

**PHD THESIS SUMMARY:**

**An incrementalist approach to political philosophy. The case of heterogeneous rationality assumptions in theories of distributive justice**

ALEXANDRU VOLACU

*PhD in political science, July 2016*

*National University of Political Science and Public Administration*

The main goal of the thesis is to provide a framework for the pursuit of normative inquiries in the field of political philosophy, which bridges ideal and non-ideal theory by accommodating both types of approaches and by providing a blueprint for the transition between ideal and non-ideal theory. This blueprint is subsequently used to examine the various ways in which principles of distributive justice are dependent upon certain idealised assumptions and the ways in which such principles would be affected by the transition from ideal to non-ideal theory. A secondary goal of the thesis is to explore to what extent we can draw on various developments from social sciences (and particularly for this thesis, from neoclassical and behavioural economics) to engage in meaningful analyses of theories belonging to the field of political philosophy (and particularly for this thesis, theories of distributive justice). Aside from the provision of a theoretical background concerning the topic of distributive justice (chapter 2), the thesis consists of two main parts.

In the first part (chapter 3), I propose and defend a novel methodological approach for pursuing normative inquiries, which aims to bridge ideal and non-ideal theory and provide a framework for the systematization of knowledge in political philosophy. The core ideas of the approach are that: (1) political philosophy should be concerned with both short term goals, such as addressing severe injustices in the present world and long term goals, such as moving towards gradually more just states of the world in the long-run and (2) an appropriate way to satisfy these aims in a systematic manner is to build a landscape of normative models through incremental derivation (i.e., making the assumptions of a model more or less fact-sensitive one at a time) and selecting to pursue the best path that we might take from mitigating problems of severe

injustice in the current state of the world to the achievement of perfect justice. The subsequent chapters, which together compose the second part of the thesis, are then designed to expose various ways in which normative principles are dependent on certain assumptions, thereby illustrating the problems which may arise in the transition from ideal to non-ideal theory. All cases studied refer to the connection between principles of distributive justice and assumptions concerning rationality.

In chapter 4 I challenge the Rawlsian assertion that each party will have a first-ranked preference for an identical set of principles of justice behind the veil of ignorance, claiming that the original position allows parties to choose on the basis of a plurality of conceptions of rationality and allows choices based on the assignment of different weights to social positions, which in turn may lead to a reasonable disagreement concerning the conception of justice selected. I then argue that this reasonable disagreement does not entail an abandonment of the contractualist project, but its reconstruction in the form of a two-stage process, where parties first construct an individual preference hierarchy for alternative conceptions of justice and then work towards the reconciliation of the divergent conceptions that are chosen in the first stage. Finally, I claim that the threshold prioritarian view is the most plausible candidate for selection in this reconciliatory stage, since it manages to address both the legitimate complaints of parties that would prefer a conception of justice focused on the most disadvantaged positions in society as well as the legitimate complaints of parties that would prefer a conception of justice in which less or no special weight is assigned to the worst-off positions. This result illustrates the more general claim that, in some cases, incrementally modifying the inputs of a normative model may lead to a change in the principles generated by that model.

In chapter 5 I seek to assess the responses provided by several theories of sufficientarian justice in cases where heterogeneous assumptions concerning individual rationality are introduced. In the first part of the chapter I draw a number of distinctions between sufficiency views and I distinguish between a maximizing conception of rationality and a satisficing one, using the latter concept in the sense developed by Michael Slote. In the second part I build two test cases and study the normative prescriptions which various sufficiency views offer in each of them. I conclude that resource sufficientarianism does not provide the correct response to the first case, since its distributive prescriptions

would violate the principle of personal good and that subjective-threshold welfare sufficientarianism as well as objective-threshold welfare sufficientarianism committed to the headcount claim do not provide correct responses to the second case, since their distributive prescriptions would violate the principle of equal importance. I then claim that an objective-threshold welfare sufficientarian view committed to prioritarianism under the threshold offers the correct response to both cases and therefore resists the challenge raised by scenarios that involve differential conceptions of rationality. This result illustrates the more general claim that, in some cases, incrementally modifying the inputs of a normative model may lead to a decrease (or, alternatively, increase) in the desirability of that model.

In chapter 6 I aim to explore the effects of relaxing rationality assumptions in respect to computational capacities on Ronald Dworkin's theory of egalitarian justice. In the first part I outline the content of resource egalitarianism, by appealing to Dworkin's distinction between option luck and brute luck and his hypothetical insurance device. In the second part I attempt to clarify an ambiguity concerning the proper site of responsibility in Dworkin's theory, arguing that it is represented by preferences—not choices—as Dworkin and other authors (such as Carl Knight), sometimes suggest. Since preferences are the proper site of responsibility for Dworkin, I go on to maintain that the process of converting preferences into choices may be affected by systematic reasoning errors, which distort individual computational capacities in a morally relevant way. I use a hypothetical case to argue that the principle of equal opportunity to insure against bad luck is undermined by the empirically plausible assumption that individuals can make reasoning errors and, therefore, that Dworkin's theory is stricken by a different and deeper strand of unfeasibility than the one which leads him to suggest that counterfactual compensation should ultimately be enacted as a matter of policy. This result illustrates the more general claim that, in some cases, incrementally modifying the inputs of a normative model may lead to a decrease (or, alternatively, increase) in the feasibility of that model.

Finally, in chapter 7 I attempt to refute a recent challenge raised by Michael Otsuka against prioritarianism, according to which the priority view is objectionable since it rejects the moral permissibility of choosing in accordance with rational self-interest, understood as maximization of expected utility, in one-person cases involving other-regarding decision-

making under risk. I claim that Otsuka's argument is bound to make an illegitimate move which is either to implausibly assume that individuals are risk-neutral or to implausibly assume that the decision-maker in his cases can have accurate information on the attitudes towards risk held by the individual on behalf of whom the decision is taken. I argue, *pace* Otsuka, that acting in accordance with rational self-interest in such cases requires that we adopt a view which takes into account general empirical facts about human nature and that prioritarianism does not conflict with this latter view. This result illustrates the more general claim that, in some cases critiques of normative models are themselves reliant on a specific combination of inputs and that they may be refuted under different, and more empirically plausible, assumptions.

**Alexandru Volacu** is an adjunct professor at the National University of Political Science and Public Administration (SNSPA), in Bucharest, Romania. Recent publications include the volumes *Modern dilemmas: understanding collective action in the 21st century* (co-editor, Ibidem-Verlag) and *Rational choice and electoral behaviour* (co-author, Tritonic), as well as journal articles published in *Contemporary Political Theory* and the *Journal of International Relations and Development*. He is currently co-authoring a chapter on the ethics of voting (together with Annabelle Lever) as part of the *Routledge handbook of ethics and public policy* (forthcoming). His research interests include theories of justice, the ideal/non-ideal theory debate, the ethics of voting, electoral behaviour, rational choice theory, and game theory.

Contact e-mail: <alexandru.volacu.14@politice.ro>