can be differentiated using the predicates "ideal" or "non-ideal" according to the following criteria:

(a) *The overall purpose or orientation of a theory of justice*, which includes whether it focuses on describing principles of justice for a 'perfectly just' society or whether it focuses on understanding injustice, comparing different states of injustice, or on recommending reforms to eliminate it.

(b) *The starting point of a theory of justice*, which includes what types of questions it asks and what types of questions it does not ask. For instance, ideal theories typically focus on the question: "what is a perfectly just society?", while non-ideal theories ought to focus on the question "what are the causes of injustice?", or, alternatively: "how do we compare different (unjust) social worlds?".

3.2: ON SCALAR AND CATEGORICAL DISTINCTIONS

Many theorists, however, disagree with the presumption that there must be a categorical distinction being made in the ideal/non-ideal theory debate. For instance, Hamlin and Stemplowska (2012), in their review of the literature, argue that the various criteria put forward to date only place theorists along a continuum of more to less ideal, as opposed to engaging in distinct theoretical enterprises. They present yet another set of criteria, which they argue helps to clear up the confusion over scalar and categorical distinctions in the debate.

Hamlin and Stemplowska adopt a distinction between 'a theory of ideals' and a theory of 'institutional design'. A theory of ideals has two components: "one devoted to the identification and explication of individual ideals or principles (equality, liberty, etc.), the other devoted to the issues arising from the multiplicity of ideals or principles" (Hamlin and Stemplowska, 2012, 53). A theory of ideals is concerned with the definitions or characterizations of ideals such as liberty and equality. It also takes the relative weighting of ideals against each other as its secondary focus. A theory that focuses on institutional design, on the other hand, "is concerned with the identification of social arrangements that will promote, instantiate, honour or otherwise deliver on the relevant ideals" (Hamlin and

Stemplowska, 2012, 53). Most understandings of the ideal/non-ideal distinction, they argue, apply primarily to problems of institutional design as opposed to ideals themselves.

The reason Hamlin and Stemplowska give for thinking that both ideal theories and non-ideal theories apply to institutional design is that one main worry in the debate is over the 'impracticability' of envisioned institutions, and practicability applies to the workings of institutions not ideals. To be sure, the practical contrasts with the theoretical. For some idea to be more practical than theoretical, it would have to focus less on why-questions and more on how-questions. The practicable, on the other hand, contrasts with the impossible or the infeasible because to be practicable is to be capable of being put into practice. An ideal, strictly speaking, cannot be practicable. That is because an ideal is not a behavior or practice but rather a way of evaluating practices and behaviors.

Further, Hamlin and Stemplowska point out that what is at the heart of the disagreements between proponents of ideal theory and proponents of non-ideal theory are often criteria that indicate scalar rather than categorical differences. Take (b), from the last chapters summary of various criteria, as an example. This set of criteria contrasts simplified models of phenomena (abstractions) with falsified models (idealizations), taking falsified models to be the mark of ideal theories. Abstraction without idealization, on the other hand, was proposed as the defining characteristic of non-ideal theories. What if we wish to include the motivations of agents in our theory; at what point does a simplified model become an idealized model? It is unclear, Hamlin and Stemplowska argue, whether there is a single point at which we can neatly divide idealized theories and non-idealized theories, which suggests a scalar distribution as opposed to a categorical distinction. The matter of strict compliance versus partial compliance, as well, seems to suggest a continuum because there is no useful point at which it is possible to measure 'formal full compliance' as opposed to complete full compliance or complete non-compliance. Further, given the multiplicity of variables associated with formal full compliance, like the number of compliers and the extent of compliance by each, the assumption of full compliance results in a varied distribution of (non) compliance.[11]

What is of particular interest for our present discussion is Hamlin and Stemploska's response to Sen's characterization of the ideal/non-ideal divide:

> "we accept that the distinction between the transcendental and comparative approaches can itself be categorical. But we also suggest that almost all of the work in making this distinction track any ideal/non-ideal distinction is being done by the assumptions of localness and realism that are imported into the comparative approach; and both localness and realism are surely better conceived as matters of degree." (Hamlin and Stemploska, 2012, 52).

In other words, the comparative/transcendental distinction by itself does not map onto the ideal/non-ideal distinction without the additional assumptions of localness and realism. Localness refers to the focus on feasible alternatives, or those societies within our own 'neighborhood' of possible worlds, as candidates for comparisons. Realism refers to the assumption that actual agents, as opposed to abstract individuals, will be the subjects of the theory. But these assumptions track the idealization/abstraction debate and the desirability/feasibility debate respectively, which Hamlin and Stemplowska argue are matters of degree and not categorical. To clarify, they think that a comparative approach to justice is not by itself an example non-ideal theory: "It would be equally 'comparative'

[11] For similar arguments on 'fact-sensitivity/insensitivity', and other proposed criteria, see: Hamlin and Stemplowska (2012), "Theory, Ideal theory, and the Theory of Ideals". *Political Studies Review,* Vol 10, p. 51.

to address the relative justice of two hypothetical societies, neither of which approximated the world as we know it and where the comparison was independent of any notion of the feasibility of implementing reforms" (Hamlin and Stemplowska, 2012, 52).

According to their view, ideal and non-ideal theories lie on a continuum whereas the theory of ideals and theories of institutional design are categorically separate because problems of feasibility, facts, and policymaking only apply to theories of institutional design and not theories of ideals. This argument seems to fall in line with Estlund's (2014) and Cohen's (2009) view that there are fundamental ideal principles. They may not go so far as to defend the 'normative ultimate', but it appears that their main worry is about feasibility constraints and the possible negative effects these can have on a normative theory's ability to be critical of the status quo:

> "even if we accept, as most political theorists probably do, that the value of justice is constrained by what is feasible – so that a truly unfeasible requirement cannot be a requirement of justice– it would still not follow that in specifying the ideal of justice we must not venture beyond what is feasible" (Hamlin and Stemplowska, 2012, 55).

This implies that philosophers focused primarily on defining values like equality, liberty, and reason, should not be worried about criticisms from the non-ideal side of the fence (including those from political realists, critics of idealization, and utophobes) concerning the infeasibility of the institutions that would instantiate the relevant values. That may go a long way to explain how and (precisely) why the debate between Cohen and Rawls is separate from the debate over ideal and non-ideal theories. One reason for thinking that it is a separate issue is that Cohen is primarily concerned with defending an epistemic thesis about how we come to endorse certain propositions with ethical content. Rawls, on the other hand, is clearly concerned with issues of institutional design and tries to separate the justification for the principles of justice from a comprehensive moral outlook.

Hamlin and Stemplowska's response to other philosophers, such as Ingrid Robeyns' (2008), attempts to use Sen's characterization of the comparative approach to explain the function of a non-ideal theory of justice is that it "leaves out forms of theorizing that we expect to be able to categorize as ideal or non-ideal theory (or both)" (Hamlin and Stemplowska, 2012, 52). That is because, they argue, there is still a possibility that a transcendental approach could identify a 'local maxima' that best represents justice. What they might mean here is that a transcendental theory could conceivably do the work of a comparative theory. But that claim, as Sen clearly shows, does not imply that one type of judgment (comparative or transcendental) necessarily follows from the other.

Is the distinction between a theory of ideals and a theory of institutional design, which Hamlin and Stemplowska propose, itself a categorical distinction? They think that it is because "there is a genuine distinction between ideals (which one believes in, or not) and social arrangements (which one adopts, or not)" (Hamlin and Stemplowska, 2012, 53). One reason we might think that their distinction is not categorical, on the other hand, is that it is being conflated with the distinction between evaluative and normative principles, which I discussed in section 2.4.1 of the last chapter. Notice that their main argument for thinking that a theory of ideals is separate from a theory of institutional design is that issues of feasibility do not restrict the explication of values as such. But the same is true of evaluative principles, which are 'open-ended' when it comes to exploring what options might be considered best according to some account of value. Since normative principles tell us what to do in certain circumstances, however, they are always in relation to some set of feasible possibilities. If it turns out that a normative principle recommends something that is not feasible (either being impossible to bring about by normal agents or because it is ‘too far’ from our known neighborhood of social worlds), then the normative principle must be rejected for a different one. Thus, it appears that Hamlin and Stemplowska’s attempt to categorize a theory of ideals as a separate area of inquiry is not right unless a theory of ideals is explicitly limited to evaluative considerations. But I see no reason why the thoughtful explication of an ideal cannot have a normative element.

Take the ideal of freedom, for example. It seems to be the case that freedom, if it really is an ideal worth pursuing, is a normative ideal. One reason for thinking that freedom is a normative ideal is that un-restricted freedom can be self-defeating. We can’t go pursuing any and all possible wants or desires at all times because, in the long run, this would make us less able to enjoy our freedoms. This would seem to imply that a proper explication of the ideal of freedom would have to include its limits and proper restrictions. The limits of freedom would be expressed as normative principles like “we ought not to steal from others”, “we ought to respect each person’s right to speak their mind”, or simply “don’t drive too fast”. If these judgments are correct, then the explication of an ideal is not limited to evaluative considerations alone, but rather ideals can contain normative elements. Moreover, with the above example of the ideal of freedom and its limits, I have tried to show that feasibility considerations can and do impact the explication of ideals by telling us whether and when they are worth pursuing un-restricted.

It appears, then, that the theory of ideals/theory of institutional design distinction is not adequate to resolve the tensions within the ideal/non-ideal debate. We have seen that this is the case because this distinction conflates a similar one between evaluative and normative principles. Further, because Hamlin and Stemplowska's reasons for denying Sen's approach (or a similar one) only shows that it is possible to have a 'conglomerate theory', which makes both comparative and transcendental judgments. I agree, however, with their view that the other criteria place theorists on a continuum rather than into two separate categories.

Another reading of the ideal/non-ideal distinction that presents it as a scalar type of distinction is Alexandru Volacu's (2018). He proposes that it is possible to reduce the ideal/non-ideal debate down to a two-dimensional framework with "desirability versus feasibility" on the one axis, and with "fact-sensitivity versus fact-insensitivity" on another (893). Fact-sensitivity and abstraction are collapsed into one dimension in his reading of the ideal/non-ideal distinction. This would place theorists along a continuum as either more or less ideal with respect to their use of factual assumptions and the types of limits placed on their normative principles by certain feasibility considerations. To the extent a theorist accepts that feasibility ought to place limits on the types of moral principles that are acceptable, this will place her closer to the non-ideal side of the spectrum, along the axis of feasibility versus desirability. Whereas, along the axis of fact-sensitivity versus fact-insensitivity, if a theorist takes empirical facts to be relevant in formulating normative theories, she will be closer to the non-ideal side of the spectrum. Interestingly enough, Rawls, who is taken to be the paradigm example of an ideal theorist, does not fall all the way on the extreme side of the ideal on either axis because his assumptions are neither fully idealized nor entirely utopian.

I think this framework can be useful in understanding the various aspects of the debate. But if it is possible is to show that the use of fact-based assumptions and the limits that are placed on normative principles depends on the purposes of a particular theorist, then this would only tend to suggest that starting-point and orientation are categorical in the way they divide theorists, whereas other criteria are not. For example, if I want an evaluative theory that can measure the "overall justice" of my current society, then fact-independent principles might seem appealing to me because I want to be critical of the status quo. If, however, I want a theory that can be used to guide actual behavior, i.e. a normative theory for real people, then I might want to consider more factual assumptions. This shows, I think, that other criteria are dependent on the overall orientation of a theorist, meaning their goals, purposes, and focus.

My view can be seen as a broadly teleological view of the distinction between ideal and non-ideal theories, which, it could be countered, is not going to be a solid enough foundation to rest the kind of categorical distinction required for it to make sense. For example, one might claim that a theorist's motivations can change from one time to another or that they could be multi-dimensional. This would mean that a theorist focused on transcendental justice might also occasionally venture into non-ideal territory, as Rawls does. If it is possible for a theorist to simply change their minds or to focus on more than one goal at a time, so the objection goes, then this set of criteria would not result in a categorical distinction because there may be a highly ideal focus at one time and a less-ideal focus at another time by the same theorist doing the same work.

I would respond that although the criteria I have proposed are broad, it is still possible to determine the *overall* focus of a theorist. I do not wish to deny, however, that a theorist may have multi-dimensional focus, which would mean that they are interested in doing both ideal and non-ideal work. I would also claim that the vagueness of a criterion such as ‘orientation’ does not by itself suggest that, given a particular theorist, it is impossible to identify an ideal focus over a non-ideal focus overall. Thus, even though the criteria I have proposed are broad, this does not show that they do not result in a categorical distinction.

3.2.1: WHAT DO WE WANT FROM A THEORY OF JUSTICE?

There is another, bigger-picture reason, why the overall orientation of a theory of justice, along with its stated starting point, is the best way to distinguish between ideal and non-ideal theories. That would be the very reason behind why we want to have a theory of justice at all; i.e. what we want it to do for us. The focus in much political philosophy in recent years has been over what philosophers would call ‘distributive justice’. The intuition behind this idea is that justice is, as Cicero said, “giving each their due”. Presumably, then, when everyone has what is entitled to them, society is just. Entitlements can come in many forms. Commonly we speak of them in terms of rights and freedoms, shares in wealth or material goods, or social positions. Of course, the issue of specifying what each person is “due”, as a matter of justice, is left entirely unspecified by these types of intuitions. A theory of justice, then, attempts to tell us what type of society would bring about a distribution of entitlements in which each person gets their due.

But there are some intuitions about justice that are left out of this picture. For example, Charles Mills criticizes Rawls’ theory of distributive justice for leaving out the question of a just rectification for historical harms based on race (such as slavery, colonialism, genocide, etc):

> “If, post-Rawls, the central question for political philosophy has become the justice or injustice of the “basic structure” of the *polis*, think how radically this question must be rethought for those whose non-consent completely

> undercuts the contractarian underpinnings of contemporary distributive justice theory, demanding instead that rectificatory justice should be our focus" (Mills, 2015, 7).

Racial justice, it seems, demands at least some level of rectification for past wrongs, which first requires acknowledgement of them. A "color-blind" system of distributive shares, which might be considered impartial and fair, neglects to acknowledge these past historical injustices. Liberal theorists often debate over whether policies of affirmative action can be justified on the basis of past historical wrongs within a framework of "procedural justice", which treats each person as an abstract individual and not as someone with a race and a gender. The point I want to draw from this, however, is that there can be many things that a person could want a theory of justice to do. In general, though, it seems that we want our theories to include as many of our intuitions about justice and injustice as possible.

Other types of intuitions, that do not fit neatly under an account of distributive justice, might bring us to develop theories with altogether different focal points.[12] There can be 'restorative justice' to deal with the lasting effects of warfare, colonialism, and other historical traumas that a society undergoes. "Gender justice" is a concept used by some feminist political philosophers, such as Ingrid Robeyns. She describes her view as a 'partial' theory of justice because it focuses on only one domain: "we may defend principles of justice telling us what is required for complete gender justice, while remaining silent on all other domains of justice." (Robeyns, 2008, 344). Some speak of 'climate justice' to address issues concerning the deterioration of the environment caused by human

[12] For an expanded critique of the "distributive paradigm" in political philosophy, see Iris Marion Young, (1990), *Justice and the Politics of Difference,* Princeton N.J. Princeton University Press. Or, for a discussion of the tendency to "reduce" injustice to the breakdown or rejection of justice, see Shklar, Judith. (1989). "Giving Injustice Its Due". *The Yale Law Journal.* 98:6, 1135-1151.

civilization. Still others refer to ‘animal rights’ as a means to achieving a more just arrangement between humans and other species.

These remarks suggest that the orientation of a theory of justice is determined in part by what one sees as the most important set of intuitions at stake. If that is the case, then it will be possible to differentiate theorists on the basis of their approach because not everyone shares the same intuitions.

So far, I have argued that orientation and starting point are the best criteria for making the ideal/non-ideal distinction because they offer a categorical distinction and because that they properly differentiate various theorists as either ideal or non-ideal. The divide between ideal and non-ideal theories is best interpreted as being about the orientation of a theory of justice. Insofar as a theory focuses on justice, it is ideal, insofar as it focuses on *injustice* it is non-ideal.

3.3: THE ARGUMENT FROM PATH DEPENDENCE

Now that the matter of whether the ideal/non-ideal distinction is categorical or not can be settled, in this section I approach the question of whether there are any necessary connections between ideal and non-ideal theory. Two opposing views are discussed. First, I outline A. John Simmons’ reconstruction of the distinction between ideal and non-ideal theory in Rawls’ work. Simmons argues that ideal theory must have priority over non-ideal theory because without a conception of the perfectly just society, there would be no goal to orient our path towards it. Second, I discuss Gerald Gaus and Keith Hankins’ view that there can be no accurate knowledge of far-flung social worlds which embody ‘perfect justice’, and hence, no clear path towards such a society.

My view, which draws heavily from Amartya Sen's formulation of the ideal/non-ideal distinction, lies somewhere in-between these two opposed positions. Both Simmons and Gaus et. al. point to problems with Sen's view that there is no necessary connection between comparative and transcendental judgments. If Sen's view cannot be rescued from these criticisms, then it might have to be modified in order to maintain his central thesis regarding the distinction between ideal and non-ideal theory, which is that neither type of judgments (comparative or transcendental) follow necessarily from the other.

Simmons, however, is mostly concerned with interpreting Rawls' basic framework for the distinction between ideal and non-ideal theory correctly. But his argument is also indirectly aimed at Sen's view of comparative justice. Simmons states that there is general confusion and lack of detailed analysis on Rawls' intended use of the distinction: "those who have adopted the Rawlsian language of ideal and nonideal theory seem mostly to have taken Rawls's version of the distinction to be obviously correct. And those who criticize the Rawlsian approach seem mostly to dismiss it as flawed in relatively simpleminded ways" (Simmons, 2010, 6).

In Simmons' reading of Rawls, ideal theory provides the framework for what types of institutions and behavioral norms are to be considered just. A just society is one in which there is a full realization of the principles of justice in the basic structure of a society through full compliance by all citizens. Simmons comments on the necessity of assuming strict compliance in ideal theory as (possibly) a way of assessing the content of the principles of justice in the original position: "if we compare the operation of societies ordered by competing principles of justice while assuming strict compliance with those principles, the different effects we observe can reasonably be taken to be wholly the responsibility of the different ordering principles themselves" (Simmons, 2010, 8). Under this interpretation, the assumption of strict compliance allows a way to imagine what it would be like to have institutions that embody the principles of justice, thus setting up a goal to orient our path towards achieving a just society.

Non-ideal theory, on the other hand, is concerned with *transitions* from the status quo to the ideal, or in other words, transitions from partial-compliance to full-compliance. Rawls also provides us with a further division of non-ideal theory into principles that deal with "unfortunate circumstances", resulting in non-compliance, and principles that deal, on the other hand, with "deliberate non-compliance" (Simmons, 2010, 16). For the first part, Rawls has in mind societies with "unfortunate cultural or economic histories", which prevent them from achieving just institutions. Rawls also considers societies in which there are new "contingencies" or "crises" that must be dealt with before going on to work on achieving just institutions (Simmons, 2010, 16). In these cases, it may be permissible to restrict the overall extent or set of liberties while still maintaining a minimal scheme of equal liberties. Societies in non-ideal circumstances such as these, Rawls explains, ought to be governed by "the general conception of justice", which does not have the lexically ordered principles in mind. The second part of non-ideal theory, which I touched on in section 1.1, is concerned with justifications for civil disobedience, just war theory, and other forms of "deliberate non-compliance" (Simmons, 2010).

What is clear from Simmons' analysis of the ideal/non-ideal distinction in Rawls is that non-ideal theory is concerned primarily with transitions, not comparisons. He also brings us the concept of *transitional justice*, in contrast to Sen's idea of comparative justice, which is meant to guide social reform towards a just society: "A good policy in nonideal theory is good only as transitionally just - that is, only as a morally permissible part of a feasible overall program to achieve perfect justice" (Simmons, 2010, 22). Understood in this way, non-ideal theory is concerned with justifiable transitions towards a completely just society, using the principles of ideal theory as a goal to orient our path and as a measure for the acceptableness of policies to be pursued.

Simmons disagrees with Sen on his characterization of the purposes of a non-ideal theory in terms of comparisons. He states: "We can hardly claim to know whether we are on the path to the ideal of justice until we can specify in what that ideal consists" (Simmons, 2010, 34). Simmons points to the possibility of a policy or reform that is intended to 'advance' justice by eliminating a particular form of injustice could have the opposite effect by making an ideal society harder to achieve in the long run. If we took "dramatic possible legal steps" towards achieving a racially just society, for example, we might suffer "a conservative backlash that would leave us stranded forever or for a much longer period of time in a deeply unjust society" (Simmons, 2010, 21). In other words, if we want to achieve *justice,* we must not only look at whether we are advancing ourselves towards a (locally) better state of affairs. In some cases, Simmons argues, that could lead us further away (measured in terms of feasibility) from achieving overall justice.

Whether we are able to take Sen's claims seriously or not, it seems, depends on what he means by 'advancing justice'. If advancing justice means putting us on a path towards a local optimal, then unless the local and global optimal are the same, it is not always clear that local improvements, even if they are justified for various reasons, will always take us closer to the ideal of a just society. If, on the other hand, 'advancing justice' means putting us on a path towards a completely just society, then it is simply not true that ideal theory and non-ideal theory are analytically disjoined.

Simmons' argument appeals to the plausible idea of path-dependence, which is the claim that we must always begin from the status quo when we are recommending reforms or policies for institutional design. If non-ideal theory is concerned with 'advancing justice' from the status quo to a more just world, then it must specify some goal by which to orient our path. Thus, on this reading of the distinction, non-ideal theory must be dependent on ideal theory because without a goal there is no specifiable path and, for all we know, we might be headed towards a worse-off situation by focusing primarily on comparing better or worse (local) alternatives. Simmons responds, more directly, to Sen's view in this way:

> "which of two smaller "peaks" of justice is the higher (or more just) is a judgment that matters conclusively only if they are both on equally feasible paths to the highest peak of perfect justice" (Simmons, 2010, 35).

Now, what does Sen really mean when he says that a policy or reform can "advance justice" without moving us closer to a transcendental ideal? Unless 'justice' is to mean something else in the comparative case, it cannot be an advancement in justice to move to a (locally) better state of affairs, which, in the long run, would actually take us further away from a just society. Simmons' response to Sen seems to illustrate this difficulty because, although it is true that we do not need to mention Everest in order to compare the relative height of two different mountains, it is not quite as clear that we do not need some standard of justice to compare the relative just-ness of different societies or policy proposals.

Simmons' reading of Rawls suggests that non-ideal theory is essentially a theory of *transitional justice,* which is supposed to govern the movement from the status quo closer to the ideal of a perfectly just society. It uses the principles of justice as a benchmark to judge the relative 'grievousness' of deviations from perfect justice. I take Simmons' interpretation of Rawls to be mainly correct because it stresses the importance of the orienting function of ideal theory, and thus, of what I think should be thought of as the proper way to draw the ideal/non-ideal distinction. But there are at least two different ways of understanding the orientation or function of a theory of justice.

In one sense, we can view an ideal theory of justice as setting up a long-term goal that should 'orient' our search for justice in the real world, even if it doesn't necessarily tell us how to get there. In another, however, we can view a theorist as having a particular 'orientation' towards their subject matter when thinking about justice. Gerald Gaus and Keith Hankins explore some difficulties with the first approach to the 'orienting' function of ideal theory in their article "Searching for the Ideal: The Fundamental Diversity Dillemma" (2017).

First, they outline Simmons' response to Sen as making the basic claim that the terrain of social justice is a 'rugged landscape', as opposed to a 'single peaked' landscape. By which they mean that when making comparative judgements between social worlds (a), (b), and (c), if (c) is thought to be best with respect to some conception of justice and all three share the same 'justice-relevant' features (i.e. features that make comparisons between the societies possible, like what types of institutions there are, how they operate, etc), then it is possible to plot each social world along a single (x) axis, in which a move from (a) to (b) and from (b) to (c) yields an increase in overall justice (along the y axis) (Gaus and Hankins, 2017, 180-181). Sen seems to be saying that the 'terrain' of social justice is much like this simplified 'single-peak' distribution. If, on the other hand, we imagine a set of social worlds (a) through (n), in which there are multiple peaks and valleys that relate to justice and injustice, then we are presented with a ‘rugged landscape’ (or ‘NK’) optimization problem:

> “*Pace* Sen, in rugged landscapes…a constant series of pairwise improvements can (*i*) lead to a local optimum that is far inferior to the global optimum and (*ii*) can also lead us away from the global optimum social world” (Gaus and Hankins, 2017, 184).

Next, Gaus and Hankins introduce a further problem called the ‘neighborhood constraint’. Given the complexity of social systems, they argue, we have reason to doubt the truth of claims to have knowledge of the global optimal: “as we evaluate more far-flung social worlds, with features very different from those we are familiar with, our understanding of those worlds institutional dynamics, and so their justice, becomes more speculative” (Gaus and Hankins, 2017, 184). Thus, the neighborhood constraint articulates the plausible idea that we can only claim to have accurate knowledge of worlds within our own ‘neighborhood’. But if we accept the ‘rugged landscape’ analogy along with the neighborhood constraint, then this leads to a dilemma.

One the one hand, if we can only have accurate knowledge of the justice-relevant features of social worlds within our own ‘neighborhood’, then in a ‘rugged landscape’ pair-wise improvements might take us further way from the global optimal. That is because it might be an improvement relative to the actual world to move to the local optimal within our ‘neighborhood’. But this improvement might actually be worse-off relative to the global optimal. On the other hand, if ideal theory is meant to orient our “quest for justice”, but the ideal is not within our own neighborhood, then it is hard to see just how this would be possible (Gaus and Hankins, 2017, 188).

One way to avoid the problems raised by this dilemma might be to encourage diverse perspectives into our search for the ideal. In this view, when a more diverse set of evaluative criteria are put forward to determine which features of a set of social worlds might be considered justice-relevant, it might be the case that there will be some single perspective that can identify a global optimal within our neighborhood. Imagine a philosopher who, after arduous years of weighing principles of justice against feasible social systems and moral intuitions in the Rawlsian fashion, claims to know exactly what steps we might take to move ourselves directly from the status quo to the ideal of a perfectly just world. Gaus and Hankins call this idea the 'utopia is at hand theorem', which states: "there are always perspectives that show that the ideal is within our own neighborhood" (Gaus and Hankins, 2017, 195). Thus, if we include more diverse perspectives into a discussion about social justice, then there might be a single perspective that we could all agree includes the global optimal into our own neighborhood.

But an increase in diverse perspectives will also make it more difficult for people to effectively communicate which features of a social world are justice-enhancing or justice-relevant. That is because an evaluative perspective, if it is to communicate a meaningful ordering of social worlds, must have at least five components:

(a) A set of evaluative standards;

(b) An identification of the relevant features of social worlds;

(c) A mapping relation from (a) to (b);

(d) An ordering of the underlying structures that meaningfully relates them in terms of similarity; and

(e) A distance metric (Gaus and Hankins, 2017, 181).

If we include diversity in distance metrics, which measure the relative distance from one social world to another, then we are likely to increase our chances of finding a global optimal within our own neighborhood because some perspectives might view the best option as feasibly close while others might view it as far away. But once we allow for diversity at this level, there will also be diversity at other levels, such as (d) the ordering of worlds in terms of similarity and (b) the identification of relevant features of a social world, because how we order different social worlds will depend on what features we take to be relevant or not. The relative distance from one social world and another, measured in terms of feasibility, will depend in turn on levels (b) and (d) because social worlds that share more in common with our own will come out as 'closer' than others that are more dissimilar. But diversity at more 'fundamental' levels like (b) creates "intractable problems" of communication according to Gaus and Hankins: "the utopian must be saying that our current perspective leads us to believe that the ideal lies outside our neighborhood only because we are mistaken about what the fundamental features of the social worlds are" (Gaus and Hankins, 2017, 196). When we are presented with a utopian view such as this, our reactions (if we do not share the same perspective) must be either "(*i*) the utopian perspective is simply erroneous, or (*ii*) that it is using a different set of evaluative standards" than ours (Gaus and Hankins, 2017, 197). Thus, we reach the dilemma: either (a) get stuck in our own neighborhood by not allowing diversity in evaluative perspectives or (b) be unable to agree on which features of a social world are relevant to their ordering within a rugged landscape. Gaus and Hankins conclude that no approach to political philosophy, ideal or non-ideal, can avoid this problem and the choice that it presents.

Gaus and Hankins' critique illustrates the problems associated with the long-term goal view of the orientation of a theory of justice. If we view the 'orienting' function of ideal theory as setting up a long-term goal, then there may be no solution to the problem of diversity in evaluative perspectives, as Gaus and Hankins illustrate. But there is a broader sense of the orientation of a theory of justice, which I think is more important. Orientation can also refer to the overall purpose and focus of a theory of justice, which includes more than just long-term and short-term goals for social reform.

In common between both of the views just discussed is the assumption that an ideal must serve some kind of normative 'orienting' function. In other words, it is often presumed, either for the sake of argument or as a genuine belief, that an ideal conception of justice is necessary for any philosophical perspective on social justice. Moreover, it is presumed that ideal theory can be helpful for actual work in policymaking and activism. But many of the criticisms of ideal theory already discussed suggest that it is unable to perform this function adequately. For instance, Mills asks: "why should anyone think that abstaining from theorizing about oppression and its consequences is the best way to bring about an end to oppression?" (Mills, 2005, 171). Weins, similarly, suggests that ideal theory is only normative for a society of angels and Gaus and Hankins hold that neither approach is going to be adequately normative for activism and policymaking. Have political philosophers simply been wasting their time? What is the use of an ideal theory of justice? In the next chapter, I will discuss these questions by attempting to relate them to debates from feminist philosophy and critical race theory.

3.4: CONCLUSION

By way of a summary to this chapter, I will return briefly to the two methodological questions posed in section 3.1: (1) "which criteria best distinguish between ideal and non-ideal theories of justice?" and (2) "is there any necessary connection between these theories?". In discussing responses to Sen's thesis, which denies (2) and proposes the comparative/transcendental distinction as the best answer to (1), I gestured towards an argument for what I call *the orientation thesis.* I claimed that a theory of justice's starting point and overall focus or orientation provide the best criteria to distinguish between ideal and non-ideal theories. This is a broader formulation of Sen's view, which I think can avoid the problem of path-dependence. I argued that these criteria are promising because (a) no other set of criteria result in a categorical distinction and (b) other criteria do not differentiate paradigm examples. I showed that when we take the ideal/non-ideal distinction to indicate a distribution of theories/theorists ranging from more to less ideal, paradigm examples, such as Rawls, don't get placed as fully ideal. Moreover, I argued that when we take a broad view of the orientation of a theory of justice, as referring to the overall purpose of a theory, the ideal/non-ideal distinction can be drawn in a way that implies a categorical distinction.

But what was the point of arguing for this view of the ideal/non-ideal distinction? Or, alternatively, what advantages can be drawn from such a view? For one thing, the conclusion that ideals and injustices can be studied independently from each other is relevant to the practice of political philosophy because it is often presumed that the opposite is true. Most people intuitively accept the idea that one cannot even think about injustice without somehow presupposing some view of justice. But if that's not true, then a theory of justice taking on a certain directionality- i.e. either beginning with injustice and working towards and account of justice or *vice-versa*- might affect how effective it can be at recommending solutions to problems. In other words, one might either begin with an attempt to understand justice, as such, and then derive a conception of injustice, or one might do the opposite by trying to understand injustice first. This is important because one method or approach might be more effective than the other. In the next chapter, I will show that there are considerable disadvantages to beginning with an account of justice and then trying to apply that account to difficult cases. Working from injustice instead, although perhaps a messier endeavor, does not obviously have the same difficulties because, as I have suggested already, the non-ideal method that begins with injustice leaves room for a plurality of principles that relate to justice and injustice.

CHAPTER 4: RACE, GENDER, AND IDEAL THEORY

4.1: INTRODUCTION

In this chapter I return to some important critiques of ideal theory that were touched on in sections 2.1, 2.2, and 3.3. In the last few chapters I provided an exposition of the distinction between ideal and non-ideal theories of justice and I gave an argument for at least one way of viewing that distinction. The main question I wish to address here, however, is: "How useful is Rawls' ideal theory approach for political theorists and philosophers interested in race and gender?". In particular, I am interested in discussing whether Rawls' ideal theory can be applied to issues of racial and gender *injustice.*

But, in order to do so, I will lean on arguments that utilize their own conceptions of what makes the distinction between ideal theory and non-ideal theory. For example, I will discuss critiques from Susan Moller Okin and Thomas McCarthy concerning fact sensitivity and idealization. I will also discuss Charles Mills' critique of ideal theory as an ideology, which he develops using his own characterization of the abstraction/idealization debate using the concept of modeling. Both critiques, however, address issues that are related to applying Rawls' ideal theory to non-ideal cases. The important point to draw from them is that, regardless of how we want to draw the ideal/non-ideal distinction, certain problems will still arise from the attempt to work from an ideal conception of justice towards non-ideal, action-guiding, principles.

4.2: FEMINIST CRITIQUES OF IDEAL THEORY

Since Rawls first published *A Theory of Justice* in 1971, feminist philosophers have criticized his work for ignoring the important issue of justice between the sexes and for relying too heavily on patriarchal norms and values. For instance, feminists interested in gender differences have contrasted an ethics of care with ethics of justice and rights. These theorists, including Carol Gilligan and Nel Noddings, argue that the focus on universalizable principles of justice obscures the particularistic, contextual, and care-based ethics of women: "justice has been much overrated as the fundamental virtue, and principles have been overvalued as a tool for thinking about ethical problems…[j]ustice itself, according to this view, should be at least supplemented, if not supplanted, by an ethic of caring" (Okin, 1989, 247). Perhaps Rawls would agree that justice is but one of many important virtues for a political community. But the criticism rests on the claim that presenting justice as "the first virtue of social institutions" obscures the fact that justice depends on the care-based ethics of women. Hence, justice should be viewed as the *second* virtue of social institutions, which would lower its status as a normative priority.

Others, like Iris Marion Young, have also argued that "the ideal of impartiality and universality in moral reasoning is misguided and works in opposition to feminist and other emancipatory politics because it attempts to eliminate otherness and difference and creates a false dichotomy between reason and feeling" (Okin quoting Young, 1989, 247). Rawls' focus on arriving at principles of justice through impartial reasoning, then, is actually far more biased than it initially comes across because it is influenced in part by patriarchal norms and do not apply universally across different cultural expressions of gender. These feminists further criticize Rawls for his use of the false dichotomy between reason and emotion. Reason depends on a stable emotional character, which cannot be developed without the appropriate amount of care. The care many receive as children (often from women who are primary caregivers) is often particularistic and partial. Without it, individuals would not develop the self-respect and sense of self-worth that is necessary to make autonomous rational decisions. Hence, the dichotomy between reason and emotion, where emotions have a negative influence on one's ability to make rational choices, is unfounded. This would mean that the ethics of care is just as important, if not more important, than the ethics of justice.

In contrast to these critiques, however, feminists like Susan Moller Okin think that under the best possible interpretation of Rawls' views (particularly his views on moral development in *Theory* p. 431-492), deliberation about the principles of justice relies on developed moral feelings of empathy and care. Without these moral feelings, Okin argues, Rawls' theory would not produce principles that have such benevolent effects. The two principles of justice require us to take into consideration the prospects of as many lives as possible, especially that of the "least advantaged" life.

The problem with interpreting Rawls as excluding an ethics of care comes from an over-emphasis of his Kantian influences. Okin writes: "central aspects of the Kantian heritage especially the presentation of moral subjects as, above all, rational, autonomous, and freed from contingency-influence Rawls in the direction of perceiving what he is doing as a branch of rational choice theory" (Okin, 1989, 240). But despite Rawls' own characterization of his theory, the rational choice interpretation is implausible, according to Okin, because there is no aspect of rational choice theory that fully encompasses what Rawls is up to in his formulation of the original position. Neither choice under certainty, choice under risk, nor choice under uncertainty seem to completely describe the deliberations that take place in the original position (Okin, 1989, 240-244). Rather, Rawls requires more of his deliberators than simple rational, self-interested, decision making. The veil of ignorance, along with other assumptions, together combine to generate principles that would have been required of perfectly benevolent subjects without actually assuming that human subjects are in fact benevolently motivated.

Okin argues that one must take the whole host of assumptions within Rawls' theory, without focusing too heavily on any one in particular, in order to arrive at the correct interpretation. In this view, Rawls' requires subjects in the original position to be 'functionally benevolent'. Or, in other words, the original position achieves the same thing as benevolence by restricting the type of knowledge available to the deliberators, which in effect 'forces' them to be considerate of others. When we are not able to know which social position we are going to occupy, we must then consider all possible positions as our own and choose from that perspective. Thus, the original position owes less to Kant's view of the human subject as autonomous and rational than it does to the requirement that subjects be altruistic in their moral deliberations.

To be sure, this reading of Rawls amounts to a considerable *revision* on the reasoning for the principles of justice. In Section 40 of *Theory*, Rawls explicitly states that his formulation of the original position is "a procedural interpretation of Kant's conception of autonomy and the categorical imperative" (Rawls, 1971, 256). Rawls sees the original position as specifying the conditions under which rational agents, who view themselves as free and equal persons, would choose principles by which to act. Moreover, his two principles of justice are an elaboration of Kant's categorical imperative, and in particular his idea of a kingdom of ends. One way to see this point is by looking at Rawls' response to Sidgwick's critique of Kantian ethics:

> "He remarks that nothing in Kant's ethics is more striking than the idea that a man realizes his true self when he acts from the moral law, whereas if he permits his actions to be determined by sensuous or contingent aims, he

> becomes subject to the law of nature. Yet in Sidgwick's opinion this idea come to naught…Kant never explains why the scoundrel does not express in a bad life his characteristic and freely chosen selfhood in the same way that a saint expresses his" (Rawls, 1971, 255).

Sidgwick's critique is that Kant does not tell us how to differentiate between the justice of thieves or "scoundrels", as it were, from "true justice" because both are equally in accordance with freely chosen principles. Rawls thinks that his contribution to Kant's thinking is to provide an account of exactly *which* principles one would have to live by in order to fully express their moral self and live in accordance with the moral law. The principles of justice, then, are deeply connected with Kantian ethics and with a view of morality as a type of rational choice. Insofar as we take Rawls' own words into account, then it seems that Okin will have to bear the burden of proof in showing that benevolence really plays a larger role in his system than rational choice and autonomy. It is beyond the scope of this paper, however, to delve into that question. For now, suffice it to say that Okin's argument at least initially seems plausible and would imply that the dichotomy between reason and feeling is overstated by some other feminist critiques.

Okin's reading of Rawls also suggests that his theory can be modified so that gender is included among the relevant social positions that would have to be taken into consideration in the original position. One reason Okin gives for thinking that Rawls should consider gender in the original position is that the "monogamous family" is included in "his initial list of major institutions that constitute the 'basic structure' to which the principles of justice are to apply" (Okin, 1989, 235). Since the family is part of the basic structure and is arguably where moral education takes place, the family structure ought to be a subject of justice with the "least-advantaged" positions in the family- typically occupied by women- being considered in the original position. But Rawls in fact does not follow this route because the deliberators in the original position are taken to be "heads of families" (Rawls, 1971, 146). Since their relative position as heads of families is already settled, the deliberators are not able to settle questions of justice *within* the family. In other words, treating the deliberators in the original position as heads of families prevents them from taking up a perspective of someone with a lesser position in the family.

But if the family is an institution that makes up the basic structure, then it seems to make sense to talk about the justice or injustice of families. Moreover, if it makes sense to talk about the family as a distributor of benefits and burdens, then it would seem that the principles of justice, including the difference principle, ought to apply to the structures of the family. But Rawls does not include families as a primary subject of justice in his theory. Why, then, if the family is a major institution in the basic structure of society and if it is a place where benefits and burdens get allocated, should it not be a subject of the two principles of justice? Rawls eventually did come to realize that the stability of justice as fairness would depend on taking the relative positions of men and women in the family into account, but it is important to see exactly how he came around to that view.

In *Justice as Fairness: A Restatement,* Rawls argues for not including the family as a primary subject of justice while acknowledging its status as a basic institution: "political principles do not apply directly to its internal life but they do impose essential constraints on the family as an institution and guarantee the basic rights and liberties and fair opportunities of all its members" (Rawls, 1999, 164). This is partly because, despite being an important political institution, the principles of justice are "out of place" in the family, where altruism and love are supposed to take precedence. Apart from prohibiting obvious wrongs like child abuse, domestic abuse, and other harms, society at some point must simply trust "the natural affection and goodwill of parents" (Rawls, 1999, 165).

But Okin' point is that Rawls' view actually requires that gender and the family be taken into consideration from behind the veil of ignorance. Since families are incubators of the sense of justice that would be required of citizens in a well-ordered society, unjust distributions within the family must be recognized: "If children see that sex difference is the occasion for obviously differential treatment, they are surely likely to be affected in their personal and moral development" (Okin, 1994, 12). Therefore, in order to avoid the continuation of gender-based inequalities in society, the principles of justice would have to apply directly to the family.

Citing Hirschman's "differential exist potential theory" of spousal relationships, Okin argues that the family is in fact a place in which benefits and burdens are distributed. The theory states that relative bargaining power in a relationship is gained whenever more material wealth or social status is held by one of the partners (Okin, 1994, 16-17). This implies that the lack of work opportunities for women outside the home is going to contribute directly to their lack of bargaining power in the home. When we consider this in turn with the gendered division of labour, in which women are expected to be primary caregivers within the family, the theory helps to explain women's unequal positions in society compared to their male counterparts. The family, it seems, unjustly allocates a greater share of burdens onto women. Yet this is not considered a violation of justice as fairness because the deliberators in the original position are viewed as "heads of families", which obscures their gender (or rather assumes its maleness) as well as their relative position within the family.

This is a clear example in which the limitations of ideal theory are brought to the foreground. The problem Okin seems to be pointing to is related to fact-sensitivity- i.e. “how should facts about gendered divisions of labour affect our choice of action-guiding principles?”- and abstraction- or “how accurate do our empirical models of social phenomenon have to be in order to effectively inform our normative principles?”. Viewing deliberators in the original position as heads of families, for example, abstracts away from the differential positions of men and women in the family. Hence, theorizing in this way involves using bad idealizations which run the risk of obscuring real injustices. This is a problem for ideal theory not only because it obscures gender injustice as an essential problem in society but also because it could have a justificatory effect on existing patriarchal norms. Moreover, if we accept that abstracting away from gender in our theories is a bad thing, it is not entirely clear how much including facts about gender will lead to changes in other parts of the theory.

By the time Rawls wrote *Political Liberalism,* however, he was deeply concerned with the types of feminist critiques that I have just outlined: “after a second round of criticism by Okin and other feminist theorists, Rawls briefly addressed the matter in his 1997 piece ‘The idea of Public Reason Revisited’” (McCarthy, 2009, 29). In that piece, Rawls recognizes that the gendered division of labour in the family has been implicated in causing inequalities between men and women all over the world. He argues that similar divisions of domestic labour would be acceptable in a well-ordered society only if they were: “‘fully voluntary’ and arrangements were made to ensure that it did not undermine the equal liberties and opportunities of women” (McCarthy, quoting Rawls, 2009, 29). So, later in his life Rawls seems to admit that the position of women in the family would have to be made a part of ideal theory. Thomas McCarthy attributes this later inclusion to Rawls' recognition that gendered divisions of labour are "general" facts about society. This would seem to make it possible to include gender and gendered social positions within the framework of ideal theory.

Recall that the general facts about human psychology, society, economics, and social organization that would be necessary to design a just and feasible basic structure are all accessible to deliberators in the original position. The purpose of using these facts in an ideal deliberative process, as we have seen, is to construct a 'realistic utopia' that takes "men as they are" and "laws as they might be" (McCarthy, 2009, 30). "Men as they are" is taken to mean their "moral and psychological natures". "Laws as they might be" are really laws as they should or ought to be, which still requires deliberators to have access to knowledge of feasible social structures, the basic conditions of social cooperation (i.e. the circumstances of justice), and other types of general knowledge.

Particular knowledge, however, of such things as "the particular circumstances of their own society", "which generation they belong", or "the relative good or ill fortune of their generation" (McCarthy, 2009, 31) is systematically excluded from ideal theory on the grounds that it is morally irrelevant. Rawls writes: "One reason why the original position must abstract from the contingencies- the particular features and circumstances of persons- within the basic structure is that the conditions for fair agreement between free and equal persons on the first principles of justice for that structure must eliminate bargaining advantages that inevitably arise over time within any society as a result of cumulative social and historical tendencies" (Rawls, 1999, 16).

One important question to ask here is: “should gendered divisions of labour really be considered as general facts about society and thus, accessible to ideal theory?”. Doing so might imply that they are somehow natural or inevitable. But this is not what we want a theory of justice to do because it would tend to normalize or rationalize the status quo between men and women. Moreover, we might ask: “can an ideal theory of justice adequately formulate normative principles for issues like gender at all?”.

Charles Mills thinks that Okin’s attempt to include gender from behind the veil is “not (somehow) Rawls’ real view- certainly not the Rawls who did not even mention sex as something you know behind the veil!” (Mills, 2005, 179). Mills’ point is that extending the principles of ideal theory into non-ideal territory does not thereby make ideal theory less exclusionary. If ideal theory does not properly allow us to consider whether there are unjust relations within family structures or whether such relations contribute to the unequal treatment of women in society, then the problem is not simply that ideal theory needs to be modified or expanded but rather that it is essentially exclusionary.

The problem of applying Rawls’ ideal principles to issues they were not meant to deal with is brought to light in a debate between Mills and Tommie Shelby. Shelby has extensively argued that Rawls’ principles of justice can be applied to problems of racial injustice. The arguments presented by Mills against Shelby’s view can apply to ideal theories of gender justice as well. Mills claims that following Rawls’ ideal method will prevent understanding the nature of racial injustice, and hence it will lead to recommending a solution that does not adequately address the problem.

4.3: APPLYING RAWLS TO RACIAL INJUSTICE

Shelby sees no problem with applying Rawls' principles of justice, as they are, to correct for and prevent racial injustices. Mills, on the other hand, rejects Rawls' approach on the grounds that it ultimately detracts from efforts to understand race and racial injustice. He recommends doing away with ideal theory and focusing instead on non-ideal theory, which he defines in terms of non-idealized assumptions about human nature, accurate descriptive models of social phenomenon, and fact-sensitive normative principles (Mills, 2005, 174-176). Here I will defend Mills' position that ideal theory is essentially ideological and exclusionary.

It is important to note two distinct senses of "ideology", which are used in everyday language. Ideology is sometimes used merely as shorthand for any belief system that one might hold. This sense of ideology is derived from the strict meaning term as "the logic of ideas". This sense of ideology, however, is rather weak and could be applied to almost any theory or set of beliefs. But there is another sense, taken from the Marxist tradition of social theory, which uses the term ideology to mean a belief system that obscures reality in a way that makes structures of power seem rational or inevitable. For example, the capitalist economic system treats each person as essentially equal. Although this reflects a formal truth about societies in which capitalist markets exist, it obscures a substantive truth about the greater bargaining power of wealthier individuals in a class system. Thus, according to the Marxist critique of ideology, the study of non-Marxist economics functions as an ideology that obscures and allows for the exploitation of lower classes by obscuring their lesser position in the system. The more interesting claim to defend, then, is that ideal theory functions as an ideology in the Marxist sense. But is this claim true? In particular, can we show that it is true by showing that ideal theory obscures and misrepresents racial injustice?

In order to answer these questions, we must first look at an honest attempt at applying Rawlsian social justice principles to issues of racial injustice. According to Shelby's reading of Rawls, preventative measures against racial discrimination could be introduced into the basic structure of society during the "constitutional phase", where the veil of ignorance is partially lifted. This is the second phase of the four-part implementation of the principles of justice that Rawls envisions as an extension of the original position (Rawls, 1999, 46-49). In this phase, general facts about society become part of the knowledge that representatives in the original position have access to. Shelby thinks that this would have to include facts about race. Note here the grounds for stating that race should be considered from behind the veil are the same as with gender. Apparently, because general facts about society might entail certain facts about race or gender, these facts would have to be accessible to the deliberators in the original position. Shelby argues that if the representatives know facts about their society's "natural resources, level of economic advance [*sic*], political culture, and so on [including beliefs and interests of citizens]", then they would have to know about:

> "(1) whether racial identity engenders conflict in the society; (2) whether there are some in the society who have, or are prone to develop, racist beliefs or attitudes; and (3) whether some racial groups in society are or have been politically, socially, or economically disadvantaged" (Shelby, 2004, 1707).

Representatives would then be able to organize a constitutional regime that explicitly prohibits racial discrimination and distributes rights and duties equally, regardless of race.

Thus, according to Shelby, preventative measures would be put in place so that the basic structure of society does not systematically disadvantage any racial group.

Moreover, Shelby argues that Rawls' principle of fair equality of opportunity (FEO) could serve as a principle of corrective justice if it were ever implemented in a society with past racial injustices. In the United States, for example, the socio-economic condition of blacks relative to whites can be explained by a history of racial oppression, slavery, and so on. FEO would mitigate or correct for this type of inequality if it "were to be institutionally realized in a well-ordered society in which the basic liberties were secure and their fair value guaranteed" (Shelby, 2004, 1711). Now, clearly the US is not a well-ordered society in the Rawlsian sense. Shelby acknowledges that the principle of fair equality of opportunity would require considerable institutional adjustments in order to be realized in the United States. He accepts, as well, that he doesn't know what set of reforms would lead to these institutions. But if they were somehow realized in the US, he argues that the principle of fair equality of opportunity should entail a "considerable redistribution of wealth, the expansion of educational and employment opportunities, and aggressive measures to address discrimination in employment, housing, and lending" (Shelby, 2004, 1711). So, according to Shelby, Rawls' principles justify a whole set of possible institutional reforms, some of which would involve massive reparations (in the form of wealth transfers) to black folks in the United States.

Shelby's attempt at rescuing Rawls for racial injustice, however, misses the mark. For one thing, if the goal of Rawls' non-ideal theory is to generate principles which justify reforms that are supposed to show us the way to perfect justice, then one cannot simply assume that the correct set of institutional arrangements would somehow arise. Non-ideal theory ought to tell us which of the possible institutional reforms are justified by the general conception of justice as fairness. Hence, one must deal more directly with issues of feasibility within the context of a particular political system. Moreover, overlooking the fact that FEO was designed as an ideal principle for a well-ordered society confuses the type of wrong being addressed in reparations. In a response to Shelby's article, Charles Mills makes this point clear.

Mills (2013) begins by making a distinction between measures that prevent racial injustices from occurring in the basic structure of society and refractory measures that correct past wrongs. He thinks that racial justice is always going to require the latter kind of solution. For this reason, he focuses on Shelby's attempt to apply FEO to justify reparations. Mills argues that Shelby is committing a type of category mistake[13] when he tries to apply the principle to issues of racial injustice. He thinks that Shelby is confusing distributive and "rectificatory" justice. Mills gives four arguments that attempt to show that Shelby is confusing these two distinct kinds of justice. I am going to focus on two of them: the argument from Rawls' focus on ideal theory and the argument from Rawls' lexical ordering of principles.

First, Mills argues that if Rawls' conception of ideal theory is focused on principles for a well-ordered society, and a society that has a history of racial oppression cannot be well ordered, then ideal theory will not tell us what to do to correct for these past wrongs. One reason that a society with a history of racial oppression cannot be well-ordered is that there still exists racial injustice in many of those societies. Moreover, it is difficult to see

[13] Perhaps "category mistake" isn't the correct phrase insofar as different types of justice don't really amount to independent logical categories. The critique might be stated as an unwarranted or misguided use of FEO to justify reparations.

how races- as we know them- would have come into existence at all in a society with no history of racial oppression: “As the huge and ever-growing body of literature over the last decade in critical race theory and critical white studies demonstrates, race is socially constructed, and without systematic discrimination, race would not even have come into existence” (Mills, 2009, 179). In a well-ordered society with no past racial injustice, and perhaps even no races (that point is more contentious), FEO would only apply to the unfair distributions of social positions that result from the “natural lottery” given at birth.

Given the fact that many disadvantaged neighborhoods are disproportionately inhabited by non-white races, measures to ensure fair equality of opportunity in those neighborhoods would surely counteract some racial inequalities. But reparations are not only meant to address unfair distributions in access to opportunities and public services. The idea is to “repair” a past wrong, not simply to correct for contingencies in the distribution of opportunities that would result in any society, even in a well-ordered one. That is why Rawls meant to include FEO as part of the difference principle.

According to Mills, then, applying FEO to racial inequalities that are a result of past oppression and violence confuses rectificatory (*sic)* justice with distributive justice. Mills further argues that if Rawls had intended his principles to be used this way, he would not have thought it necessary to split the theory of justice into two parts. It is clear from reading Shelby, however, that he is well aware of the departures he makes from Rawls’ view. So, it would be somewhat pointless to accuse Shelby of misusing Rawls’ normative system unless doing so was somehow bad for other reasons. That seems to be Mills’ view, as we will see in the next section. Mills’ argument against Shelby is primarily grounded in a view about the limitations of ideal theory. The claims he makes about the metaphysics of race, though interesting, are less important to his critique of Shelby than they appear. Further, they are outside the scope of this paper to discuss. The second argument I want to discuss is more centrally focused on the limitations of ideal theory.

The argument begins with the claim that Rawls' lexical ordering of principles places restrictions on when and how they can be applied. Basic liberties are lexically prior to fair equality of opportunity, which means that violations of basic liberties will have to be dealt with first. In other words, Rawls' lexical ordering of principles implies what Mills calls "deontological constraints" for any proper Rawlsian non-ideal theory. These constraints include: "(i) consistency with the 'general' conception of justice…(ii) the 'reflect[ion of] priority relations of ideal theory' in the attempt to bring about ideal conditions… and (iii) 'consisten[cy] with the [deontological] spirit of the ideal theory'" (Mills, 2013, 15). Mills thinks that Shelby disregards (ii) and (iii) because he tries to apply the principle of fair equality of opportunity to correct for wrongs that are more properly seen as violations of basic liberties. Shelby tries to justify reparations- or a massive transfer of wealth from whites to blacks- on the grounds that historical injustices have created class divisions along racial lines. Though this may be true, it is not enough to justify reparations. If a redistribution of wealth from whites to blacks was to be justified, it could not merely be on the grounds that there is an unequal distribution. It is only because this unequal distribution was a result of violence and oppression-which is primarily a violation of basic "negative' liberties- that reparations make sense. Otherwise any redistribution of wealth (not necessarily along racial lines) would be enough to solve the problem.

Mills states that racial oppression and discrimination "is a violation of negative rights", which consist in "noninterference with life, liberty, and property" as well as a violation what he calls “weak egalitarianism”, which only recognizes “moral, legal, and political equality” as legitimate norms (Mills, 2013, 19). Now, there may be a much more complex story to be told about the origins of black socioeconomic disadvantage and what’s wrong with it. But I take Mills’ point to be that by applying FEO to justify reparations amount to a kind of category mistake because “Shelby is blurring the difference between wrongs that involve violation of (left-liberal) norms of opportunity and wrongs that involve the violation of personhood” (Mills, 2013, 19). Insofar as we accept that racial class disadvantage in the United States was primarily a result of exploitation and racial violence, as opposed to resulting from the “natural lottery”, we should be able to accept the conclusion that Shelby is misapplying FEO.

Mills writes: “[Shelby] is using FEO *as if* it were a principle of rectificatory justice authorizing wealth transfer. But so far as I can see, he is not entitled to do this, because such an extrapolation goes far beyond what Rawls himself intended” (Mills, 2013, 15). Since Shelby’s use of fair equality of opportunity does not reflect Rawls’ priority rules, it isn’t going to work as a Rawlsian non-ideal principle.

The disagreement between Mills and Shelby hinges on two different interpretations of ideal theory and its function. For Shelby (and Rawls) ideal theory is a conception of the best possible society. It gives us a blueprint for identifying deviations from perfect justice, and thus for recommending reforms that would take us closer to the ideal. Mills, on the other hand, views ideal theory as a tacitly designed empirical model (Mills, 2005). He thinks that anyone who has a picture of an ideal society must also have a pre-formed empirical model of the way their society *actually* looks. Mills calls this “ideal theory as idealized model” (Mills, 2005, 167), which is a picture of an actual society with all the imperfections abstracted away. Shelby sees ideal theory as an indispensable tool for understanding injustice, whereas Mills sees ideal theory as a kind of injustice in itself.

Recall from section 2.2 the discussion of idealization versus abstraction in the initial factual assumptions of a normative theory. Idealization was taken to mean falsification in the theory's conception of the human subject or of society. Abstraction, on the other hand, was thought to consist in simplification of social and psychological phenomenon without falsification. Mills' view seems to be that ideal theory represents an ethically noxious form of idealization. An ideal social ontology is presupposed that underrepresents marginalized groups and thereby perpetuates their disadvantage. Thus, ideal theory is "really an ideology, a distortional complex of ideas, values, norms, and beliefs that reflect the nonrepresentative interests and experiences of a small minority of the national population" (Mills, 2005, 172).

Now, clearly this is a very strong claim to make. Does Mills' argument make sense? If we accept Mills' distinction between ideal-as-model and ideal-as-idealized-model, then it would seem that ideal theory could be viewed as an extrapolation on empirical assumptions. In order to extrapolate to a normative ideal from a descriptive model, then, one must make use of an idealized social ontology that will depart dramatically from one's initial descriptive ontology. Does it follow from this that ideal theory is ideological? Shelby (2013) has responded to Mills by saying that even if we accept all of Mills' claims about ideal theory, it does not follow that it is thereby ideological in the sense of a belief system that perpetuates group privilege. He states: "In particular, they do not establish that ideal theory necessarily obscures or misrepresents racial injustice, conceals the need for rectificatory justice, or perpetuates the racial status quo." (Shelby, 2013, 153).

In order for the inference to follow, I contend, there would first of all have to be relations of oppression already in place that are systematically ignored by most or all ideal theories. This seems plausible and is emphasized by the complaints that ideal theories tend to ignore important injustices. But just because there is oppression and a theory doesn't talk about it, that doesn't mean the theory will necessarily contribute to it. There would also have to be a premise in the argument about the connection between silence in academic discourses and its effect on relations of oppression. How might silence contribute to relations of oppression? One way to answer that question would be to look at the ways in which silence can be advantageous to some but not others. Further, one might consider how epistemic norms surrounding the transfer of knowledge can, in effect, "silence" oppressed groups by systematically excluding them from attaining status as "knowers". In the next section, I will explore some of these possibilities.

Mills, however, would respond that the connection is contained in the definition of ideology as a belief system which obscures reality and thereby perpetuates group privilege. The best way to understand how ideology perpetuates systems of oppression is through "social privilege and resulting differential experience" (Mills, 2005, 172). The privileged group takes their particular experience to be constitutive of the real world. Mills observes that political philosophy as represented by mostly upper to middle class white men (Mills, 2005, 172). This over-representation of a privileged group within academic circles plus the absence of countervailing views leads to a limited perspective of reality. In particular, the privileged group's inability to recognize their place in oppressive social structures leads to the continuation of those structures because they will be viewed as natural or inevitable. Presumably, then, ideology in this sense would have to include any structural bias in the structures of knowledge production that privileges some groups and disadvantages others. It is important to note that this is a considerably expanded notion of ideology that is somewhat detached from the pejorative connotations that are typically associated with it.

4.4 IDEAL THEORY, IDEOLOGY, AND HERMENEUTIC INJUSTICE

As a way of supporting the argument for ideal theory as a type of ideology, I will now explore a possible link between theories of social epistemology and theories of social justice by referring to Miranda Fricker's concept of epistemic injustice. I argue that an analysis of epistemic injustice, in particular hermeneutic injustice and its implications, can provide the missing premise for Mills' argument.

Epistemic injustice occurs when a person is harmed in her capacity as a knower (Fricker, 2007, 1-3). There are two types: testimonial and hermeneutic injustice. The first type occurs when an undue credibility deficit is afforded to a speaker because of an identity prejudice on the part of the hearer (Fricker, 2007, 17-28). Identity prejudice is explained in part by the influence of distorting stereotypes on the hearer's credibility judgments. One example Fricker uses is a white police officer not believing the testimony of a black person due to the influence of a negative stereotypes (Fricker, 2007, 4). Fricker views stereotypes as simplifying heuristics that allow us to navigate the social world without stopping to rationally assess every interaction we have. In general, she states they are "widely held associations between a given social group and one or more attributes" (Fricker, 2007, 30). They can be mostly empirically reliable, like the stereotype of the "dependable family doctor" (Fricker, 2007, 32), or they can be unreliable. Some unreliable stereotypes involve a kind of ethically noxious prejudice. Prejudice, in the sense Fricker employs, is a kind of pre-judgment: "it is most naturally interpreted in an internalist vein as a judgment made or maintained without proper regard to the evidence" (Fricker, 2007, 33). More accurately, however, Fricker thinks that stereotypes are like images that display the characteristics of social types. Stereotypical images work at the level of the collective imagination, as opposed to directly influencing rational deliberation (Fricker, 2007, 36). Identity prejudice, then, will be mediated in part by an individual's false beliefs about social groups but also by stereotypes at work in the collective 'social imagination'.

When ethically noxious prejudice impacts a hearer's attribution of credibility to a speaker such that the speaker is no longer able to convey information and participate in testimonial exchanges, the speaker is then wronged in her capacity as a knower. The primary harm of testimonial injustice is conceived of as a kind of objectification, where a subject or agent is undermined as a "giver of knowledge" and relegated to a "source" of information (Fricker, 2007, 132-133).

The second type of epistemic injustice is caused by a gap or 'lacuna' in the shared hermeneutic (interpretive) resources of a community. Fricker uses the example of women's experience of sexual harassment prior to there being a name for that phenomenon (Fricker, 2007, 150). The lack of a name for something might not seem like a problem by itself. For instance, it is quite likely that women were well aware that sexual harassment was a problem before there was a word for it. But when the lacuna is viewed from a structural perspective, there is an obvious "asymmetrical disadvantage", which can allow for the wrong of sexual harassment to go un-checked:

> "In the present example, harasser and harassee alike are cognitively handicapped by the hermeneutical lacuna- neither has a proper understanding of how he is treating her- but the harasser's cognitive disablement is not a significant disadvantage to him. Indeed, there is an

> obvious sense in which it suits his purpose…By contrast, the harassee's cognitive disablement is seriously disadvantageous to her…Her hermeneutical disadvantage renders her unable to make sense of her ongoing mistreatment, and this in turn prevents her from protesting it" (Fricker, 2007, 151).

There is a deeper sense, however, in which hermeneutic injustice is a kind of structural injustice. Fricker explains that the primary harm of hermeneutic injustice is exclusion from sharing knowledge, particularly knowledge of one's own experiences, due to a lack of interpretive resources in a community. Fricker calls this process of exclusion hermeneutic marginalization.

To be marginalized in this way usually involves being excluded from the types of professions that are epistemically esteemed, such as "journalism, politics, academia, and law" (Fricker, 2007, 152). Oppressed groups have historically been excluded from these types of jobs, which seems to explain the gap in hermeneutic resources that would be relevant to interpreting their experiences. Thus, hermeneutic marginalization, which involves an inequality in the sharing of knowledge, usually has an antecedent material or social inequality. In other words, historic oppression results in hermeneutic lacunas, which then unjustly disadvantage oppressed groups in their status as knowers. But the causal dependence between social inequality and epistemic inequality also seems to work in the other direction, creating a kind of vicious cycle. When there is hermeneutic marginalization, interpretive resources will be "structurally prejudiced" because they favor the experiences of dominant groups. But that will lead to the continued material and social inequality of oppressed groups because they will continue to be denied access to privileged professions. Thus, not only does social inequality lead to unjust hermeneutic lacunas, but also: "hermeneutic marginalization *entails* marginalization of a socio-economic sort" (Fricker, 2007, 155). Fricker is not talking about strict logical entailment, of course, but rather a kind of causal dependence or correlation.

Fricker's view of epistemic injustice, I think, provides some grounds for thinking that there are ways in which academic silence can contribute to structures of oppression. If gaps in the shared interpretive resources of a community can result in socio-economic inequalities, then since academic communities are primarily responsible for the maintenance of these resources, they are also responsible, in some way, for the continued existence of oppressive social structures. One might respond that, as a matter of fact, academics are less responsible for the maintenance of interpretive resources than say, politicians, schoolteachers, and journalists. This might be so, however, the abilities of academics to influence these professions in various ways is considerable if not actually quite high. After all, teachers and politicians have to be educated themselves. So, I view the inference as warranted by the way the academic world is currently set up. Therefore, the critique of ideal theory as a kind of ideology needs to be explored further.

In his self-proclaimed work of non-ideal theory, *The Racial Contract,* Mills speaks of epistemologies of ignorance, which are proscribed by the Racial Contract (Mills, 1997, 18-19; 96-101). Mills' view is that the Racial Contract, a primarily descriptive (non-idealized) model of global white-supremacist political systems, better explains what is wrong with the world and how to fix it than traditional contract theory approaches. He explains that the Racial Contract involves a set of epistemological norms that work to promote ignorance and misrecognition of racially structured political systems. The epistemic conditions of the Racial Contract include:

(a) Concepts that legitimize the racial order will be favored over ones that call into question the superiority of whites;

(b) Later on, concepts that "*derace* the polity" will be promoted in order to misrepresent its actual racial structuring;

(c) Measures of well-being will be dependent, in some way, on comparative judgments between whites and non-whites, since "the essence of whiteness is entitlement to differential privilege vis-a-vis nonwhites" (Mills, 1997, 95);

(d) On a whole, whites will develop patterns of affect and empathy that are only weakly, if at all, influenced by non-white suffering (Mills, 1997, 95).

The general acceptance of these epistemic norms results in whites being on the whole unable to recognize the political system that affords them differential privilege on the basis of their race. Further, a general social ontology of white humans and non-white sub-humans is presupposed or unconsciously accepted in order to lessen the cognitive dissonance experienced by whites who accept the terms of the Racial Contract. Ideal theory can be seen as a symptom of white epistemologies of ignorance because it (i) disables the recognition of racially structured political systems, (ii) obscures the need for rectification of past wrongs, and (iii) contributes to hermeneutic lacunas that disadvantage oppressed groups both materially and in their status as conveyors of knowledge.

In a sense, epistemologies of ignorance buttress or support hermeneutic lacunas, which disadvantage oppressed groups in their status as knowers. Those within oppressed or marginalized groups begin to speak another language, as it were, and fail to get their message across to dominant groups because of structural prejudices working against them. If hermeneutic lacunas disadvantage groups in their status as conveyers of knowledge, then we cannot treat all agents who claim to have knowledge in the same way. There could be a structural prejudice that supports one group's claims to knowledge and discounts others', leaving behind gaps in the collective resources within a knowledge community. The production and transfer of knowledge itself, then, ought to be subject to evaluations from the perspective of justice since these processes result in social inequalities. This point calls into question Rawls' method, which treats general knowledge of political systems as basically morally neutral. Since, not every knowledge claim should be treated as equal, especially when a would-be knower is asymmetrically positioned with respect to some hermeneutic lacuna, Rawls' neat division between general and particular facts is actually more problematic than it initially appears.

One might respond, however, that these implications are not essential to any and all ideal theories, but rather that they are only contingent aspects of some particular theories. For instance, Rawls' theory might succumb to one or more of these criticisms while other ideal theories might not. This objection seems inadequate because Rawls' method is rather influential and is considered by many as the paradigm example of an ideal theory of justice. Moreover, even if the objection is true in the sense that the argument against Rawls does not establish that any and all attempts at ideal theory are necessarily exclusionary in a morally objectionable way, the critique strongly suggests that we ought to find another way to do normative political theory. In other words, it may be true that not every ideal theory is, in essence, ideological. But insofar as theories of justice should really be critical of the status quo- and being critical requires being able to recognize forms of injustice which might not be apparent to someone who sits at the top of a social hierarchy- theories of justice should be non-ideal theories.

Another objection might state that the critique of ideal theory as an ideology presupposes intent on the part of white academics to maintain the racial status quo. But it is false that the exclusionary aspects of ideal theory must exist because of the intentions of theorists like John Rawls. In other words, the argument that ideal theory functions as an ideology does not depend on the assumption that it was constructed with the intent that it should perform that function. Rather, when we begin from the idea of a situated knower- an epistemic agent who is part of a larger community with a set of norms and rules governing the transfer and production of knowledge- then we can arrive at a conception of how structural bias in the knowledge community can result in winners and losers without the need for intentional prejudice. Fricker's view of hermeneutic injustice provides an institutional or structural perspective from which to view the function of ideal theory in academic discourse that does not involve invoking the malicious intentions of academics.

4.5 CONCLUSION

At the end of the preceding section, I suggested that the arguments presented here would also implicate attempts to include gender- and by extension other ascriptive identity categories- into the framework of ideal theory. This point rested on two questions: (1) "should we consider gendered divisions of labour as general facts accessible to ideal epistemic agents in the original position?" and (2) "can ideal theory adequately formulate normative principles for issues like gender?".

We have seen that an appeal to the general nature of facts is less reliable than previously thought. An agents' relative position in being able to convey knowledge differs dramatically when there are structural prejudices in the collective interpretive resources that favors their group over others. Hence, we cannot treat all knowledge claims in the same way, even if they are of a general sort. This would mean that we need to consider further aspects about epistemic agents- in particular their race or gender- as well as a whole host of other "situated" relational qualities. But once we accept the notion of a situated knower, the simplistic division between "general" and "particular" social facts flies out the window. Moreover, calling gendered divisions of labour or histories of racial oppression "general facts" about society runs the risk of normalizing the status quo, which I believe almost everyone can agree is not what we a want a theory of justice to do, at least not in the societies in which we currently live.

Further, we saw that attempts to apply Rawls' ideal theoretical principles to non-ideal cases, in particular those involving ascriptive identity categories, will miss the mark because they will fail to fully capture what is wrong them. For instance, racial injustice is not only about fair equality of opportunity but also about rectification for past wrongs. Failure to see it as such results in the misrepresentation of racial injustice as a problem that can be dealt with using "color-blind" or race-neutral policies.

Rawls' attempt to include gendered divisions of labour from behind the veil, on the condition that they must be freely chosen and that measures be in place to mitigate inequalities which result from them, seems to be a similarly misguided solution. For instance, the concept of a "free choice" is itself laden with masculine norms that don't apply equally to all agents. Moreover, putting measures in place to mitigate inequalities that result from gendered divisions of labour is like putting a bandage on a problem rather than finding a real solution.

CHAPTER 5: CONCLUSION

5.1: ON JUSTICE AND INJUSTICE

The preceding four chapters can be roughly divided into two separate discussions relating to ideal and non-ideal theories of justice, dealing with methodology and then with application respectively. The first and most lengthy part examined the ideal/non-ideal distinction by explaining how theorists in fact tend to draw that distinction. I concluded the first part by arguing for how I think we should draw it according to broad teleological factors. I proposed that we ought to look at the overall purpose of a theorist, as expressed in their theoretical starting point, their normative goals, and their particular orientation towards their subject matter. In short, I presented a simplified view of Sen's distinction- between comparative and transcendental theories- which referred to the purposes and goals of a theorist, as opposed to more "concrete" or internal aspects of their normative method and/or theory. Although these more objective characteristics are important to determining what the overall purpose of a theory of justice is, I argued that they only tend to show that things like purposes and goals are primary. For instance, the way that a theory includes empirical data into its normative framework can signify what the starting point of a theory is. How a theory of justice conceives of societies and citizens- i.e. from an abstract individualistic point of view or from a situated relational point of view- as well, can indicate a king of normative starting point. Moreover, whether a theory of justice is responsive to issues of feasibility, or whether it intends to compare different feasible social structures, can indicate a theorist's overall normative orientation or purpose. Hence, I think that the distinction between ideal theories and non-ideal theories of justice is primarily one about the purposes and goals of a theorist in tackling issues of social justice.

But what does this view amount to? If one views the ideal/non-ideal theory distinction in this way, then theories that try to give an account of Justice are ideal and those that give an account of injustice and its related problems are non-ideal. But does this even make sense, and if so, what might the problems be for holding a view like this? One problem my view encounters is that when we suppose that there is no necessary connection between ideal and non-ideal theory- as Sen maintains and I agree- while also claiming that non-ideal theories are theories of *in*justice, then we must conclude that injustice can be studied apart from justice. But insofar as injustice is merely a negation of justice, this hardly seems to make sense. How could we even begin to identify injustice without a presupposed ideal theory of justice working in the background?

Simmons' argument from path-dependence, which I discussed in section 3.2, seems to state a similar worry. Simmons claimed that Sen's view, which is that comparative and transcendental theories of justice are analytically disjoined, is incorrect if we take him to mean that knowledge of how to advance justice does not require knowledge of justice itself. Similarly, Rawls' claim that ideal theory provides the only systematic way to grasp the problems of non-compliance that exist in everyday societies is a reminder that justice and injustice are interdependent concepts. Someone with a particular conception of justice will see some social arrangements as just and others as unjust. It is less clear, however, whether someone who lacks a conception of justice (in the transcendental sense) would be able to pick out as many or perhaps even more individual situations that we could all agree are unjust. That thought leads one to the possibility that justice is actually a negation of injustice and that whatever social arrangements most closely eliminate injustice are just social arrangements. That is not the same as saying that whatever social arrangement instantiates a principle, or which has the right outcome measured in terms of well-being, is a just social arrangement. One might accuse me here of abusing language. But the concept of justice is constructed out of the need to remedy problems and to resolve conflicts. It does not track any natural kinds. Hence, its very nature is dependent on the types of problems that exist in society.

Moreover, when we consider that having knowledge of justice in the abstract- i.e. justice itself- requires having knowledge of the relevant justice-bearing features of a (possibly endless) set of social worlds, we are forced to conclude that it is impossible to give an account of justice by describing a fully just society. That is because such an account requires making every justice-relevant feature of that society intelligible to someone else, and this might not be possible. That is the essence of Gaus and Hankins "diversity dilemma", also discussed in section 3.2, which stated that knowledge of far-flung social worlds is restricted by our ability to identify and evaluate features of our own social world. This is problematic because perspectives that claim to have knowledge of the "global optimal" social world can be fundamentally incompatible with perspectives that are useful in identifying and ranking various local optima.

But the difficulty with giving an account of a perfect society that would thereby make justice- and the path towards it- intelligible to others is only a problem for someone whose purpose is to give an account of Justice by listing its qualities. If it turns out that justice in fact has no qualities or positive characteristics, but rather is only a negative concept that refers to a lack of qualities or characteristics[14], then the account I have given

[14] For a perspective that considers the possibility of justice being a "negative concept", see David Schmidtz's discussion of comparative justice in Schmidtz, David. (2015). "Nonideal Theory: What It Is and What It Needs to Be". *Ethics*. 121:4, 772-796.

of the distinction between ideal and non-ideal theories would make perfect sense. The only problem would be that the overwhelming focus on trying to understand Justice has been almost entirely in vain. The correct method would instead begin with an account of the problems posed by injustices in the world, and then by referring to their causal influences, we can recommend a solution or a remedy.

A key advantage to this view is that it remains silent on the actual nature of justice and injustice. If it turned out that I am wrong, and justice could be understood by a set of qualities or characteristics that would be instantiated by an ideal society, then the study of Justice itself would still be a proper part of ideal theory, and the study of injustice would then be dependent on it. But if not, then there is room to say that there are many forms of injustice, each requiring its own non-ideal theory of justice, and the collective elimination of each would be what justice requires.

The second part of this thesis, concerned more with application than with methodology, discussed some negative consequences that could be attributed to attempts at applying Rawlsian ideal theory to difficult cases. The cases I chose to discuss were Susan Okin's argument for including gendered divisions of labour in the framework of ideal theory and Tommie Shelby's attempts to apply Rawlsian social justice principles to issues of racial injustice. Both attempts were taken to be ideological in the sense that certain important features of gender and racial injustice were left out of the picture in a way that seems morally problematic. For instance, including general facts about gender and gendered divisions of labour from behind the veil of ignorance seems to make them appear natural or inevitable, which prevents us from being able to critically assess them in the first place. Further, Shelby's attempt at applying FEO to justify reparations overlooked a key feature of racial injustice- the historical wrongs of slavery- in a way that risked misdiagnosing the problem and misrepresenting the solution.

I then presented an expanded version of Mills' critique of ideal theory as a form of ideology by using Miranda Fricker's account of hermeneutic injustice to show that ideal theorizing involves a pernicious form of academic ignorance. In short, I argued that Mills' argument is missing a premise that can be provided by an analysis of hermeneutic lacunas in academic discourse. I think that the claim that ideal theory is really an ideology, if interpreted correctly, provides enough reason to discount Rawls' method and to motivate the search for a new method for thinking about justice.

5.2: IMPLICATIONS OF THE ARGUMENT

As a way to conclude the arguments presented here, I would like to briefly discuss some possible avenues of future research. If what I have said here is correct, then there may be a fruitful non-ideal method that works backwards from problems of injustice to normative principles as solutions. But any such project would have to adequately address the issues raised by debates surrounding fact-sensitivity, idealization, feasibility, and comparative justice. One such issue comes from the need for models of social phenomenon that allow us to better understand and analyze injustices in the world. The problem is that when we ask things like "how much fact-sensitivity is adequate?" or "how much idealization (or what kind) is a bad thing?", there seems to be little consensus about what amounts to a good answer. Hence, there should be more collaboration between social theorists and philosophers interested in mapping injustices so that our non-ideal normative principles actually respond to the problems at hand.

Second, the relationship between social epistemology and theories of justice needs to be explored more fully. In the last chapter, I suggested that once we begin with the concept of a situated knower- a more promising non-ideal epistemological concept than the abstract rational individual- our principles of justice will have to respond to inequalities in the structures of knowledge production and transfer. Since viewing epistemic agents as situated in a community of norms and as being engaged in a kind of economy of knowledge transfers is to place them in positions of advantage and disadvantage relative to each other, theories of social justice must explore the consequences of these inequalities. Theories of epistemic injustice, in my opinion, do not go far enough in explaining the material and social consequences of knowledge production and transfer. Neither do they go far enough in developing the intuition that "knowledge is power" into an intelligible view.

Finally, a major question that I think confronts anyone interested in developing a non-ideal theory of justice is "how can we identify injustice if we do not have (or want to have) a comprehensive theory of justice?". One way to answer this question is to explain the recognition of injustice in terms of non-inferential judgments. From the perspective of virtue epistemology, which takes a page from Aristotles account of ethical sensibility, one might argue that ethical judgments are, at base, non-reflexive, non-inferential, and to some extent, spontaneous. That is to say that we are just somehow "hard-wired" to make ethical evaluations and that our judgments are a result of a complex arrangement of factors, including upbringing, education, cognitive biases, and some kind of "ethically laden" form of perception (Fricker, 2007, 84). Hence, the failure to identify injustice seems more a result of cognitive bias rather than having the wrong theory of justice. A non-ideal theory of justice, then, should help us to see through bias to more accurately perceive injustice.

BIBLIOGRAPHY

Anderson, Elizabeth. (1999). “What’s The Point of Equality?”. *Ethics.* 109:2 287-337

___. (2012). “Epistemic Justice as a Virtue of Social Institutions”. *Social Epistemology.* 20:2, 163-173.

___. (2010). *The Imperative of Integration.* Princeton: Princeton University Press.

Boot, Martin. (2010). “The Aim of a Theory of Justice”. *Ethical Theory and Moral Practice.* 15:1, 7-21.

Cohen, G.A. (2008). *Rescuing Justice and Equality.* Cambridge: Harvard University Press.

___. (2003). “Facts and Principles”. *Philosophy & Public Affairs.* 31:3, 211-245.

Estlund, David. (2017). “Prime Justice” in *Political Utopias: Contemporary Debates.* Ed. Michael Weber and Kevin Vallier. New York: Oxford University Press.

___. (2014). “Utophobia”. *Philosophy & Public Affairs.* 42:2, 113-134.

___. (2011). “Human Nature and the Limits of Political Philosophy”. *Philosophy and Public Affairs.* 39:3, 207-237

Fuller, Lisa. (2012). “Burdened Societies and Transitional Justice”. *Ethical Theory and Moral Practice.* 15:3, 369-386.

Fricker, Miranda. (2007). *Epistemic Injustice: Power and the Ethics of Knowing.* New York, Oxford University Press.

___. (2003). “Epistemic Injustice and a Role for Virtue in the Politics of Knowing”. *Metaphilosophy.* 34:1/2, 154-173.

Gaus, Gerald and Hankins, Keith. (2017). “Searching for the Ideal: The Fundamental Diversity Dilemma”. In *Political Utopias: Contemporary Debates.* Ed. Michael Weber and Kevin Vallier. New York: Oxford University Press.

Gilabert, Paulo. (2012). “Assessments of Justice, Political Feasibility, and Ideal Theory”. *Ethical Theory and Moral Practice.* 15:1, 39-56.

___. (2017). “Justice and Feasibility: A Dynamic Approach” in *Political Utopias: Contemporary Debates.* Ed. Michael Weber and Kevin Vallier. New York: Oxford University Press.

Gilabert, Paulo and Lawford-Smith, Holly. (2012). “Political Feasibility: A Conceptual Exploration”. *Political Studies.* Vol. 60, 809-825.

Gurrero, Alexander. (2017). “Political Functionalism and the Importance of Social Facts” in *Political Utopias: Contemporary Debates.* Ed. Michael Weber and Kevin Vallier. New York: Oxford University Press.

Hamlin, Alan and Stemplowska, Zofia. (2012). “Theory, Ideal Theory, and the Theory of Ideals”. *Political Studies Review.* Vol 10, 48-62.

Knight, Jack. (2014). “The Imperative of Non-Ideal Theory”. *Political Studies Review.* 12, 361–368

McCarthy, Thomas. (2009). *Race, Empire, and the Ideal of Human Development.* Cambridge: Cambridge University Press.

___. (2007). “The Natural Order of Things: Social Darwinism and White Supremacy”. *Contemporary Pragmatism. 4:1, 7-24.*

Mills, Charles. (2015). “Decolonizing Western Political Philosophy”. *New Political Science*. 37:1, 1-24

___. (2009). “Rawls on Race/Race in Rawls”. *The Southern Journal of Philosophy.* Vol XLVII.

___. (2005). “Ideal Theory as Ideology”. *Hypatia*. 20:3, 165-184.

Okin, Susan, Moller. (1989). “Reason and Feeling in Thinking about Justice”. *Ethics.* 99:2, 229-249.

___. (1994). “Gender Inequality and Cultural Differences”. *Political Theory.* 22:1, 5-24.

O’Neill, Onora. (1987) “Abstraction, idealization, and Ideology in Ethics” in *Moral Philosophy and Contemporary Problems.* Ed. Evans, J. D. G., Cambridge University Press, New York.

Rawls, John. (1971). *A Theory of Justice.* Cambridge: Harvard University Press.

___. (1999). *Justice as Fairness: A Restatement.* Cambridge: Harvard University Press.

___.(1999). *The Law of Peoples*. Cambridge: Harvard University Press

Roberto Frega. (2014). “Between Pragmatism and Critical Theory: Social Philosophy Today”. *Human Studies.* 37:1, 57-82

Robeyns, Ingrid. (2008). “Ideal Theory in Theory and Practice” *Social Theory and Practice.* 34:3, 341-362.

Schmidtz, David. (2015). “Nonideal Theory: What It Is and What It Needs to Be”. *Ethics*. 121:4, 772-796.

Schwartsman, Lisa. (2006). *Challenging Liberalism: Feminism as Political Critique*. Pennsylvania: Pennsylvania State University Press.

___. (2009). "Non-Ideal Theorizing, Social Groups, and Knowledge of Oppression: A Response". *Hypatia.* 24:4, 177-188.

___. (2006). "Abstraction, Idealization, and Oppression". *Metaphilosophy.* 37: 5, 565-588

Sen, Amartya. (2006). "What Do We Want From a Theory of Justice?". *Journal of Philosophy*. 103:5 215-238.

___. (2009). *The Idea of Justice.* Cambridge: Harvard University Press

Shelby, Tommie. (2004). Race and Social Justice: Rawlsian Considerations. *Fordham Law Review.* 72:5, 1697-1714

___. (2013). Racial Realities and Corrective Justice: A Reply to Charles Mills. *Critical Philosophy of Race.* 1:2, 145-162.

Sherman, Benjamin, R. (2016). "There's No (Testimonial) Justice: Why Pursuit of a Virtue is Not the Solution to Epistemic Injustice". *Social Epistemology.* Vol. 30, No. 3, 229-250.

Shklar, Judith. (1989). "Giving Injustice Its Due". *The Yale Law Journal.* 98:6, 1135-1151.

___. (1991). "The Faces of Injustice". *Law and Philosophy.* 10:4, 433-446

___. (1965). "The Political Theory of Utopia: From Melancholy to Nostalgia" *Daedalus.* 94:2, 367-381

Simmons, John. (2010). "Ideal and Nonideal Theory" *Philosophy and Public Affairs.* 38:1, 5-36.

Talisse. Robert, B. (2017). "Can Nonideal Theories of Justice Guide Action?" in *Political Utopias: Contemporary Debates.* Ed. Michael Weber and Kevin Vallier. New York: Oxford University Press.

Valentini, Laura. (2017) "On the Messy 'Utophobia vs. Factophobia Controversy: A Systematization and Assessment". In *Political Utopias: Contemporary Debates.* Ed. Michael Weber and Kevin Vallier. New York: Oxford University Press.

___. (2012). "Ideal Vs. Non-Ideal Theory: A Conceptual Map". *Philosophy Compass*. Vol. 7, No. 9, 654–664.

Volacu, Alexandru. (2018). “Bridging Ideal and Non-Ideal Theory”. *Political Studies*. Vol. 66, No. 4, 887 –902

Weins, David. (2017). “Will the Real Principles of Justice Please Stand Up” in *Political Utopias: Contemporary Debates.* Ed. Michael Weber and Kevin Vallier. New York: Oxford University Press.

___. (2015). “Against Ideal Guidance”. *The Journal of Politics.* 77:2, 433-446.