ERASMUS JOURNAL FOR PHILOSOPHY AND ECONOMICS
VOLUME 16, ISSUE 1, SUMMER 2023

The Erasmus Journal for Philosophy and Economics (EJPE) is a peer-reviewed bi-annual academic journal supported by the Erasmus Institute for Philosophy and Economics, Erasmus School of Philosophy, Erasmus University Rotterdam. EJPE publishes research on the methodology of economics, history of economic thought, ethics and economics, and the conceptual analysis of inter-disciplinary work relating economics to other fields. EJPE is an open-access journal. For additional information, see our website: <http://ejpe.org>.

EDITORS
Savriël Dillingh
Elisabetta Gobbo
Ruth Korte
Benjamin Mullins
Erica Celine Yu

ADVISORY BOARD

EDITORIAL ASSISTANTS
James Grayot, Gergana Boncheva

ISSN: 1876-9098
TABLE OF CONTENTS

ARTICLES

Ideologies and Utopia: A Ricoeurian Reading of Thomas Piketty
*BENOÎT WALRAEvens* [pp. 1–27]

BOOK SYMPOSIUM on *Exploitation as Domination*

Editorial Note
*TULLY RECTOR, ELISABETTA GOBBO, AND BENJAMIN MULLINS* [pp. 28–30]

The Injustice of Domination
*S.M. LOVE* [pp. 31–42]

What Makes Exploitation Wrongful?
*LUCAS STANCZYK* [pp. 43–55]

Exploitation and Domination in Application
*GULZAAR BARN* [pp. 56–68]

Responses
*NICHOLAS VROUSALIS* [pp. 69–76]

SPECIAL CONTRIBUTIONS

Reflections on the 2021 Nobel Memorial Prize Awarded to David Card, Joshua Angrist, and Guido Imbens
*LENNART B. ACKERMANS* [pp. 77–96]

Reflections on the 2022 Nobel Memorial Prize Awarded to Ben Bernanke, Douglas Diamond, and Philip Dybvig
*JENS VAN ‘T KLOOSTER* [pp. 97–118]
BOOK REVIEWS

Conrad Heilmann and Julian Reiss’ (eds.)
The Routledge Handbook of Philosophy of Economics
D. Wade Hands
[pp. 119–129]

Robert Skidelsky’s What’s Wrong With Economics?:
A Primer for the Perplexed
Ella Needler and Maria João Pimenta
[pp. 130–136]

Susumu Egashira, Masanori Taishido, D. Wade Hands, and
Uskali Mäki’s (eds.) A Genealogy of Self-Interest in Economics
Stavros A. Drakopoulos
[pp. 137–143]

Mariusz Maziarz’s The Philosophy of Causality in Economics:
Causal Inferences and Policy Proposals
Fernando Varela Levy
[pp. 144–151]

Michael A. Wilkinson’s Authoritarian Liberalism and
the Transformation of Modern Europe
David Hollanders
[pp. 152–155]

Anna Horodecka’s Human Nature in Modern Economics:
Structure, Change and Perspectives
Andres Lazzarini
[pp. 156–163]

PHD THESIS SUMMARIES

Gender in the Labor Market
Anne Sophie Lassen
[pp. 164–167]

Elucidating the Role of Value Judgments in Normative Economics
Nestor Lovera Nieto
[pp. 168–171]

A Model-Based and Mechanistic Approach to Social Coordination
Matti Sarkia
[pp. 172–176]

Uncertainty for Uncertain Decision-Makers
Malvina Ongaro
[pp. 177–179]
Ideologies and Utopia:  
A Ricoeurian Reading of Thomas Piketty

BENOÎT WALRAEVENs  
Université de Caen Normandie

Abstract: In his most recent books, Piketty offers a global history of inequality in its economic, social, political, and intellectual dimensions, arguing that history is moved by the struggle of ideologies. To take part in this battle of ideas, he conceives a new ideal model of society, ‘participative socialism’, as an egalitarian alternative to the dominant neoproprietarian ideology and to the dangerous resurgence of nationalism and populism. This paper provides a new interpretation of Piketty’s view of history and of his participatory socialism in light of Paul Ricoeur’s study of the dialectics of ideology and utopia. First, I present Ricoeur’s singular analysis of ideology and utopia, which he sees as two inseparable facets of social imagination. Then I show how Ricoeur’s concepts can be fruitfully applied to Piketty’s conception of history and to his conception of a new form of socialism for the 21st century, drawing lessons from history and forming a ‘good’ or ‘realist’ utopia. Finally, I demonstrate that this interpretation of Piketty’s socialism can help to better understand some of the criticisms he has received.

Keywords: Piketty, ideology, utopia, realism, Ricoeur, participatory socialism

JEL Classification: B5; P1; P3

I. INTRODUCTION

Soon after the fall of the Berlin Wall and the disintegration of the Soviet Union, some were quick to proclaim ‘the end of history’ (Fukuyama 1989, 1992), and along with it, the death of socialism (Hodgson 2019, Introduction) and of utopian thinking (Hodgson 1999, xv). For many, and especially for the proponents of neoliberalism, there would be no alternative to liberal democracy and market-based capitalism. Concomitantly, socialism and communism were seen as ideologies of the past and relegated to the dustbin of history. However, the combination of the economic crises

AUTHOR’S NOTE: I’d like to thank the two referees and the editor for their comments and suggestions on previous versions of this paper which have led to significant improvements.
(the 2008 financial crisis and the Covid crisis), social and political crises (rising inequalities, the crisis of the social state, the rise of populism), health crisis (Covid-19), and the ever more pressing environmental crisis urge us, on the contrary, to rethink our model of society. More precisely, it pushes us to rethink its ideological representations in the social imaginary, and thus, to imagine new alternatives, or ‘utopias’. The definition and meaning of ‘utopia’ have long been hotly debated questions (see Levitas [1990] 2010). I will follow Levitas in defining utopia in a broad sense, reconciling its various definitions in terms of forms, contents, and functions, as the expression of “the desire for a better way of being and living” ([1990] 2010, 8).

While the nineteenth century has been seen as the golden age of political utopia, the twenty-first century might be a new age of utopia, when it is more necessary than ever. The time is ripe for utopia as the question of the purpose of society or social life arises because of the limits of our way of living. But the word ‘utopia’ still often connotes an inaccessible, unfeasible, and therefore useless ideal. It is used to discredit any radical project of reform or any alternative project of society. Yet, utopian thinking is experiencing a certain return to favour (Wright 2010; Macherey 2011; Chrostowska and Ingram 2016; Bregman 2017), which owes as much to the material circumstances that were aforementioned as to contributions in the world of ideas by philosophers who have tried to rehabilitate the notion of utopia, such as Abensour (Cervera-Marzial and Poirier 2018), Ricoeur (Roman 2021a), or Levitas ([1990] 2010), among others (Chrostowska and Ingram 2016). Utopian desire, or the hope for a better way of living, has never completely disappeared, probably because it is a significant part of human culture. Yet, it is being reborn today. Some critical social scientists are working to rethink the boundaries of the possible (Gueguen and Jeanpierre 2022), and to reinvest the utopian space (Levitas [1990] 2010, chapter 7), long neglected, through a new Marxist, anarchist, or radical democratic perspectives (Chrostowska and Ingram 2016), reinventing liberalism (Gamel 2021), communism (Badiou 2009, Friot [1989] 2012; Friot and Bernard 2020; Lordon 2021; Friot and Lordon 2021), and socialism (Cohen 2009; Wright 2010; Honneth 2017; Hodgson 2019), or imagining an ‘ecologic City’ by building on a multiple heritage (Audier 2020).

Piketty follows this trend in his recent books (Piketty 2020, 2022a), I argue, provided we first carefully define what kind of utopia he envisions. Although he never explicitly presents his ideal model of society, i.e.,
‘participatory socialism’ (Piketty 2020), as a utopia in Capital and Ideology (henceforth C&I) or later, in A Brief History of Equality, I think that there are good reasons to interpret it as such, and in a positive and constructive manner. With C&I, Piketty clearly intended to overcome what he sees as the main limitations of his preceding book (Piketty 2020, Preface), the now famous Capital in the Twenty-First Century. He does this by providing (i) a more ‘global’ history of inequality, and (ii) an analysis of the ideologies that served to legitimize inequality throughout human history. Additionally, in C&I, Piketty no longer simply offers a few normative policy proposals for reducing inequality, such as a global wealth tax, that he interestingly presented as an “ideal” and a “useful utopia” (2014, 515). He now formulates explicitly some underlying principles of social justice that are embodied in a model of society, a new form of socialism called participatory socialism seen as one possible model for reforming society and surpassing capitalism. Fighting explicitly against fatalism and disenchantment, Piketty fuels the utopian spirit, showing that another world is not only possible but necessary, calling for new alternatives.

The purpose of this paper is to provide a new interpretation of Piketty’s vision of history as the struggle of ideologies and of his project to rethink socialism for tackling the issues of the twenty-first century. First, it aims to enrich its philosophical foundations and its normative framework, on which he has been criticized. In doing so, this article reinterprets philosophically Piketty’s vision of history in C&I and studies his ideal model of society, participatory socialism, through the prism of Ricoeur’s innovative but little-known analysis of the dialectics of ideology and utopia. It thus provides a new interpretation of Piketty’s participatory socialism as a ‘realist’ utopia and as a ‘good’ utopia in the sense of Ricoeur, i.e., as a practical and political utopia, or as an alternative and ideal model of society that relies on both history and on real, existing institutions (and also, presumably, on unexploited possibilities of the present to formulate a society that is desirable and, at least in part, achievable and realizable). Second, and relatedly, I will show too that this perspective on Piketty can help to better understand some of the criticisms he has received, and that it finds some echo in current debates in normative political theory on political feasibility, utopianism, and realism.

1 This crossed reading of Piketty and Ricoeur is not motivated by a willingness to unveil an unknown influence of the latter on the former, nor even by the observation of references to Ricoeur in Piketty’s works, since I did not find any. Of course, it does not mean that Piketty did not read Ricoeur’s works.
The paper is organized as follows. Section 2 deals with Ricoeur's singular analysis of the dialectics of ideology and utopia, which he sees as two inseparable facets of social imagination. His views on the links between ideology and utopia underlines the main characteristics of his approach, and its distinction between good and bad utopias. Sections 3 and 4 try to apply Ricoeur's reflections on ideology and utopia to Piketty's thought to enrich its normative framework, first, in view of the latter's conception of history as a struggle of ideologies, and second, with respect to the alternative socio-economic system he describes in his latest works, participatory socialism. As to the first point, I question whether Piketty defends an idealist, teleological, and dialectic view of history (section 3). In section 4, I discuss the multiple origins, the foundations, and the nature of Piketty's socialism and its utopian character through the lens of Ricoeur. I claim that the categories of 'utopianism' and 'realism', which are harshly debated today in political theory, helps in understanding the reception of Piketty’s project. I conclude by summarizing the results of this ‘Ricoeurian’ interpretation of Piketty, adding further thoughts on how his works may help economists to (re)consider economics as a critical, engaged, and emancipatory social science, cultivating the utopian spirit.

II. RICOEUR AND THE DIALECTICS OF IDEOLOGY AND UTOPIA

Ricoeur's singular analysis of ideology and utopia is part of his wider and long-lasting reflections on imagination (Amalric 2021, 37–38). For Ricoeur, ideology and utopia are two conflicting though complementary “idealities” (Roman 2021c, 134). Inspired by Mannheim, Ricoeur thinks about ideology and utopia together. But what characterizes his approach is to think of them as two inseparable facets of our “social imagination” (Ricoeur 1986a, 1), or more precisely, as transhistorical structures of social imagination working at, respectively, the conservation and the transformation of social order. Ricoeur builds upon an innovative three-levels analysis for ideology and utopia, these levels echoing and responding to each other, and he underlines a double polarity both within ideology and utopia and between them.

III. The Two Inseparable Facets of Social Imagination

These two idealities at the heart of our social imagination establish our collective identity. In order to exist, every society needs a representation of its own identity. On the political level, “the social imaginary is reflected in the conflicting tension between ideology and utopia” (Roman 2021a,
23). No representation of social order is possible without ideology and utopia. For Ricoeur, ideology needs utopia, and vice versa, though it is often difficult to think of them together because each of these two poles is often presented “in a polemical and sometimes pejorative sense that prevents us from understanding the social function of the collective imagination” (1984, 53). There is a “crossed game” and “mutual exchanges” between ideology and utopia (Ricoeur 1986b, 260; 1984, 1963). We always need utopia in its fundamental functions of contestation and projection into a radical ‘elsewhere’ to cure the rigidity and petrification of ideologies. But we also need ideology to cure utopia from the “madness” it may fall into and to provide a historical community with a narrative identity (Ricoeur 1984, 63–64; 1986a, 310). In the end, there is always an unsurpassable tension between ideology and utopia that are dialectically linked, so much so that it might be difficult sometimes to decide whether a mode of thinking is ideological or utopian (Ricoeur 1986b, 261). According to Ricoeur, “we are always caught in this oscillation between ideology and utopia,” and thus, we have no choice but to “let ourselves be drawn into the circle and then try to make it a spiral” (1986a, 310).

Both ideologies and utopias are necessary for our collective narratives and identity. Ricoeur provided a rehabilitation and redefinition of utopia at a time when it was decried. Accordingly, utopia is neither reduced to a mere literary style, nor to a critical function of reality. Following Mannheim, Ricoeur gives priority to ‘practical’ and political utopias over literary utopias, favoring Saint-Simon and Fourier over More and Campanella, because the former aim at the contestation and concrete transformation of our historical and social reality. Utopia has a projective and practical function. That is why Ricoeur sees the foundation of utopia in Müntzer and in the works of the first socialists rather than in More. There is in human beings an original desire for utopia which is an original desire of the right and the good, prompting the variations of utopian imagination and underlying the protests and transgressions of practical utopia (Amalric 2021, 52–53). In Abensour’s words, man is a “utopian animal” who always craves for emancipation (2013, 17-18, 20, 31, 60). For Ricoeur, it is impossible to imagine a society without utopia because it would be a society without purpose (1986a, 283). Utopia is an escape from, and the ‘weapon of criticism’ of, social order. It helps us to understand the contingency of existing order (299–300), to take a fresh look at our reality,

---

2 For more details on Mannheim’s views on utopia and its limits, see Levitas ([1990] 2010, chapter 3).
“in which nothing can be taken for granted anymore”, and to open the field of possibilities beyond reality (Ricoeur 1986b, 257–258). Utopias are imaginative variations about power and authority, and they are indispensable to achieving a vibrant democracy driven by the immanent tension between the power that is exercised and the possibility of challenging it (Roman 2021b, 21). Utopian pluralism and the conflict of utopias (and ideologies) are necessary conditions of the democratic project (Amalric 2021, 55).

II.II. Commonalities and Differences between Ideology and Utopia
For Ricoeur, what ideology and utopia have in common, and what makes it possible to think of them together, is their necessary “non-conformity” (“noncongruence”) with social and historical reality, as Mannheim had noticed (Ricoeur 1986a, 3). But ideology exercises a resistance to change and aims at preserving the past order, while utopia is oriented toward the future and brings about change, breaking through “the thickness of reality” (309). Ideology is mainly linked to dominant groups, whereas utopia is supported by emerging groups. Utopia can be seen as a kind of transformative and developing ideology. Ricoeur stresses, in an original way, that the polarity between ideology and utopia illustrates the two sides of social imagination (310). Ideology corresponds to the conservation or reproductive function of imagination, our memory of things through “paintings”, “portraits”, and “pictures” (Mongin 2021, 61). Utopia, by contrast, represents “fiction”, the constructive or productive function of imagination (Ricoeur 1986a, 310; Mongin 2021, 61). It reconfigures reality through the figuration of a nowhere, a u-topia.

There is, in fact, a double polarity between ideology and utopia, both among them and within each of them (Ricoeur 1986b, 254). Ideology and utopia have both a positive and constructive role, and a negative and destructive side, or sound and pathological forms (Ricoeur 1984, 54). Inspired by Geertz, Ricoeur emphasizes that the positive side of ideology is to preserve collective identity and social order—its integrative function—which is pre-existent to any distortion. But against Geertz, he claims that we still need to criticize ideologies, and this is precisely the role of utopia (Amalric 2021, 42). Its positive side is to explore the possible, or its subversive, critical function. On the negative side, ideology is conservative and distorts reality, often to the advantage of the dominant groups,

---

3 For Hodgson, too, “the word ‘utopia’ fosters a likelihood of change, and points to an unfulfilled future that differs from the present” (1999, 4).
whereas utopia can appear as mere fantasy, as ‘an escape from reality’ or as the chimerical dream of an unreachable ideal.

Ricoeur goes further than Mannheim\(^4\) into the analysis of the dialectics of ideology and utopia by identifying, through a specific method which he calls a “regressive method” (Ricoeur 1986a, 311), three parallel levels of depth for each, from the most superficial level to the most profound. The motivation for this method is to investigate beyond the apparent meanings of ideology and utopia, i.e., their “pathological” meaning, to reach their more essential and positive meanings (Amalric 2021, 46). It thus innovatively combines, rather than separates, different meanings of these concepts and formulates a hierarchy among them.\(^5\) Moreover, it allows Ricoeur to highlight a dynamic correlation between ideology and utopia.

Let’s start with ideology. There are, Ricoeur claims, three uses of the concept, all of them equally legitimate, corresponding to three levels of depth. On the first, most superficial level, remaining at the surface of the phenomenon, we find ideology as distortion and dissimulation. This is Marx’s early view of ideology as presented in his *Economic and Philosphic Manuscripts of 1844* and in *The German Ideology*, in which ideology is supposed to offer an inverted view of reality as opposed to praxis (Ricoeur 1984, 54–56). On the second, deeper level, ideology is the legitimation or justification of domination (56–57). It is close to Marx’s notion that the ideas of ruling classes become the dominant ideas by being presented as universal ideas. Domination always wants to justify itself, Ricoeur notes, and ideology is necessary to the process of legitimation of power and authority. Inspired by Weber and Arendt, Ricoeur claims that there is always “a gap to close, a kind of surplus-value of belief that every authority needs to extort from its subordinates” (57). On the third, deepest and most fundamental level, ideology should be seen as “integration into social memory” (58-60). In Ricoeur’s words: “Any group is held together by the stable and lasting image it creates of itself. This stable and lasting image expresses the deepest level of the ideological phenomenon” (59). For him, this illusion is not the most important phenomenon, it is only a corruption of the process of legitimation which is rooted in the integrative function of ideology.

---

\(^4\) For more details on Ricoeur’s analysis and critique of Mannheim on ideology and utopia, see Ricoeur 1986a, chapters 10 and 16.

\(^5\) On the multiple meanings given to ideology in political theory, see Geuss (1981) and Freeden (2003).
Ricoeur then offers a parallel analysis of utopia. The three levels of ideology and utopia complement one another. While under its three forms ideology reinforces, redoubles, and preserves a social group and order, providing an interpretation of real life, utopia is meant to project the imagination out of reality in an elsewhere which is also a nowhere, in a spatial and temporal exteriority (60). There are three parallel levels of utopia that Ricoeur studies in a reverse order, from the bottom up because, he argues, it’s easier to show how utopia, in its fundamental sense, is the necessary complement of ideology in its fundamental sense. On the third, deepest and most fundamental level, utopia is the proposal for an alternative society: “If ideology preserves and maintains reality, utopia essentially questions it” (60). All areas of life in society are concerned, be it family, modes of appropriation and consumption, or political and religious life. Utopia cannot be defined by its content, but rather by its main function, which is always to imagine an alternative society, against the integrative function of ideology. On the second level, utopia represents a questioning of power, responding to the parallel middle level of ideology, which aims at legitimizing authority (61–63). Utopias are seen by Ricoeur as ‘imaginative variations’ of power and human emancipation. At last, with the third, most superficial level of utopia, we reach its pathological form, which is the opposite of that of ideology: “Where ideology reinforces real life, praxis, utopia makes the real itself vanish in favour of perfectionist, ultimately unattainable schemes” (62). It is the negative side of utopia, when the ‘crazy’ all-or-nothing logic replaces the logic of action. Thus, for Ricoeur it’s important to come back to the positive, liberating function of utopia, “concealed by its own caricatures” (63). Utopia imagines the no-where and opens the scope of possibilities.

II.III. Good and Bad Utopias
But what are the criteria for distinguishing ‘good’ from ‘bad’ utopias? Ricoeur explores this issue, too, and provides his own answer that I will use later to assess Piketty’s proposal of an alternative socio-economic system, his participatory socialism. Inspired by Bloch’s concept of “concrete utopia” (1986, 157), Ricoeur distinguishes good from bad utopias, or utopia from utopianism in the following manner: good utopias are ‘practical’, political utopias actualizing latent potentialities from the past and present and aiming at effectively transforming social order, while bad ones are an escape from reality, or the chimerical dreams of a perfect though unrealizable society, often based on an unrealistic view of human nature and
without regard for history. The “utopian mentality” is dangerous when it leads to “the absence of any reflection of a practical and political nature on the support that utopia can find in the existing reality, in its institutions,” and when it is unable to “designate the first step that should be taken in the direction of its realization from the existing reality” (Ricoeur 1984, 62).

Utopia must be a space of exploration of other possible worlds, societies, institutions, relying on the lessons of history. Following Bloch, Ricoeur decides to re-historicize utopias, seen as “unfulfilled promises of the past,” or “variations on the possible, which history keeps in mind” (Pierron 2021, 115). As Roman rightly noted, for Ricoeur utopia is “the production of a new form determined by an already-there, by an a posteriori appropriation of not yet actualized possibilities which gives them a new meaning” (Roman 2021a, 18), and which counterbalances the “nowhere” of utopia, otherwise condemned to be a mere reverie (Foessel 2021, 93). The end of the good utopia is to find a proper balance between what Ricoeur calls the “desirable absolute” and the “achievable optimum,” or between the “expectation horizon” and the “experience space” (Roman 2021a, 10). A good utopia reconciles the ‘ethics of conviction’, which is defined by Ricoeur as the ‘excellence of the preferable’, with the ‘ethics of responsibility’ defined by what is realizable in a specific historical context (Ricoeur 1986b), while a bad utopia is impracticable. Utopia defines an ideal, but if it wants to serve as a regulating horizon and motivate action, it must be based upon experience and be practical and feasible, at least in part. It has more to do with finding the best compromise than with the posture of radical conflict (Roman 2021a, 13–14). But Ricoeur does not explain to us how to identify these latent possibilities of the past and present. That’s where Piketty’s work, and critical social science in general, might play an important role. The careful study of social experimentations throughout history, of their successes and failures, offers a wide range of experiments of human inventiveness from which we can draw some lessons to better understand which kind of institutions are possible and necessary to revive what Piketty sees as a long-term movement towards equality.

III. PIKETTY’S VIEW OF HISTORY: A PHILOSOPHICAL PERSPECTIVE
Ricoeur's analysis of the dialectics of ideology and utopia can be fruitfully applied to Piketty’s most recent works (Piketty 2020, 2022a), and more specifically, to his views on ideologies, which Piketty considers to be the
driving forces of history. His interest in history is of crucial importance in his works. He regards his work as history and social science, as a continuation of the new economic and social history promoted by, among others, the French Annales School of history (Piketty 2022a, 5). What interests us most here, however, is Piketty's views on ideologies, which take center stage in his big narrative upon the global history of inequality. He presents history as the “struggle of ideologies” (Piketty 2020, 1035-1036) and a quest for justice and equality, with different “inequality regimes” succeeding over time (2-3). Piketty’s history of inequality is not merely descriptive (see by contrast Scheidel 2017). This section will thus study Piketty’s view of history from a philosophical perspective. Even though Piketty does not try to offer a new philosophy of history, it is interesting to identify the main features of his history of inequality and to question its idealist, teleological, and dialectical dimension.

III.I. An Idealistic View of History?
I explore first the traditional opposition between the idealist and the materialist views of history, which Piketty uses to clarify his position and, more generally, to distance himself from what he sees (somewhat distortedly) as the Marxist view of history. Contrary to most of his fellow economists, Piketty claims that inequality is mostly due to political and ideological causes, setting the degree of acceptability of society towards the level (and forms) of inequality, rather than to economic or technological causes (2020, 7). He starts from the idea that every society needs to make sense of and justify its inequalities to perpetuate itself (1). In doing so, it develops different and competing discourses and narratives, or ideologies, with a dominant ideology structuring the existing “inequality regime”, based upon a “political regime” and a “property regime” (4).

In “a self-conscious inversion of Marx” (Dennison 2021, 162), Piketty argues that: “The history of all hitherto existing societies is the history of the struggle of ideologies and the quest for justice” (2020, 1035). Ideas and ideologies “count in history,” and have their own autonomy (1035, 7–10). For Piketty, the political and ideological superstructure is not (fully) determined by the economic infrastructure: “Given an economy and a set of productive forces in a certain state of development [...] a range of possible ideological, political, and inequality regimes always exists” (8). These clear albeit provocative statements very much sound like an idealist view of history (Brisset and Walraevens 2020; Motadel and Drayton 2021). But Piketty’s position is more nuanced than it seems at first sight, especially
in his latest book, *A Brief History of Equality*, maybe in response to some of the criticisms he faced for *C&I*. Moreover, even in the latter book, there are Marxist overtones in his view of history and ideology, which is not without creating some tensions (Dennison 2021; Morgan 2021). Piketty does not deny that power relationships and social struggles play an important (though limited) role in history, especially in the transition from one inequality regime to another (Piketty 2022a, 13–15). What he fights against, I think, is an overly mechanistic and deterministic view of history that very few would endorse today. For Piketty, the study of history teaches us that similar levels of material and productive forces can always lead to different outcomes and paths, depending on the prevailing ideological balance of power. History exhibits different ‘trajectories’ and astonishing ‘bifurcations’ (or ‘switch points’) (Piketty 2020), as in Sweden in the 20th century with the emergence of social democracy. Eventually, Piketty seems to adopt a kind of reasonable mixed view between idealism and materialism (Piketty 2022a, Introduction), giving credit to both, though he appears more on the side of the former than of the latter with his constant emphasis on the role and independence of ideas in history.

Now, what is his definition and view of the notion of ideology, so central to his history of inequality? It is defined “in a positive and constructive way”, but somewhat vaguely, as “more or less coherent discourses” and “a set of a priori plausible ideas” on “how society should be structured”, having economic, political, and social dimensions (Piketty 2020, 3, 9). Or again, as “an attempt to respond to a broad set of questions concerning the desirable or ideal organization of society,” expressing an idea of social justice (3, 9).

The conceptual limits of Piketty’s notion of ideology have been repeatedly underlined (Brisset and Walraevens 2020; Motadel and Drayton 2021; Reddy 2021; Zevin 2021). Here it seems that Ricoeur’s analysis of ideology can be helpful to complete and enrich Piketty’s view by providing stronger philosophical foundations to it. For the latter, ideologies are essentially discourses and narratives serving to justify inequalities “more or less truthfully” (Piketty 2020, 3; 1035), and defining two things: political borders (the “political regime”) and property (the “property regime”). In other words, ideologies define who has rights and who owns what, both of which suggest “the exclusion of others” (Witztum 2021, 3). Thus, Piketty’s view of ideology seems closer to Ricoeur’s second level of ideology, i.e., as a legitimation of existing power and inequality or domination. Sometimes Piketty’s analysis of ideology might also be reminiscent of Marx’s
early view of ideology as distortion and dissimulation, especially in his important critique of the fallacy and illusion of meritocracy in neoproprietarianism, which he sees as today’s dominant ideology (Piketty 2020, 1–2, 709–713). More generally, there are Marxist overtones in Piketty’s overall presentation of proprietorarian societies being born out of and held by an ideology supported by the elites, the rich and powerful to confer legitimacy on their rights to property, which is then adopted by society (Dennison 2021). What reading Ricoeur fundamentally adds to Piketty’s mostly descriptive and neutral view of ideology is the idea that there are different meanings and levels of ideology which combine, and further, that ideologies have—first and foremost—a structural and constructive though conservative power, creating and supporting a certain view of collective identity and memory. They help to maintain the stability of social order and identity or, in Piketty’s words, of the inequality regimes, and mainly serve the interests of dominant groups. That’s why it is difficult to replace them. Piketty’s view has sometimes strikingly ‘Ricoeurian’ overtones, for example, when he emphasizes that “every ideology attempts […] to impose meaning on a complex social reality” because “human beings will inevitably attempt to make sense of the societies they live in, no matter how unequal or unjust they may be” (2020, 16). Ricoeur also importantly adds that ideologies are inextricably linked with the imagination of alternative views of society and social justice, or utopias.

III.II. A Teleological View of History?
There is also in Piketty’s view of history an apparent teleological dimension. In several passages in C&I and maybe even more in A Brief History of Equality, Piketty seems to assign to modern history a specific end or telos, namely equality. History, he tells us, is a relentless quest for justice and equality (Piketty 2020, 1035–1037). The conclusion he draws from his historical analysis of inequality regimes is that “what made economic development and human progress possible was the struggle for equality and education and not the sanctification of property, stability, or inequality” (3). There is “a historic trend toward ever greater equality” (33). The title of Piketty’s latest book (Piketty 2022a), as well as the title and first lines of its first chapter, “The Movement toward Equality,” are quite explicit: “there has been a long-term movement over the course of history toward more social, economic, and political equality” (1).

Thus, at least since the 18th century, there is for Piketty a real, long-term though unachieved tendency toward equality, an advance toward the
equalization of conditions that is reminiscent of Tocqueville, and a faith in the progress of society. But it does not mean that for him history is predetermined and linear. Quite the contrary: His works on highest incomes and subsequent inequalities have long highlighted the rise of inequality since the 1980s, leading to a rejection of the famous inverted-U Kuznets curve (Piketty 2014). There is now a form of consensus in the economic and political sphere on the ‘revival of inequality.’ But in Piketty’s narrative, it leads to “reversing the historic trend toward ever greater equality” (Piketty 2020, 33). It is seen as an exception to the rule, to the general levelling trend in modern history, and as a consequence of the failures of social democracy and of the fall of communism, which led to the rise and dominance of neoproprietarian ideology. The history of equality is not “a linear one” and is “punctuated by multiple phases of regression and identitarian introversion” (Piketty 2022a, 1). At every moment of history, multiple “trajectories” are always possible (Piketty 2020, 1037). There is a fundamental contingency of history.

Piketty’s highlighting of a trend of history toward equality is meant to be a descriptive point of view, a statement based upon a careful analysis of scientific data on inequality by social scientists, rather than a philosophical or normative analysis of the predetermined telos of human history. But he switches more clearly to a normative stance when he imagines the future and argues that the historical quest for equality and justice can and will continue (Piketty 2022a, viii), at least if we enter the battle of ideas and support a new ideology of equality with its attendant socio-economic system—participatory socialism—though its implementation would not be the ‘end of history’. Piketty’s idealism and his claimed optimism join here.

III.III. A Stadial and Dialectical View of History?
Piketty’s narrative about human progress toward equality is based upon the observation and distinction of different and successive inequality regimes, supported by a specific ideology and institutional framework: from the trifunctional societies of the premodern era, to the ownership societies with first proprietarian societies and then communist societies and social democracies, to today’s hypercapitalist societies founded on the neoproprietarian ideology. This typology of inequality regimes might at first sight look like a stadial theory of history, in the spirit of the great narratives about the progress of society in the 18th century. But these ‘stages’ of modern history and distinct inequality regimes are not meant
to be a natural course through which every society should necessarily pass. Rather, they help us to identify the numerous forms that social regulation has taken over time and space, the infinite inventiveness of human societies to justify and regulate their inequalities through different institutional schemes. For Piketty, there is a big fund of “historical experimentation” (Piketty 2020, 10, 41, 116, 119, 495) from which we can draw some lessons to imagine new forms of ideology and inequality regimes because we face an open future with multiple possibilities for regulating inequality.\(^6\) Piketty’s history looks at the past, but always with an eye on the future.

Finally, Piketty’s view of history seems to have a dialectical aspect. Each inequality regime has its own limits and contradictions, or internal sources of instability, which tend to exacerbate themselves leading to a crisis and to the need for of a new narrative about inequality—a new inequality regime with institutions implementing a different political and property regime. The rise of social democracy is seen as a reaction to the increasing concentration of incomes and wealth in proprietor societies which became more and more difficult to justify in proprietor societies, and to the threat of communist societies. Likewise, today’s continued increase of inequality under neoproprietarianism and its growing inability to justify them by the meritocratic narrative—not to mention its inability to answer to climate crisis—all combine to fuel peoples’ dissatisfaction, which translates into the rise of social nativism, populism, and nationalism, and calls for an egalitarian alternative (Piketty 2020, 1–3, 821, 825).

**IV. Piketty’s Utopia: Participatory Socialism**

Piketty does not simply describe the characteristics and limits of the different regimes of inequality which existed throughout human history; he also wants to actively participate in the struggle of ideologies by imagining a new egalitarian ideology, or a utopia, in Ricoeur’s sense, which would continue what he sees as the long-run march of modernity toward equality. Some passages of the conclusion of C&I sound quite Ricoeurian in spirit, for example, when he writes that “human societies have yet to exhaust their capacity to imagine new ideological and institutional solutions” (1034). Or when he claims that not only do we have to “carefully scrutinize today’s inequality regimes and the way they are justified.

\(^6\) As remarked by Morgan (2021), it means that there are no laws of capitalism for Piketty, what seems to be in contrast with his position in *Capital in the 21st Century* in which he argued for two fundamental ‘laws of capitalism’.
Above all, we need to understand what institutional arrangements and what types of socioeconomic organization can truly contribute to human and social emancipation” (1035). Imagining a new form of egalitarian inequality regime is the topic of the last chapter of C&I in which he provides, for the first time, well-defined normative foundations for his repeated policy proposals to reduce inequality. We find here the elements of a theory of social justice, or an economic theory of justice (Brisset and Walraevens 2021), defining an ideal socio-economic model—his theory of participatory socialism. For Ricoeur, “to develop a critique of ideology is thus always to oppose a utopian imagination to an ideological imagination: it is to bet on an imagination of the new against a dominant imagination already there and already operating” (Amalric 2021, 44). This is what Piketty does, I argue, with his critique of the dominant neoproprietarian ideology, of its sacralization of markets, private property and billionaire entrepreneurs, and of its meritocratic discourse, which only serve to naturalize inequality (Piketty 2020, 709–716). He does it too with his imagination of an alternative, participatory socialism, based upon the lessons of history.

IV.I. The Multiple Origins of Piketty’s Participatory Socialism

In order to elaborate his theory of social justice, Piketty draws on multiple sources (Piketty 2020, 970, fn. 3), among which one can identify the French “solidarism” of Fouillée, Bourgeois, and Durkheim with their ideas of “social debt” (562). This movement first appeared in France at the end of the nineteenth century and played a fundamental role in the emergence of its social state. It considered property to be essentially social, the product of a collective and intergenerational effort and cooperation, which should therefore (in part) be returned to society. Solidarism can be seen as a form of ‘liberal socialism’.

Another important source of influence on Piketty’s thought of social justice is Rawls’s theory of justice (Brisset and Walraevens 2020; Morgan 2021, section 5), which had a major influence on 20th century social science; though Piketty seems to underestimate this filiation (Brisset and Walraevens 2020). In particular, he appropriates Rawls’s difference principle, arguing that the just society “organizes socioeconomic relations, property rights, and the distribution of income and wealth in such a way as to allow its least advantaged members to enjoy the highest possible life conditions” (Piketty 2020, 968). Further, by advocating for a broader distribution of property through an increased progressivity of the
taxation of incomes and wealth and a capital endowment, and by defending a more democratic corporate governance, Piketty’s socio-economic model recalls the egalitarian version of “property-owning democracy” supported long ago by James Meade ([1964] 1993, 41–60), by Rawls himself, who refers to Meade and who considered property-owning democracy to be compatible with his own principles of justice (Rawls 1971), and more recently by Atkinson (2015) (see also Morgan 2021), praised by Piketty (2017).

Lastly, Piketty aligns with democratic socialism, which he sees as having reached a limited, unachieved equality (Piketty 2020, chapter 11). He wants his participative socialism to be seen as the next step toward equality, and thus, as a form of radical social democracy, extending what Piketty sees as its successes: progressive taxation and the social state (1036).

**IV.II. The Characteristics of Piketty’s Participatory Socialism**

Piketty defines a just society as “one that allows all of its members access to the widest possible range of fundamental goods” (Piketty 2020, 967). So, his theory of justice is based on a metric of ‘fundamental goods’ rather than utility, capabilities, or primary goods; and the rule of distribution on which it rests is to maximize these fundamental goods and ensure equal access to them for all (Brisset and Walraevens 2020). Piketty’s egalitarian theory of justice is founded on the equalization of resources. His concept of fundamental goods is not precisely defined, but he offers a preliminary list, open to debate and likely to change with the evolution of deliberation and new pressing problems for society, including education, health care, the right to vote, “and more generally to participate as fully as possible in the various forms of social, cultural, economic, civic, and political life” (Piketty 2020, 967–968). Piketty’s fundamental goods have an air of universal human rights, most of them being part of the United Nations 1948 Universal Declaration of Human Rights. As I already indicated, Piketty also endorses Rawls’s difference principle and its intent to maximize the situation and opportunities of the least well-off (968), and so he makes a plea for deliberative and participatory democracy, and social federalism. He sees the quest for justice as a collective learning process, based upon extended public deliberation over time and space, with strong cosmopolitan overtones (970).

That said, for Piketty, “it is wise to be wary of abstract and general principles of social justice” because what matters most are the “concrete
policies and institutions” reducing inequality and injustice (969). His participatory socialism is based on the idea of combining private, “temporary,” and “social” property (493). It pleads for a wide de-concentration and constant circulation of private property made possible by a progressive taxation of carbon emissions, of incomes, of wealth, and of its transfers with very high top marginal tax rates for each, allowing for the payment of a substantial capital endowment to everyone, which would guarantee more equality of opportunity, as well as through measures for promoting economic democracy or the power of workers inside firms, inspired by Swedish and German congestion.

Piketty’s project also contains a whole set of measures to ensure a truly equal participation of all in political life. He advocates a fair democracy, one which is not reduced to electoral processes but rather based on democratic and egalitarian deliberation founded on the development of non-profit media companies, on some control of the media, on stricter rules on party and campaign financing, and on the granting of “democratic equality vouchers” (1016), an idea he takes up from Julia Cagé, and which aims to establish a more egalitarian and participatory democracy in order to reduce the influence of private financing on political processes. On the international level, Piketty promotes the creation of transnational assemblies and of co-development treaties to fix crucial issues about borders, fiscality, international trade and investments, and climate change, focusing on global public goods. Piketty’s alternative project of society is meant to redistribute power in society and to provide maximum equal access to fundamental goods, allowing for the greatest participation of everyone to economic, social, and political life.

IV.III. Piketty, Socialism, and Utopia

Even though Piketty often claims to be a socialist, he neither defines socialism in his books, nor does he discuss the “complicated intellectual and conceptual history of socialism since the 19th century” (Langenohl 2021, 132; see also Zevin 2021, 78). Yet, it is interesting to try to identify the type of socialism that he defends because it helps to better understand some of the criticisms he has received. Certainly, he rejects both the authoritarian state socialism with central planification of the communist era (Piketty 2021; 2022a, 167) and today’s post-communist Chinese model, which he considers “an authoritarian mixed economy” or “a perfect digital dictatorship” (Piketty 2020, 606–611; 2022a, 230–237). What he offers instead is a “participative socialism” (Piketty 2020, chapter
17), or a decentralized, self-managed, democratic, ecological, and multicultural socialism for the 21st Century (Piketty 2022a, chapters 7 and 10). Though this new socialist model is turned towards the future, it has some roots in the past, and especially in solidarism and social democratic thought. We can also establish a filiation with the decentralized and self-managed socialism of the first socialists, as he himself acknowledged (Piketty 2021). Frobert (2019, 2021) has recently distinguished between two views of justice and equality, and thus, two kinds of socialism among the first, so-called ‘utopian’ socialists. Associated with Saint Simon and his followers, the first one relies on the equity criterion and aims for equality of opportunity, while the second one promotes a more radical view of economic equality, beyond mere equal opportunities, and can be found in the works of Blanc, Pecqueur, Raspail, and Sand. Frobert calls these strands of socialism, respectively, “below socialism” (“socialisme de l’en-deçà”) and “beyond socialism” (“socialisme de l’au-delà”) (2021, 7). Piketty belongs to the latter, I argue; but does it mean that he should be seen as a ‘(neo)utopian’ socialist? And if so, in what sense?

The answers to these questions will depend on the meaning we give to the word ‘utopian,’ especially in its association with socialism, which has a long and complex history. As noted by Levitas, “The Marxist tradition has for the most part been strongly antipathetic to utopianism, which it has understood as the construction of blueprints of a future society that are incapable of realization”, though ironically this charge has also been levelled against Marxism by its opponents ([1990] 2010, 41).

Marx and Engels presented a caricatural and dismissive view of what they called in The Communist Manifesto the utopian socialism of the first socialists, as opposed to their scientific socialism. But the real bone of contention between them and the utopian socialists “is not about the merit of goals or of images of the future but about the process of transformation, and particularly about the belief that propaganda alone would result in the realization of socialism” (Levitas [1990] 2010, 41). The main problem is their idealistic, as opposed to a materialistic, concept of social change. Likewise, the common meaning of the word is still pejorative today, representing utopia as an escape from reality, as an unrealizable project of society. Piketty’s project is not utopian in the latter sense. But it is, if we place Piketty in the filiation of the radical egalitarianism of the utopian socialists of ‘beyond socialism’, promoting an important

---

7 There are a few important exceptions within the Marxist tradition, like Bloch and Marcuse. See Levitas [1990] 2010, chapters 4 and 6.
redistribution of property and wealth and limited levels of inequality. All the more so knowing that in ‘beyond socialism' the theoretical principles of socialism, considered to be synonymous with the science of society, must be embodied in practical proposals and social reforms (Frobert 2019, 199, 206).

More importantly, Piketty’s participatory socialism is a utopia in Ricoeur's positive sense of the word. The latter wrote that “there are two kinds of utopia: those which are escapes and those which are programs and want to be realized” (Ricoeur 1986a, 289). Piketty’s participatory socialism clearly belongs to the latter. Utopia is meant here to be a possible future, rather than an impossible dream. Even though it is not fully implemented or implementable, his project of society is meant to be a useful, practical, and evolving guide for action, for reforming society towards more equality and justice. Participatory socialism is for Piketty a project to overcome capitalism, an alternative narrative about inequality and a new egalitarian ‘horizon’ having two related characteristics which make it a good, practical utopia in Ricoeur’s sense: (i) it is based on a more realistic view of human nature than communism, while being by contrast more “optimistic” on this front than neoproprietarianism, and (ii) “it is firmly rooted in the lessons of global history” (Piketty 2020, 3, 1037).

On the first point, Piketty argues that a fully public or state property, denying any private property, is unrealistic, untenable, and undesirable because it makes no room for the legitimate differences and necessarily non-identical aspirations of human beings (591–594). He privileges a social, “temporary” and “shared” vision of property, as a kind of middle way between the sacralization of private property and respect for individual aspirations in proprietarianism, and their denial in Soviet ideology (592). Thus, Piketty seems to offer an ideal model of society based on a realistic view of human nature, or to use Rawls's words, a “realistic utopia,” taking people as they are, not as we want them to be (Rawls 1999, 4, 5-6, 11-12). Piketty does not base his participatory socialism on the presence or development in each of us of noble, disinterested feelings of benevolence, solidarity, or fraternity, as one can find recently in Cohen (2009) or Honneth (2017).

On the second point, Piketty draws lessons from his analysis of the history of inequality regimes, rejecting ‘historical experimentation’ that did not work like state socialism, while keeping and extending what for

8 Reddy calls Piketty's participative socialism a “limited utopia,” but he does not explain what he means by it (2021, 18).
him did work: the benefits of social democracy, namely progressive taxation and the social state, or Ricoeur’s latent potentialities of history. In line with the latter, who thought that good utopias are always historically determined, because they are a reaction to one or several other ideologies of a given period and inherit from the history of utopias, Piketty tries to reinvent and revive socialism after the fall of communism in order to find egalitarian solutions to the issues of the 21st century and to continue the long-run progress of human societies towards justice and equality.

**IV.IV. Utopianism, Realism, and the Overcoming of Capitalism**

Interestingly, my presentation of Piketty’s project of society as a ‘Ricoeurian’ and ‘realistic’ utopia situates his work at the heart of currently heated debates in political theory and philosophy on the relative merits of ideal and non-ideal theory and on the possibility and/or necessity of combining them (Valentini 2012). In these debates, the seeming opposition between ideal and non-ideal theory might take the form of a distinction between idealistic or utopian theory and realistic theory. Then the discussion focuses on whether considerations of feasibility should constrain normative political theorizing, and if so, what kinds of (realistic) feasibility constraints should be considered as relevant (Valentini 2012; Gilabert and Lawford-Smith 2012). The compatibility of utopianism with realism is widely debated and some authors think that it is possible to fruitfully combine them (Raekstad 2016, 2020a), as Piketty seems to try to do. This is the case of Gilabert and Lawford-Smith who offer a model combining evaluations of desirability and feasibility, reconciling ideal and non-ideal theory, or utopianism and realism. In a general sense, “some state of affairs is feasible if there is a way we can bring it about”, now or later (Gilabert and Lawford-Smith 2012, 809). They distinguish “soft” constraints (economic, social, political, or religious constraints) and “hard” constraints (biological) (813–814). Then, they describe a three-step approach on the way evaluations of desirability and feasibility should be combined. Step 1 must focus on the formulation and definition of the main principles, and here only hard constraints should be considered (819–820). Step 2 should deal with proposals of institutions implementing the principles of step 1. Here questions of stability and soft constraints must be considered (820). Finally, step 3 must deal with strategies of political reforms leading to the realization of the institutional framework defined in step 2, or aspects of accessibility (820).
This theoretical framework can help to better understand some of the criticisms that have been levelled at Piketty’s project of society. In particular, some commentators have raised doubt about the capacity of his participatory socialism, if it was implemented, to truly ‘overcome’ capitalism, as Piketty claims (Bihr and Husson 2020; Brisset and Walraevens 2020; Witztum 2021; Zevin 2021; London 2021). It is rather seen as a reformist project, not going much beyond social democracy or egalitarian forms of property-owning democracy because it would tend mainly to redistribute private property more equally. In a sense, it would not be transformative, radical, or utopian enough for those wishing to transcend capitalism.

Of course, statements about the post-capitalist character of Piketty’s project of society might depend, at least in part, on the definition of capitalism we adopt. Piketty provides a definition which is embodied in his own framework of the reasoned history of inequality regimes. Capitalism is seen as an advanced, historically situated form of proprietarianism, emerging “in the era of heavy industry and international financial investment, that is, primarily in the second half of the nineteenth and early twentieth centuries,” and culminating in today’s “globalized digital hyper-capitalism” (Piketty 2020, 154). It is defined as “a historical movement that seeks constantly to expand the limits of private property and asset accumulation beyond traditional forms of ownership and existing state boundaries” (154). Hence Piketty’s insistence on his concept of ‘temporary’ property, or a permanent and more widespread circulation of private property that would be allowed by his redistributive measures. More generally, capitalism can be defined as a socio-economic system having some or all of the following features: (i) the bulk of the means of production is privately owned and controlled, (ii) markets are the main mechanism allocating inputs and outputs of production, (iii) there is a class division between capital owners and workers, and (iv) production is oriented toward capital accumulation and profit rather than to the satisfaction of human needs (Gilabert and O’Neill 2019). Socialism, as an alternative to capitalism, would involve, by contrast, removing at least one if not all of these features, and especially the first one (see Roemer’s market socialism in Roemer 1994, or Rawls’s liberal socialism in Rawls 1971, §42). It has traditionally been defined upon the public or social property and control of the means of production. Socialism is about extending social empowerment and democratic control over the economy (Gilabert and O’Neill 2019; Wright 2010). Now, with the definition I have given of capitalism, a

---

9 https://www.youtube.com/watch?v=dDY3aczW0d0
fairer and equal distribution of social wealth is not enough for transcending capitalism. It doesn’t seem to deeply question features (ii) and (iv), and it would merely weaken features (i) and (iii).

In Piketty’s defense, however, we can think that guaranteeing maximum equal access to fundamental goods to everyone might require that a (quite) significant part of social wealth and of the means of production be publicly owned or under social control. Moreover, his participatory socialism is meant to be an open model subject to change with the evolution of public debate and new data, in line with his view of justice as a deliberative and never-ending process animated by the democratic public sphere. Precisely, in his latest book, A Brief History of Equality, Piketty provides further thoughts and policy proposals for implementing his participatory socialism, among which one finds some motivations for the development of commons and cooperatives, and positive comments on Bernard Friot’s communist proposal of a life wage (169). He also endorses the support of a system of job guarantee (Tcherneva 2020), the strengthening of minimum income schemes, and the development of the non-profit sector (chapter 7). Finally, he pleads more explicitly for the de-commodification of society and for the collective redefinition of needs. All these elements seem more in agreement with his stated intention to transcend private property and to overcome capitalism, giving more substance to the post-capitalist, socialist character of his project.

Another type of criticism levelled at Piketty’s social project concerns what can be seen as its alleged insufficient ‘realism,’ in the sense that he does not tackle enough with the conditions of its feasibility and accessibility in today’s world (Bihr and Husson 2020; Zevin 2021; Reddy 2021). His project is thus deficient on step 3 of the process of combination of utopianism with realism sketched above, i.e., on the process leading to a socialist society, a key issue in debates about socialism. This is how we can interpret Bihr and Husson’s view of Piketty’s participatory socialism as an ‘utopian socialism’ (Bihr and Husson 2020, chapter 6; see also Motadel and Drayton 2021, 36). The word ‘utopian’ is used by them in its pejorative, Marxist sense, to mean that Piketty’s project is too idealistic, or in other words, too confident about the power of ideas to change the world, and hence, not attentive enough to the real conditions of political change or to obstacles that his project would face in the material world from those who believe in, benefit from, and defend neoproprietarian ideology. Specifically, Piketty does not really explain why or how a new majoritarian political coalition would or could form, endorsing his socialist
project and trying to implement it when in power (Zevin 2021, 76). Yet, I do think that his project on political and economic differences should be seen as one step in that direction (Piketty 2020, chapters 14 and 15; Gethin et al. 2021). It is meant to inform us better about the historical dynamics of electoral cleavages and its recent changes, and thus could be used to draw some lessons from history for reconstructing a majoritarian, socialist, or egalitarian coalition.

V. Conclusion

In this paper, I tried to show why and how Ricoeur’s rehabilitation of utopian thinking, and his innovative analysis of the dialectics of ideology and utopia might help both to shed new light on and give stronger philosophical foundations to Piketty’s view of history as the struggle of ideologies and to his alternative project of participatory socialism, which I interpreted as a ‘good’ utopia in Ricoeur’s sense. Both Ricoeur and Piketty plead for taking into account the feasibility of utopias. We must not oppose utopianism with realism, but rather try to find a proper balance between them to awaken the utopian desire and mobilize people for action. Piketty’s project allies the critical and transformative functions of utopia. It is not meant to be the final word on ideal society, but only an alternative or a ‘possible world’, which should be discussed and compared with others in the endless public debate about social justice. Presenting Piketty’s project in this way as a ‘Ricoeurian’ and ‘realistic’ utopia helps, I think, to better understand some of the criticisms levelled against him. Additionally, these discussions find an echo in current debates in normative political theory over the benefits of ideal and non-ideal theory and of their possible reconciliation, combining utopianism with realism by defining feasibility constraints that political theories should respect.

Good utopias can make useful contributions to social science and be fundamental guides for public policies. Piketty’s latest books, I argue, are a prime example of this. But they show that the reverse is also true, in the sense that social science can help to extend the frontiers of the possible and to imagine new utopias. His works dispel an unorthodox view of economics as a historical, social, political, and moral science. The historical study of inequality and of their discursive and institutional structures of justification, or inequality regimes, combining intellectual and quantitative history, allows to highlight the political and ideological foundations of inequality. More than that, it offers important lessons for building a new egalitarian socio-economic model or utopia. He repeatedly underlines
the role of statistics for fighting injustice and promoting equality (see Piketty 2022b). The development and widespread diffusion of economic data on inequalities are seen by Piketty as a means to improve democracy through citizens’ enlightenment and their appropriation of economic knowledge (Piketty 2020, 1039-1041; 2022a, 244; Piketty, Saez and Zucman 2022, 2), providing the necessary epistemic conditions of a successful deliberative and participatory democracy and of an open and inclusive public debate about social justice. As rightly noted by Langenohl, “Thomas Piketty has written a book [C&I] that is truly dedicated to the idea of economic enlightenment” (2021, 121), relying on and calling for other researchers to follow his steps in developing and using economic data to foster human emancipation. For Piketty, “more just and egalitarian societies are always possible”, but if “progress exists […] it is a struggle, and it depends above all on rational analysis of historical changes and all their consequences, positive as well as negative” (2020, 20). Under Piketty’s pen, economics becomes a critical, engaged, and emancipatory social science. At the end of his books, Piketty quite naturally pleads for a decompartmentalization of knowledge and for an interdisciplinary perspective on economics as a normative and moral science (Piketty 2014, 2020).

To conclude, in line with Ricoeur’s recommendation, Piketty’s project of participatory socialism seems to reconcile the ‘ethics of conviction’, or his socialist sympathies and his ideal of radical equality, with the ‘ethics of responsibility’ and the ‘logic of action’ by defining concrete, implementable, and realistic though ambitious proposals to advance social justice and equality, trying to find a proper balance between idealism and realism. Piketty’s participatory socialism unites the critical and transformative functions of utopia. Through his work in history and social science, he seems able to identify some latent, unfulfilled possibilities of the past and present which are necessary for building ‘good’, practical utopias as Ricoeur recommends it.

REFERENCES


Benoît Walraevens is Assistant Professor in Economics at the Université de Caen Normandie’s Center for Research in Economics and Management, Caen, France.
Contact e-mail: <benoit.walraevens@unicaen.fr>
Editorial Note

Concomitant with the surging inequality of recent decades has been the development of exploitation theory, which concerns the form and moral import of transactions between stronger and weaker parties. As capital is concentrated in fewer hands, more people are relegated to demeaning and precarious work. Sweatshops multiply. Labor migrants are denied basic rights. Surrogacy and prostitution proliferate. Such exploitative practices can be consensual and mutually beneficial—they typically are. Yet that does not seem to vindicate them. How precisely should we account for the injustice they exhibit, and how central is that injustice to the capitalist order? The focus of our special issue is Nicholas Vrousalis’ recent book, *Exploitation as Domination* (OUP, 2023), which represents a major advance in the debate over these questions.

Vrousalis’ argument shifts our focus away from the damaging, disrespectful, or distributively unequal character of exploitative relations, towards the victim’s subordination, how her free purposive agency is compromised in ways that no compensatory welfare gains can morally redeem. Exploitation is a species of domination, the wrongful control by one agent over another’s power to set and pursue their own ends. By virtue of that control, the powerful extract productive service from the vulnerable, without ever having to extend their own labor service in return. When capital goods are scarce and privately held, as under capitalism, owners of the means of production necessarily possess this dominative control. Workers are not the masters of their own purposiveness. This is a *structural* condition: the unfreedom of workers is administered not only by capital owners themselves, but through an array of mediating regulators, who endow these social relations with their systematic, durable character. And these relations are global in scope. Political communities in the rich world still dominate those in the old colonial periphery, under a pervasive order of neo-imperialism. Whatever globalization’s effect on poverty rates, it has not freed the developing world from relations of exploitation. Or so Vrousalis argues.

Our symposium on *Exploitation as Domination* includes critical notices by S.M. Love, Lucas Stanczyk, and Gulzaar Barn, along with a reply to their arguments by Vrousalis. While broadly sympathetic to the aims
and upshot of the book, the contributors each in their own way challenge important parts of its argument.

S.M. Love is concerned with the theory of justice subtending Vrousalis’ conclusions. She affirms his conclusions, but thinks they can only be sustained via a specific route. Her article works through the Neo-republican, recognitional, and Kantian approaches to domination, and argues that only the last can consistently explain the wrongmaking features of capitalist power relations. She contends that the first approach is essentially descriptive, so it fails to substantiate the claim that unfreedom is unjust. The second approach, in her view, is properly normative, but it reaches too deeply into an agents’ motivational states to support a theory of justice, as opposed to a theory of (im)moral action. Finally, she sustains that the Kantian approach succeeds in incorporating relations of freedom into a theory of public right, and thereby anchors Vrousalis’ case against capitalist exploitation on firm normative grounds.

For his part, Lukas Stanczyk argues that the Vrousalian emphasis on power over another’s purposive labor does not register salient features of exploitative relations. Something crucial is absent from the domination conception: the normative force of reciprocity. In social interactions, people have claims on another that are distinct from their claim to reciprocity as such. It is against this independent moral standard that we should assess whether, in any given cooperative context, the parties are reciprocating appropriately. Where there is a culpable reciprocity failure on that score, there you find exploitation. In an intimate partnership, for example, A can passively accept advantages flowing from B’s besotted devotion, but deal coldly with B in return. A is a taker. She exploits B, but without directing B’s purposive actions; so unilateral control over purposive action is not—contra Vrousalis—necessary for exploitation. Stanczyk marshals different arguments to show that his reciprocity conception is also to be favored over the harm-based, distributive, and respect-based accounts of exploitation.

To close off the symposium, Barn concedes that the relation between workers and capitalists is exploitative in just the way, and for the reasons, that Vrousalis identifies. But as a general theory of domination, she thinks the book fails to capture aspects of patriarchy that fall outside the category of domestic and sexual labor. The ideological norms that uphold male social power are broader than this category allows, and their effects also harm the dominators. Barn’s claim is that Vrousalis neglects this.
Moreover, Barn contends that his account of exploitation, when applied to surrogacy, is problematic. It entails that a vulnerable woman recruited into the practice of renting out her womb, exploits the infertile couple paying her to pursue their familial ends because she possesses the means of sexual reproduction. This implication, Barn maintains, is implausible.

We hope that this collection of contributions, and Vrousalis’ book, will stimulate the readers to further discuss the problems that the current economic system is entrenched in, and their putative wrongness.

These papers were presented and discussed in a lively symposium in Rotterdam in March 2023. The editors are grateful to all the participants in that event, to the authors who contributed to this special issue, and of course to Nicholas Vrousalis for having written such a stimulating book. We also wish to thank Gergana Boncheva and James Grayot for assisting us in the editorial work and the Erasmus Institute of Philosophy and Economics and the Erasmus School of Philosophy at the Erasmus University Rotterdam for supporting the *EJPE*.

TULLY RECTOR  
*Utrecht University*  
**Guest Editor**  
<tully.rector@gmail.com>

ELISABETTA GOBBO  
*Erasmus University Rotterdam*  
**Editor**  
<gobbo@esphil.eur.nl>

BENJAMIN MULLINS  
*Erasmus University Rotterdam*  
**Editor**  
<mullins@esphil.eur.nl>
The Injustice of Domination

S.M. LOVE
Georgia State University

In *Exploitation as Domination: The Injustice of Capitalism*, Nicholas Vrousalis (2023) builds on the voluminous literature criticizing capitalism, giving us his novel and insightful argument for the injustice of capitalism: not only does capitalism necessarily involve exploitation, as many have previously argued, but this exploitation necessarily constitutes domination. As Vrousalis explains, capitalist exploitation “is a dividend of servitude: the dividend the powerful extract from the servitude of the vulnerable” (1). Because domination is unjust, capitalist exploitation must also be unjust.

*Exploitation as Domination*, while modest in length, is ambitious in its scope. As a consequence, the portion articulating the injustice of domination is brief. Vrousalis articulates three possible justifications for holding that domination is unjust:

- Kantian, broadly based on the idea of a moral right to independence;
- republican, broadly based on a conception of arbitrary power; and
- recognitional, broadly based on a conception of interpersonal recognition. (43)

Here, I argue that of these three views only the Kantian view can justify Vrousalis’ argument for the injustice of exploitation, and I give a more detailed account of the injustice of domination within the Kantian framework.

---

1 Vrousalis makes it clear that his aim “is not to provide a complete justification” for the injustice of capitalist exploitation; instead, his aim is instead “to steer the conversation about the injustice of exploitation in the direction of domination theory”, an aim he certainly achieves (43).
I. VROUSALIS ON EXPLOITATION AS DOMINATION

Again, the central claim of *Exploitation as Domination* is that capitalist exploitation constitutes domination and is therefore unjust. In this section, I will lay out the essentials of this central argument.

Vrousalis gives a novel and exciting argument that exploitation, contrary to previous accounts, is best understood as domination. Domination is defined as the “subjection of purposiveness to the choices of others” (6). On this “domination view, A exploits B if A benefits from a transaction in which A dominates B: exploitation is the dividend A extracts from B’s servitude” (30). In other words, in a situation where A has power or control over B such that B is in A’s service, the benefit A draws from this servitude is exploitation.

Vrousalis applies this domination view of exploitation to work and labor, generating the “Non-Servitude Proviso”:

> For any agents or groups engaged in mandatory mutually affecting cooperation under a division of labour, and barring any special justification that exempts them, none should possess unilateral control over the labour capacity of any other. (40)

Capitalism is defined as a system of “unequal private ownership of scarce productive assets, along with a labour or credit market conferring access to those means” (101). As Vrousalis argues, by its very nature, the “abstract capitalist relation” violates the Non-Servitude Proviso: the capitalist state’s regulation of private ownership grants those who control capital, i.e., those who control scarce productive resources, “unilateral control over alien labour capacity” (105).

According to Vrousalis, all domination, including capitalist domination, is a violation of “a fundamental commitment to freedom,” where each has a right to “nonsubjection to the choices of others” (40). Vrousalis posits three alternative accounts for justifying this fundamental commitment to freedom: Kantian, republican, and recognitional. In what follows, I will explore the suitability of each account as a justification for the injustice of exploitation as domination.

II. THE NEO-REPUBLICAN VIEW

While Vrousalis does not focus on a particular republican view, I focus here on the hugely influential neo-republicanism of Philip Pettit (1997),
who, along with Quentin Skinner (2002), is largely responsible for bringing domination to the forefront of contemporary political philosophy. Neo-republicanism is characterized by a focus on the importance of freedom as non-domination: on this view, “one agent dominates another if and only if they have a certain power over that other, in particular a power to interfere on an arbitrary basis” (Pettit 1997, 52). To prevent misinterpretation, Pettit (2008, 102) has more recently articulated the power that constitutes domination in terms of control rather than arbitrariness, defining “liberty as the absence of alien or alienating control on the part of other persons”. External interference is non-alienating, and so consistent with freedom, when it is “subject to your control” (Pettit 2012, 57).

At first glance, neo-republicanism might seem like a fine ground for the arguments Vrousalis offers. Neo-republicanism offers a robust and detailed explanation of what constitutes domination, which can be drawn on to supplement the account of exploitation as domination. Further, it seems plausible that the capitalist labor relations Vrousalis criticizes would constitute domination of the neo-republican sort, as the terms of employment that govern workers are arguably not subject to their control.²

Despite these clear affinities, this neo-republican view does not support Vrousalis’ conclusions. Again, Vrousalis argues that because capitalism necessarily involves domination, it is unjust. Within this neo-republican framework, demonstrating that something constitutes domination is insufficient to establish its injustice. To begin, the core of neo-republicanism is a merely descriptive account of freedom as non-domination (Pettit 2008, 117). A descriptive view of this sort may resonate more with those who see morality, as Marx did, as mere ideology supporting the interests of the ruling class³ or with moral skeptics more generally. But this descriptive account, which is the main focus of standard neo-republicanism, does not generate claims concerning injustice. Pettit’s broader view, though, does incorporate this descriptive account into a consequentialist moral framework where “non-domination is a goal which [political] institutions should seek to promote” (Pettit 1997, 81). But even this broader consequentialist view will not support Vrousalis’ essential premise that domination per se constitutes injustice. Within this framework, there is no deontological right to non-domination.

² Those arguing for workplace democracy often rely on neo-republicanism to ground their arguments in just this way. See, for example, Anderson (2017).
³ Recently, there has been renewed interest in the republican aspects of Marx’s thought. See, for example, Roberts (2016).
It is always a contingent and empirical question whether prohibiting any particular form of domination will minimize overall domination, and not all forms of domination should be prohibited on this view. One could argue within this framework that capitalist domination will lead to a minimum of overall domination, thereby meriting its prohibition. However, not only would establishing this require considerably more empirical argument than Vrousalis provides, but also within this framework capitalist domination could only ever be contingently (and never in principle) unjust.

Furthermore, within this consequentialist framework, the reasoning given for prioritizing freedom as non-domination over all other values is thin: Pettit asserts that if non-domination is given this priority, “then most other desiderata will look after themselves” (1997, 7). In the case of capitalism in particular, popular arguments in favor of capitalism indicate that proponents of capitalism would dispute whether eliminating capitalist domination would best serve our other values, such as prosperity and welfare.

So, while one can argue that capitalist labor relations constitute domination in the neo-republican sense, successfully establishing this does not suffice to show that capitalist labor relations are unjust. For one who would argue that permitting capitalist domination would be unjust even if it led to a minimum of overall domination, a consequentialist framework will not suffice. This framework also does not establish a right to freedom that in principle overrides any other interest we may have. As I will argue, the Kantian framework I develop below offers just the sort of deontological view that can justify a fundamental right to freedom, thus offering a much more compelling foundation for Vrousalis’ arguments.

III. THE RECOGNITION VIEW

Vrousalis offers up his own recognitional explanation of the injustice of domination. This account is premised on an understanding of human beings as agents who are capable of acting for reasons. On this view, one acts freely when one “recognizes and appropriately reacts to a normative reason” (266). When it comes to interacting with other people, in order to

---

4 According to Pettit, a free person “will avoid alien control in relevant choices,” where these “relevant choices... correspond to the important liberties, however they are understood” (2008, 103). On Pettit’s own view, the list of “basic liberties” includes freedom of thought, expression, and association (2014, 72).
act freely one must take as reasons others’ “rational intentions—inten-
tions whose contents are independently justified” (44). Free agency more
broadly requires that we be generally disposed to take one another’s ra-
tional intentions as reasons in this way (Vrousalis 2020, 283). Since dom-
ination involves directing the behavior of another person for one’s own
reasons, whatever they may be, domination consists in “misrecognized or
alienated agency”, where one person has “power over” another, “power
that does not facilitate action for reasons” (275). Capitalist domination in
particular involves misrecognition in this way, as within this system, as a
worker “my productive agency, and therefore my power to produce for
reasons, is alienated to the capitalist owner and subordinated to her will”
(278–289).

This recognitional view is a valuable supplement to Vrousalis' account
of exploitation as domination. It offers a plausible and interesting view of
something that goes wrong within relationships of domination, and
within a capitalist productive system in particular. When capitalists em-
ploy workers in order to maximize profits, whatever intentions the work-
ers may have in engaging in this process of production are irrelevant to
the process. This failure to recognize the worker as a source of reasons
thus resonates with Marx's account of alienation, where the worker's lack
of agency in the production process results in a disconnect between the
worker and her work, the product of her work, those who she engages
with, and her own essence as a freely choosing being.\(^5\)

However, this recognitive view cannot easily support Vrousalis’ con-
clusion that capitalism is unjust. To start, this recognitive view, while per-
haps plausible as an ethical view, is not a plausible theory of justice, as it
cannot distinguish between mere violations of ethical duties and injus-
tices. A view that cannot make this distinction will give rise to a grossly
over-demanding conception of justice, as it will condemn as injustice a
great many actions that are not intuitively within the domain of justice.
Misrecognition happens not only when someone acts in a way that is in-
consistent with others' desires, it also happens when someone does not
act for the right reason—even when someone acts in conformity with oth-
ers’ demands, they still misrecognize them if their action is not motivated
by the fact that others have demanded it (275).\(^6\) Typically, identifying
something as an injustice involves calling for its prohibition. Prohibiting

\(^5\) For an overview of Marx’s account of alienation, see Wood (2004, 3–15).
\(^6\) It also seems consistent with this view that you could misrecognize someone merely
by failing to regard them appropriately as sources of reasons, even in the absence of any
external actions.
misrecognition would then entail prohibiting not only actions but mental states, which would be deeply problematic.  

Furthermore, the requirement that we recognize others' rational reasons as reason-giving for ourselves is not in itself opposed to capitalist domination. Workers might rationally prefer capitalist domination to its alternatives, if, for example, capitalism best fostered economic prosperity or welfare. It seems that capitalists would not misrecognize workers, then, in playing their part in the capitalist system. In order to condemn domination outright, regardless of whether one might prefer to be dominated, a separate account of inalienable individual rights would be necessary. While the recognitive account seems to presuppose that each person has “standing”, i.e., ultimate authority, over themselves and their bodies, it does not itself offer such an account of individual rights (270).

As I will argue, the Kantian account developed below does offer an account of fundamental individual rights that also distinguishes merely ethical duties from injustice, thus offering a more compelling ground for Vrousalis' arguments.

**IV. THE KANTIAN VIEW**

Here, I argue that the Kantian framework can ground Vrousalis' arguments. In what follows, I will first briefly articulate the framework of a Kantian theory of right. I will then give a brief account of domination within the Kantian framework, explaining when and why it is unjust. I will conclude by highlighting the consequences of relying on this Kantian framework for Vrousalis' account of exploitation as domination.

**IV.I. Kant's Theory of Right**

In contrast with republican and recognitional theories, this Kantian account draws a distinction between our ethical duties and our duties of right. As embodied beings who live in the world together, our external actions affect one another. While it is not possible to compel anyone to act for any particular reason or other, we can evaluate how each person’s external actions relate to others.

---

7 While any attempt to compel particular mental states would be troubling, to put it mildly, Kant also explains why compelling ethical duties (compelling doing the right thing for the right reason) is impossible. To put the point simply, acting for the right reasons, acting from duty, is something that I can only do for myself. External compulsion cannot make someone act from duty; it can only make them act from external compulsion (Kant [1797] 1996, 6:220, 6:239).
The foundation of the Kantian theory of right is the one innate right to freedom:

*Freedom* (independence from being constrained by another’s choice), insofar as it can coexist with the freedom of every other in accordance with a universal law, is the only original right belonging to every man by virtue of his humanity. (Kant [1797] 1996, 6:237)

On this view, people deserve to be respected as the rational beings that they are. As rational beings, we set ends for ourselves and pursue them at will. Since we are embodied beings that live in the world together, the choices we make in the world can affect others. Treating others as the rational beings they are means respecting their setting and pursuing their own ends. The right to freedom is thus a right to direct your own will in the world *consistently with the rights of others to do the same*. This right is not a right to do whatever you want: each person has the right to direct only their own will in the world, and this right does not include the right to direct the wills of others. Instead, this right is universalizable: since each has the right to direct only their own will and not the wills of others, each person’s right to freedom is consistent with the right to freedom of all others.

This one innate right is the foundation for all other rights within this Kantian framework. Since all rights are derived from this one innate right, there is no conflict between the different rights that flow from it. While innate rights, such as bodily rights, are determinate in virtue of the right to freedom alone, some rights are indeterminate. We must make these rights determinate together, since they structure all of our rights. And once we make these rights determinate, they are binding—once we pick a side of the road to drive on, for example, we have a duty to drive on that side of the road.

Rather than taking the role of government to be making trade-offs between conflicting rights, within this framework the role of government is simply to secure and make determinate the right to freedom and all the

---

8 In this way, the Kantian right to freedom is fundamentally distinct from a conception of freedom as negative liberty, where I am unfree to the extent that “I am prevented by others from doing what I could otherwise do” (Berlin 2002, 169).

9 Though Kant did think such rights needed to be made through an omnilateral will, he did not think this required democratic procedures. Many Kant scholars, including myself, assert that Kantian principles do entail a right to procedural democracy. See, for example, Hanisch (2016).
rights that flow from it. Unlike a conception of freedom where any interference with persons’ activity constitutes unfreedom, on this Kantian account, not all interference with the choices of others violates their right to freedom (Kant [1797] 1996, 6:231). A prohibition on assault, for example, does not violate the right to freedom: since no one has a right to assault anyone else, prohibiting engaging in this activity instead secures and is consistent with the right to freedom. So, unlike these other freedom-based views, this view carries with it no presumption that all government activity reduces freedom and so should be kept to a minimum. Instead, on this view, the state should take all and only that action that is required to secure each person’s right to freedom.

**IV.II. Domination in the Kantian Framework**

At a fundamental level, having the right to freedom means having the right to “innate equality, that is, independence from being bound by others to more than one can in turn bind them” (Kant [1797] 1996, 6:238). This right to innate equality is not a right separate from the right to freedom; instead, it is part of what it is to have the right to freedom. If each of us has the right to direct our own wills in the world, then no one has the right to direct the will of anyone else. We must all be equal under law; if the law does not treat us equally in like circumstances, then some receive better treatment under the law in contradiction with our innate equality.\(^{10}\) Having the right to freedom also means that each of us has the innate right to be our own masters, as each person having the right to direct only their own will in the world means that no one has the right to be the master of anyone else (6:239).

Equality and self-mastery can both be undermined by domination, in violation of the right to freedom. Let us define domination as an ongoing relationship between people where one person is under the control of another (or others). In order for this dominating relationship to violate the right to freedom, one must not only have control over another; this domination must be externally enforced. The right to freedom is only violated when one’s capacity for choice is externally restricted. Dominating relationships where one is fully externally free to walk away from the relationship do not violate the right to freedom. Within a society where the state has a monopoly on force, dominating relationships will violate the

\(^{10}\) This does not entail that we must all have exactly the same rights; instead, equality under law entails that different people in the same circumstances will have the same rights. This right would be violated, for example, if some were granted fair trials while others were not or if some classes were subject to invidious discrimination.
right to freedom when they have the force of law, whether *de jure* or *de facto*.\(^\text{11}\)

If one person is externally bound to another more than that other is in turn bound to her, the person who is disproportionately bound is to that extent subjugated to the other’s will. This makes the person to which they are bound to that extent their master, a condition that is inconsistent with their right to freedom. This rights-violating domination can be complete, as in the case of slavery, or partial, as in any case where some have disproportionate legal authority over others.

Domination that does not have the force of law within a state, however, will not violate the right to freedom, even if it is ethically wrong. Imagine two friends, one a kind friend who cares very deeply about the other’s opinion, and the second a mean friend who is manipulative and denigrates the other. This is a dominating relationship, as the mean friend manipulates the other into comporting herself as the mean friend desires. This situation is certainly ethically wrong, but the dominating relationship has no legal authority, as the mean friend can only dominate the kind friend so long as the kind friend chooses to stay in the relationship.\(^\text{12}\)

Since the dominated person is not externally subordinated to the other and can exit the relationship at will, her right to direct her own will in the world has not been violated.\(^\text{13}\) Domination of this sort thus does not constitute injustice within this Kantian framework.

So, within this framework, demonstrating that a particular relationship is a relationship of domination does not suffice to establish that this relationship is unjust. To establish the injustice of domination within a state, the domination must be shown to have legal force, whether *de jure* or *de facto*.

---

\(^\text{11}\) By ‘*de facto*’ here, I simply mean that even when law does not explicitly permit domination, it may in fact permit domination through less explicit laws and policies.

\(^\text{12}\) Imagine a case where the mean friend physically detained the kind friend, thus forcing their continued interaction. As I see it, the kind friend’s rights would be violated here, but not by the relationship itself; instead, the kind friend’s rights would be violated by the act of incarceration. On this Kantian view, the right to freedom can be violated by acts that are not best understood as cases of domination. This view thus resonates with Quentin Skinner’s early republican views, where he took domination to be a form, but not the only form, of political unfreedom. See Pettit (2002).

\(^\text{13}\) Note that personal relationships, such as romantic or familial relationships, can involve *de facto* legal subordination even if the laws of a given state do not formalize this subjection; this can happen when, for example, law enforcement does not adequately enforce the law in particular situations, such as domestic violence. When the family is considered by the state to be a ‘private’ domain in which the state should not interfere, this implicitly licenses patriarchal domination. On this point, see Okin (1998).
IV.III. Exploitation as Domination Within the Kantian Framework

Vrousalis’ aim is to demonstrate the injustice of capitalist exploitation by showing that this exploitation necessarily constitutes domination. As we have seen, within the Kantian framework, demonstrating the presence of domination is insufficient to establish injustice. To demonstrate the injustice of domination, we must establish that it has the force of law.

In the case of capitalist domination, this is a task that Vrousalis has already accomplished. In *Exploitation as Domination*, Vrousalis demonstrates in detail that the capital relation, encompassing the classic case of capitalist labor relations and alternative forms of capitalist exploitation, is a relationship of domination that is structured by law (2023, 138–140). Vrousalis defines structural domination as involving “a triadic relationship between powerful agent(s) P, disempowered agent(s) Q, and regulator(s) R, such that (i) P dominates Q and (ii) R regulates that domination” (98–99). It is thus a distinctive feature of structural domination that it is regulated (99). As Vrousalis insightfully explains, capitalist domination in particular exists through regulation and is thus a form of structural domination: in claiming “exclusive title to interpret, apply, and enforce private property on the part of all of its subjects”, the capitalist state establishes and enforces through law the capitalist’s domination of the worker (104). Within the Kantian framework, this legally enforced subordination of some to others violates the right to innate equality and self-mastery of each person who is subordinated and is therefore unjust.¹⁴

However, not all instances of domination referenced by Vrousalis have been shown to be externally enforced in this way. Vrousalis’ schoolyard bully, for example, who manipulates others into giving up their toys, may be a case where an ethical duty has been violated but a duty of right has not (143). More importantly, more would need to be shown in order to establish the injustice of liberal domination, which was meant to be loosely analogous to the manipulative schoolyard bully. According to Vrousalis, liberal imperialism, which involves liberal domination, is characterized by “unilateral control without direct political control or (threats of) force” (149). By definition, one aspect of domination through law is absent with liberal domination, as the dominating country does not assume direct political control of the dominated country. Despite this lack of force and direct political control, it is still possible to show that liberal imperialism has the force of law: it requires no stretch of the imagination.

¹⁴ For more on the injustice of dominating economic relationships within the Kantian framework, see Love (2020, 150).
to think (and many have argued) that global financial systems that are “the instruments of liberal imperialism”, “such as the WTO, the IMF, the World Bank, and so on”, have the force of law globally (142).

So, while incorporating Vrousalis’ arguments into the Kantian framework requires some additional argument, the Kantian framework can support the key conclusions of Exploitation as Domination. And this framework strengthens the case against capitalist domination, as it marks out capitalist domination as a form of domination that violates the rights of others.

V. CONCLUSION
In Exploitation as Domination: What Makes Capitalism Unjust, Vrousalis (2023) gives a compelling account of the dominating nature of capitalist exploitation. However, due to its ambitious scope, the book does no more than outline three theories of justice that could explain why this domination is unjust. Here, I have supplemented his account by investigating whether each of these three accounts can provide an adequate foundation for Vrousalis’ arguments. I have argued that the neo-republican and recognitional accounts cannot. I have also argued that the Kantian framework can support Vrousalis’ arguments, as it yields a principled account of why and when domination is unjust.

Exploitation as Domination is an important contribution to a broader leftist and (loosely) republican movement arguing that capitalism is unjust because it necessarily involves domination. These accounts have focused largely on arguing that capitalism necessarily involves domination. At the same time, they have focused comparatively less on arguing that domination constitutes injustice. This may be a consequence of the nature of the broader, hugely influential contemporary neo-republican project: since on this view domination is by design a nonmoral concept, the emphasis within this framework is on making nonmoral claims that certain relationships constitute domination of this sort. It is often little more than a background assumption of the view that we care about this form of unfreedom for separate moral reasons. As I have taken steps to show here, the leftist republican critique of capitalist domination can find a better foundation in a Kantian account of the injustice of domination.

15 For an overview of this movement, see Losoncz and Tóth (2021).
Capitalist domination is not merely unfreedom—it is injustice, and the Kantian account can explain why.

REFERENCES


S.M. Love is an Assistant Professor of Philosophy at Georgia State University. In this paper and in other works, Love reconstructs a Kantian theory of economic justice and rights that draws on Marxian insights into capitalism that were unavailable to Kant himself. Love earned a Ph.D. in Philosophy from the University of Pittsburgh and a J.D. from Yale Law School. Contact e-mail: <suzannemlove@gmail.com>
What Makes Exploitation Wrongful?

LUCAS STANCZYK  
*Harvard University*

**I. INTRODUCTION**

Exploitation is among the most basic ways in which people can wrong one another. Despite this, there is still no consensus on what makes it wrongful. Various philosophical accounts have been proposed, yet all of them seem to disappoint in one of two ways. They either capture some clearly wrongful instances of exploitation but fail to appreciate others. Or else they capture all wrongful exploitation but equally condemn some actions not plausibly described as wrongful. In other words, existing analyses of exploitation either under-produce or over-produce what seem to be the required moral judgments.

Nonetheless, each of the existing accounts does seem to get something right. For this reason, there is continuing disagreement concerning what makes exploitation distinctively wrongful. Indeed, the literature appears to have reached an impasse, with some tempted to abandon the search for a unified explanation. Perhaps, what makes exploitation as an act type distinctively wrongful admits of a plurality of explanations.\(^1\)

As I see it, however, this suggestion is mistaken. We should not accept a pluralist account of the central wrong-making feature of exploitation. The reason is that a compelling unified account is available after all. The account that I will articulate starts with the same basic concept of exploitation as every other account in the literature. This is the moralized concept of wrongly taking unfair advantage of others. However, my account reduces what makes this act type distinctively wrongful to a single con-

---

\(^1\) For this suggestion, see Risse and Wollner (2013, 214): “We embrace an ecumenical approach: different conceptions capture independent concerns that nonetheless all fall under the concept of exploitation. There are multiple ways of taking unfair advantage and correspondingly many reasons why it is wrong”. See also chapter 5 of Risse and Wollner (2019).
ception without remainder: to wit, we exploit people when we fail to reciprocate appropriately for the benefits we receive from others, however directly or indirectly.

A pluralist account may seem to have the advantage of better illuminating when a given act of exploitation is especially wrongful (that is to say, when the act in question exhibits a greater number of exploitation’s allegedly-many wrong-making features). My elaboration of the reciprocity account will show that this is not really an advantage. The reason is that the reciprocity account has no trouble explaining when a given act of exploitation is especially wrongful. Nor does it struggle to explain other relevant considered judgments. Instead, once it is properly understood, the reciprocity account of exploitation can be shown to capture the force of every other conception to which theorists of exploitation have been attracted, without inheriting their problems. This advantage holds even against the best and most comprehensive alternative conception currently on offer, the domination conception of exploitation elaborated by Nicholas Vrousalis (2023). Or so I shall argue.

II. TWO WAYS OF CONCEIVING REQUIREMENTS OF RECIPROCITY

My central question in this paper is what makes exploitation wrongful. In answer to this question, I will argue that to exploit people is to fail to reciprocate appropriately. Here is the argument for this thesis stated in a nutshell. The moralized concept of exploitation—wrongly taking advantage of others—can be reduced to a single conception of its central wrong-making feature, without remainder. That is because, once it is properly understood, the failure of reciprocity conception contains the essential content of every other conception to which theorists of exploitation have ever been attracted. Moreover, when properly understood, the reciprocity conception has none of the implausible implications that have been alleged against it. Accordingly, it is a failure to reciprocate appropriately, and not the violation of other moral standards, that best explains what makes the act type of exploitation distinctively wrongful.

The line of reasoning that leads to this conclusion starts by distinguishing two conceptions of reciprocity. The first conception, prevalent in the literature on exploitation, is

Reciprocity as (mere) mutual advantage. Cooperative interaction that is in some way beneficial or productive of a surplus
ought to leave everyone who is party to it better off, relative to a baseline of no cooperation.

For reasons that have been stressed by others, this conception cannot plausibly form the core of an account of what makes exploitation wrongful. Most importantly, whenever a person is in danger or otherwise in desperate need of others’ help, there will often be numerous ways of extracting grossly disproportionate benefits from her in return for the help she needs. Moreover, interacting with her in such a patently exploitative way will often leave her better off than if she had simply been abandoned to her fate. And yet, the mutually beneficial nature of such transactions does not make them any less exploitative or wrong. Hence a failure to honor reciprocity conceived as mere mutual advantage cannot plausibly account for the wrongful nature of many paradigmatic instances of exploitation.\(^2\)

There is, however, another distinct conception of reciprocity, which has been mostly overlooked in the philosophical literature on exploitation. This is the conception of

\[Reciprocity \text{ as mutual advantage relative to an appropriate benchmark.}\]

Productive cooperative activity ought to benefit everyone relative to an appropriate benchmark, as given by the claims that people have independently of the demand for reciprocity.

To illustrate this idea, consider the conception of reciprocity articulated by Rawls in *A Theory of Justice* (1999). For Rawls, the relevant benchmark, relative to which social cooperation productive of a surplus must be mutually beneficial, is an equal share of the primary goods of social life. That is to say, mutual advantage above the benchmark of equality is both necessary and sufficient to avoid failures of reciprocity. Why is this the case, according to Rawls? It is because in Rawls’ theory of justice as fairness, persons are assumed to have a prior, free-standing, and reciprocity-independent claim to be considered equals: in particular, equal citizens and equal participants in social cooperation. What this means is that, in arranging productive cooperation, no one has a prior claim to benefit more than others from the rules that ultimately enable it. In light of this prior independent claim to equality, the appropriate benchmark relative to

---

\(^2\) This outcome should not be surprising. A conception of reciprocity as mere mutual advantage is congenial even to the egoist. It is therefore unlikely that this conception will turn out to be fundamental to any important part of morality or justice.
which productive cooperation must be mutually beneficial in order to avoid violating reciprocity is an equal distribution of the primary goods of social life. It is precisely this benchmark that is expressed in Rawls’ difference principle, which he therefore rightly calls a principle of reciprocity appropriate for a democratic society (2001, 61–65).

However, other accounts of the appropriate benchmark for a conception of reciprocity are also possible, some less and others more egalitarian than the difference principle. The point I wish to make here is simply this. Any plausible conception of reciprocity will take for granted that productive cooperation must be mutually beneficial relative to an appropriate benchmark, as given by the other claims that people have independently of their claim to reciprocity. Whether these claims are understood to be claims to certain (for example, equal) shares, or claims to certain forms of help, or claims to something else will depend on the given context, as well as on the content of our wider moral theory.

With this distinction in hand, I will now argue that a moralized conception of exploitation as failure of reciprocity, given an appropriate benchmark will no longer under-produce or over-produce judgments of wrongful exploitation. If I am right about this, then there is a single unified account of the distinctive wrong of exploitation.

III. THE EXPLANATORY POWER OF AN APPROPRIATE RECIPROCITY CONCEPTION

According to critics of the reciprocity conception of exploitation, the absence of reciprocity cannot suffice to make an action or practice an instance of wrongful exploitation because the absence of reciprocity is central to various innocuous social practices. For one thing, note the critics, failures of reciprocity are arguably essential to acts of gift-giving. Yet it seems false to say that accepting a gift without offering one in return is always objectionably exploitative. Similarly, redistributive taxation sometimes assists people who are thoroughly helpless and cannot reciprocate. And yet it would be a mistake to deem redistributive taxation exploitative for this reason. Critics conclude that the reciprocity account of exploitation must be wrong because it is over-inclusive.

This inference is mistaken. It follows only from an implausible conception of reciprocity as mere mutual advantage. However, if exploitation is failure of reciprocity as I have defined it, then neither un reciprocated redistributive taxation nor the practice of gift-giving need be exploitative.

To see this, begin with gift-giving. 1. When a proper gift is made, there is only a unilateral transfer—not cooperation productive of a surplus. 2.
Moreover, if it really is, morally-speaking, a gift, as opposed to something owed, then, in the context, there are no relevant background claims defining the benchmark against which everyone must benefit, on pain of failing to satisfy a moral requirement. 3. Of course, a gift can create a debt, or obligation to reciprocate, if it creates a surplus—for example, a fertile relationship of trust or sympathy—from which both parties can benefit in the future. However, in that case, the gift is no longer simply a gift. 4. So, I conclude, when the situation is one entirely of gift, the standard of reciprocity does not apply. But, when the situation goes beyond gift to the case of something owed in the context of continuous productive interaction over time, it is the idea of reciprocity relative to an appropriate benchmark that ultimately explains why.

The case of unreciprocated compelled assistance through the tax system is simpler. 1. In a just society, able persons have free-standing and enforceable duties to help those who could not survive without the labor of others. These duties hold quite independently of anyone’s claims to reciprocity. So, when a just society deploys redistributive taxation to help people who cannot help themselves, it merely institutionalizes a duty of assistance that holds on other grounds. 2. At the same time, if a recipient of social assistance truly cannot reciprocate, then it is not the case that she is morally required to do so. Ought implies can, after all. So, once again, the demands of reciprocity do not apply, and the failure to reciprocate is neither here nor there.

Let me now continue with this line of reasoning to explain why the reciprocity conception, when properly understood, does not under-produce judgments of wrongful exploitation. On the contrary, I will argue that the reciprocity conception contains the essential content of every other conception to which theorists of exploitation have been attracted: namely, exploitation as failure of beneficence and/or a duty not to harm, exploitation as instrumentalizing vulnerability, exploitation as involuntary or forced transfer, exploitation as distributive injustice, exploitation as rights violation, and finally, exploitation as failure of respect, and, in particular, as ‘dividend of servitude’ or form of domination.

---

3 For a defense, see Buchanan (1984). For a comprehensive discussion of all of these views, see Vrousalis (2023, 11–34 and 66–90).
4 For variants of this view, see Goodin (1985) and Wood (1995).
5 For examples, see Reiman (1987) and Peffer (1990).
6 See for example Roemer (1985) and Reiff (2013).
7 See for example Steiner (1984).
Exploitation as failure of beneficence
Consider first the idea that what makes exploitation wrongful is a failure of the duty of beneficence. 1. The reciprocity conception contains the (relevant) content of this conception because to engage in mutually beneficial conduct with others, relative to an appropriate baseline, is, *eo ipso*, to *benefit* them. 2. At the same time, when a duty of beneficence is *also* in play—for example, in the standard Rescue Case⁹—then, on the exploitation-as-failure-of-reciprocity conception, we can criticize your exploitative conduct toward a disaster victim for not one but two reasons. First, you haven’t done your duty to help a person in need. In addition, however, you have exploited a person by failing to appropriately reciprocate for the interaction-generated benefits that *you* received. Accordingly, there is no reason to replace the reciprocity account with the beneficence view.

Exploitation as instrumentalizing vulnerability
For a similar reason, we should reject the idea that what makes exploitation wrongful is the mere fact of making use of other people’s vulnerability. After all, towing broken cars or performing surgery for a living is not in general impermissible.¹⁰ What matters is whether all parties benefit from the service, relative to a baseline of how they ought to have been treated prior to any further cooperative interaction productive of a surplus.

Exploitation as harming others opportunistically
What about the idea that exploitation is, rather than a failure to reciprocate, a matter of opportunistically harming people? The first thing to note is that it is far from clear that every opportunistic way of setting back another person’s interests is morally objectionable. For example, it may sometimes be permissible to invent a sports technology that can outcompete one’s rivals only to require payment from its would-be users later. However, the second thing to note is that it is the reciprocity conception that explains why such a judgment is correct, if and when it is: namely, because the benchmark giving the set of claims relative to which people must be made better off does not include the claim not to be outcompeted

---

⁹ “Pit—A and B are alone in the desert. A finds B lying in the bottom of a pit. A offers B costless rescue, on condition that B works for A for $1/day for the rest of her life. B accepts” (Vrousalis 2023, 15).

¹⁰ Arneson (2016, 10). For discussion, see Vrousalis (2023, 81–82).
in the relevant domain. When that is so, creating and exchanging useful things for compensation need not be objectionably exploitative.

**Exploitation as violation of distributive justice**

What about the idea that what makes exploitation wrongful is the violation of a patterned principle of distributive justice? 1. Once again, the reciprocity conception contains the essential content of this conception. For the benchmark against which productive cooperation must leave everyone better off on the reciprocity conception can be *given* partly by a patterned idea of distributive justice. Recall for instance the conception of reciprocity embodied in Rawls’ difference principle. According to one formulation of this principle, mutual advantage above the benchmark of distributive equality is both necessary and sufficient to avoid failures of reciprocity.

2. At the same time, the reciprocity conception succeeds where the distributive justice conception fails. For it is perhaps true that—at t₁—the hard-working Ant, who has collected ample resources for the winter, is not unjustly better off than the happy-go-lucky Grasshopper who spent the summer living carefree. However, when in light of Grasshopper’s vulnerability, Ant entices Grasshopper into productive cooperation on highly skewed terms at t₂, then, depending on the wider truth about distributive justice, Grasshopper may well have a claim to be viewed as an equal participant to that *subsequent* cooperation. In addition, in sufficiently dire circumstances, Grasshopper may have a claim to be assisted in some material way prior to and independently of this forward-looking, reciprocity-related claim. If so, then Grasshopper can be exploited, just as long as his cooperative venture with Ant leaves him no better off relative to the benchmark of how he should have been treated prior to being inducted into productive cooperation at t₂. I conclude that, when properly understood, the reciprocity conception of exploitation explains everything captured by the idea of exploitation as distributive injustice and more.

**Exploitation as failure of respect**

What about the idea that exploitation is at bottom a failure of respect? Here, we ought to distinguish treating people with respect or ‘being respectful’—which is merely one of the claims that people can have on us—and respecting or honoring people’s moral claims, whatever these are. However, if we understand the idea of exploitation-as-failure-of-respect in the latter, more capacious terms, then the reciprocity conception that
I have outlined in this paper will capture the central thrust of this abstract idea, too. After all, to exploit someone on the reciprocity conception is to fail to respect their moral claim a) to benefit from productive cooperation, b) relative to an appropriate benchmark as given by their reciprocity-independent claims. Moreover, if we replace the word ‘claims’ with the word ‘rights’ in the sentence immediately above, then a similar conclusion will apply to the notion that what makes exploitation wrongful is the violation of people’s rights, whatever these are. In short, the reciprocity conception contains the central thrust of this abstract idea as well as the abstract idea that exploitation is a failure of respect.

Now, theorists of exploitation have been attracted to two other conceptions that are closely connected: exploitation as involuntary or forced transfer, and exploitation as the product of servitude or domination. What makes these conceptions especially interesting is that they can likewise claim to capture the thrust of various competing ideas. Thus, people who force others to work for them, or make them their servants, typically also harm their victims opportunistically, fail to provide them morally required assistance, take advantage of their victim’s vulnerability, fail to benefit their victim’s appropriately, and fail to respect the moral claims and rights of others. Notwithstanding these observations, I believe that these conceptions fail to capture the central wrong-making feature of exploitation, and for the same basic reasons.

**Exploitation as forced transfer**

To see this, consider first the idea of exploitation as involuntary or forced transfer. All around the world, millions of workers have been left with no reasonable alternative but to accept a needlessly dangerous job in exchange for appallingly low wages. These workers all surely count as exploited. Moreover, it does not matter who or what made it case that they have no acceptable alternative to laboring under these conditions. Perhaps it was the callous indifference of the so-called welfare state, or perhaps it was an unscrupulous sweatshop owner using violent threats or worse. Either way, countless workers around the world are forced to serve the will of some employer or another, on pain of failing to meet their own and their dependents’ basic needs. And yet it would be a mistake to infer that workers are exploited *just because* they are forced to do so.

1. After all, as I have already mentioned, it is not in general impermissible for a society to ask every able adult person to contribute their time and effort to meeting the set of social tasks that are properly regarded as
shared responsibilities—tasks such as educating the young, treating the sick, caring for the elderly, and so on. Moreover, since every able person has a fundamental interest in being free to choose from a range of occupations, it can be appropriate for the state to implement this contributory duty indirectly, through the design of the tax system, redistributive institutions, and other elements of the basic structure that will together determine the labor allocation choices that will realistically be left open to people. In evaluating a society’s economic structure, then, the central question is not whether anyone will ever be made to contribute to any social purpose whatsoever, nor whether anyone will ever be left with no reasonable alternative but to contribute. Instead, the question is whether in making provision for meeting the properly shared tasks of citizenship and apportioning fairly the burdens of discharging these responsibilities, the basic economic and social structure a) gives due weight to the free-standing claims of able persons as well as the claims of people who need the labor of others, b) while also ensuring that everyone benefits appropriately from mutually productive cooperation over and above this baseline. In short, it is not the case that having to contribute is sufficient for being a victim of exploitation.

2. However, being forced to contribute disproportionately is also not necessary for exploitation. To see this, one need only think of exploitative intimate relationships. With enough selflessness, stubborn loyalty, and blind love that goes unrequited, a person can be led to take on the lion’s share of the emotional labor and financial cost of maintaining an intimate relationship simply by falling for an unfeeling partner who knowingly takes advantage of them for this very purpose. To exploit such a selfless loving person, no one needs to be forced to do anything, if what this means is lacking any reasonable alternative. All that is necessary is that the caring person come to care deeply about the uncaring person, and that this attitude is not manifestly and appropriately reciprocated. I suggest that only the reciprocity conception can make sense of this—perhaps the most common and most hurtful—form of interpersonal exploitation.

**Exploitation as domination**
Consider finally the view that exploitation is a form of servitude or domination. On Vrousalis’ original articulation, domination consists in the unilateral direction of another person’s purposive activity or labor, and is, for this reason, pro tanto wrongful. On this view, exploitation is best understood as a dividend of servitude, consisting of the unreciprocated
labor that results when the exploitative desires of a dominator are actualized. Can this conception better capture the reality of exploitative intimate relationships? I suggest that the answer is ‘no’, because the view is subject to a dilemma that is unlikely to be resolved.

To see the dilemma, consider once more the exploited, deeply caring person and ask the following question. Must it be true of every such person that her will is ‘unilaterally directed’ by the partner that she cares for? Or is it possible to take advantage of a person’s love and selfless character without ever making her do anything? It seems to me that this is eminently possible. Intimate exploitation does not require unilaterally directing the selfless partner. On the contrary, all that is required is a) letting her act on her loving generous desires b) while also failing ever to reciprocate appropriately.

If this is right, however, then the domination account faces a dilemma. It can admit that the loving labors of an exploited party to an intimate relationship need not be unilaterally directed by her selfish partner. But in that case, it will have failed to capture one of the paradigmatic forms of interpersonal exploitation. Alternatively, the account can be revised to say that whenever a person wants and labors for the approval of someone else, their purposive activity should be seen as ‘unilaterally directed’ by an alien will, at least until a procedure is established to allocate the approval through an appropriately omni-lateral will. In this case, however, an implication of the account will be that exploitation and servitude are omnipresent in human life, and likely to be permanently so.

The explanatory power of the reciprocity account
As I see it, there is a better account of exploitation, with similar explanatory power and none of the revisionary implications. On this account, to exploit someone is to fail to reciprocate appropriately. Meanwhile, forcing someone into a productive cooperative exchange, or dominating them, is one standard means of exploiting them. Indeed, on the reciprocity account, it is possible to say considerably more than this. Namely, the presence of power and vulnerability is very often what makes exploitation-as-
failure-of-reciprocity possible. Moreover, exercising unaccountable power over people in a successful effort to exploit them constitutes a failure of respect in part because it is exploitative. And yet, people can count as exploited even if they are not forced to participate in their own exploitation, and even if they are not dominated—just in case they do not benefit from productive cooperative activity relative to an appropriate benchmark setting out how they should be treated independently of the demands of reciprocity.

IV. THE REAL BUT LIMITED SIGNIFICANCE OF EXPLOITATION

Let me now summarize what I think follows from these observations. First, it is not the case that there is a variety of ways in which exploitation can be wrong. On the contrary, the distinctive wrongness of this act type consists in the failure to reciprocate appropriately.

However, this is not to say particular acts that are exploitative cannot be wrong in several different ways. For example, if I get you to work in my sweatshop at gunpoint, then this act is wrong, and it is wrong for several reasons. One reason is that I fail to help you—when really I should. Another reason is that I wrongly force you to work for me—when I shouldn’t force you to do anything. A third reason is that, by forcing you to work for me, I exploit you, because I benefit from our productive cooperation at your expense, relative to an appropriate benchmark defined by all of your other claims.

And yet, even though the act of getting you to work in my sweatshop at gunpoint can be, and indeed is, especially wrongful, it can be especially wrongful other than because it is (especially) exploitative: perhaps I pay you just below an equal share of the proceeds of our labor. This particular act will then be especially wrongful for other reasons—in this case, because it involves placing particularly obscene restrictions on your liberty. Its wrongness as an instance of exploitation will be, by contrast, comparatively modest.

More generally, many seriously-wrongful exploitative acts—from robbing people at gunpoint to holding wages down by threatening capital flight—are simultaneously act tokens of several distinct wrongful act types. Such acts can therefore be doubly or triply wrongful for reasons besides the fact that they are exploitative. For instance, such acts can exploit by means of wrongly threatening the use of deadly force, or exploit while dominating workers when they ought to have been given a vote. In
such cases, the exploitative acts at issue are seriously wrongful for a plurality of reasons. Accordingly, there is no need to say that every action that is exploitative is wrongful only or even primarily because it is exploitative. Nor should we say that actually existing capitalism is unjust only or primarily because is exploitative. Economic structures that are rightly characterized as exploitative can simultaneously be even more seriously unjust for other reasons.

Having said this, it is perfectly appropriate to focus on identifying modes of exploitation when evaluating existing and possible economic systems. After all, as I have argued in this essay, economic systems that enable exploitation thereby countenance some people wrongly taking unfair advantage of other people’s labor. Moreover, economic systems such as slavery, feudalism, and capitalism typically enable exploitation by making workers dependent on socially much more privileged actors. Thus, it will normally also be appropriate to focus on how institutionalized relations of dependence facilitate exploitation understood as unreciprocated labor flow. Indeed, in performing such analyses, theorists of domination can even help themselves to the reciprocity conception of what makes structurally mediated exploitation wrongful: to wit, the social failure to institutionalize appropriate relations of reciprocity. The one thing that theorists of domination ought not to say, however, is that exploitation is wrongful because it is always and everywhere a form of servitude. For servitude is distinct from and goes beyond the central wrong of exploitation.

REFERENCES


The subtitle of Vrousalis’ *Exploitation as Domination* is *What Makes Capitalism Unjust*. The book argues that exploitation (conceived as domination) is what makes capitalism unjust.


Lucas Stanczyk is Assistant Professor of Philosophy at Harvard University. He works on topics in political philosophy. Contact e-mail: <stanczyk@fas.harvard.edu>
Exploitation seems to evade straightforward definition. Disagreements abound in the pursuit of necessary and sufficient conditions that determine when an exploitation charge is deemed valid. Even if a transaction satisfies an exploitation condition, Alan Wertheimer’s (1992) influential account separates an admission of exploitation from its moral force, or impermissibility. According to Wertheimer, exploitative transactions may be *mutually beneficial* rather than net *harmful* to the exploited party, in which case there is a strong presumption for permitting them (213). In ordinary language, however, exploitation is often used in a pejorative sense as an indictment of social arrangements. One might think that an important desideratum for an account of exploitation is that it coheres with such usage. That is, a theory of exploitation ought to reflect exploitation’s meaning as a political term, in the sense that it is used to condemn seemingly unjust forms of treatment. A plausible theory should therefore capture what are generally understood to be paradigmatic cases of exploitation in the public imagination. In addition, one might posit that exploitation ought to be defined as inherently wrongful in a manner that requires redress, for it to cohere with the moral force with which it is used in general parlance. Nicholas Vrousalis’ (2023) original account of exploitation successfully satisfies these desiderata. His theory also attends to the debate between force-inclusive accounts of exploitation and those that posit exploitation as necessarily consensual, offering an explanation as to why exploitation is wrongful in the absence of force.

Vrousalis argues that exploitation is unjust, or wrongful, because it constitutes enrichment through the domination of others. The wrong of exploitation does not consist in harm, coercion, or unfairness, but rather

---

1 For instance, one does not need to be a philosopher to use the term exploitation in reference to sweat shop labour and be on to something, and a plausible account of exploitation ought to reflect this.
in *domination*: exploitation is a dividend of servitude. In this paper, I suggest that while Vrousalis’ account provides a compelling story of why capitalist labour relations are unjustly exploitative, difficulties arise in its application to other cases. I begin by querying Vrousalis’ definition of patriarchy, using it to highlight what is missing in his understanding of the wrongfulness of domination. Following this, I turn to explore Vrousalis’ examination of surrogacy as a paradigmatic instance of domination-based exploitation. I suggest that his account struggles to capture the exploitation involved in surrogacy, and inadvertently renders surrogates the dominating party insofar as they control the means of (re)production.

I. DOMINATION IS WHAT IS WRONG WITH EXPLOITATION

Exploitation is unjust because it constitutes domination-induced unilateral servitude to others (Vrousalis 2023). Thus, Vrousalis’ definition of exploitation invites the following questions: What do domination and unilateral servitude consist in? Why, precisely, is domination wrongful? Vrousalis provides the following answers: domination is the subjection of purposiveness to the (arbitrary) choices of others (43). Purposiveness might be understood as agency or the ability to set and pursue ends (38). Unilateral servitude involves asymmetrical control, in this case of labour capacity: A possesses unilateral control over B’s labour capacity when A possesses control over the content, intensity, or duration of B’s labour process which B does not possess over A’s (41). Such control over the labour capacity of others is presumptively unjust because “there is always a presumption against unilateral residual control over purposiveness and its conditions” (41). Vrousalis discusses three possible justifications for his Non-Servitude Proviso, which could thus account for the wrongfulness of domination.

The Kantian argument for the Proviso asserts that all humans have an innate moral right to external freedom. Domination, by contrast, consists in my thwarting your agential purposiveness, subjecting your ability to set and pursue ends to mine. Similarly, the republican argument for non-dominination states that unfreedom consists in subjection to arbitrary power: “all power over others that is not compelled to track their judgements, interests, or goals is arbitrary” (44), and is thus, objectionable. The

---

2 Non-Servitude Proviso—For any agents or groups engaged in mandatory mutually affecting cooperation under a division of labour, and barring any special justification that exempts them, none should possess unilateral control over the labour capacity of any other. See Vrousalis (2023).
recognition argument for the Proviso focuses on free recognitive agency, which consists in the ability to act by having your rational intentions taken by others as reasons. Domination is wrongful in so far as it undermines this recognitive agency. In outlining these accounts, Vrousalis does not seek to provide a complete justification for the Proviso, but rather seeks to “steer the conversation about the injustice of exploitation in the direction of domination theory” (43). This could be interpreted as him grounding the injustice of exploitation in the injustice or wrongfulness of domination. It is therefore worth considering whether the wrongfulness of domination is appropriately captured by these three accounts.

While it is clear that Vrousalis’ list of possible justifications for non-domination is not intended to be exhaustive, there is an important omission. Domination as a concept is in part philosophically interesting due to its relational properties. That is, domination speaks to a hierarchical relationship between dominating A and dominated B. The accounts presented above focus solely on why this relationship is wrong for B, making focal B’s demands to agency and freedom from arbitrary power. Yet these accounts say nothing about why domination might be wrong for A. Why might subjecting someone else’s agency to their own, or holding arbitrary power, be bad for A? And how is this relevant to the wrongfulness of domination? I will turn to the importance of these considerations next.

II. What Is Wrong with Patriarchal Domination?

It is notable that Vrousalis invokes patriarchy as a key instance of domination, which is defined as involving coercive control, by men, over the (sexual) labour of women. Individual women can apparently meaningfully exit this coercive control by avoiding men, but “for as long as the role of women within the (legal) description of the family remains subordinate to the role of men, women as a whole remain subjected to the choices of men” (42). Fleshing this out further, patriarchal norms “make women unfree by making them either abstain from familial relations altogether or participate on men’s terms” (99). This plausibly refers to the coercive control men have over women’s domestic and reproductive labour, according to Vrousalis’ definition of patriarchy. Vrousalis continues that “when pornographic norms restrict access to empowering sexual possibilities for women, women have to choose between nonsex and sex in conformity

---

3 The methods Vrousalis suggests are divorcing or ‘retaining’ the socioeconomic where-withal to never marry.
with the model of male fantasy” (99). This is plausibly what Vrousalis means by male coercive control over women’s ‘(sexual) labour’. 

Patriarchy has been too narrowly conceived in Vrousalis’ account in terms of control over women’s domestic and sexual labour. While this definition might be consistent with the focus and contour of Vrousalis’ servitude-oriented framework, it is insufficiently attentive to the expansive reach of patriarchal norms. The gap this definition leaves has notable implications for understanding domination. The nuclear family unit (wherein women’s domestic and sexual labour is largely but not exclusively situated) may be a chief institution of the patriarchy, but the patriarchy is not confined to this realm. Rather, it is evident that our society is a patriarchy:

If one recalls that the military, industry, technology, universities, science, political office, and finance—in short, every avenue of power within the society, including the coercive force of the police, is entirely in male hands. (Millett 1970, 25)

An increase in formal equality notwithstanding, much of the same holds true since Kate Millet first made this claim in 1970. One way that patriarchy has been upheld is through its systematised conditioning towards an ideology. That is, sexual politics are upheld “through the socialization of both sexes to basic patriarchal policies with regard to temperament, role, and status” (26). This interiorisation of ideology has effects on the comportment of both sexes.

The systemic way in which patriarchy maintains gender norms, more broadly conceived, is arguably one of its key features. Restrictive ideas about masculinity and femininity abound in our culture, and these construct the development of both men and women, inhibiting self-actualisation. The gendering of characteristics such as capability, assertiveness, strength, vulnerability through norms and patterns of behaviour, seems to speak to a phenomenon that is extraneous to male coercive control of women’s (sexual) labour. Further, to understand pornographic norms and the choice between sex and nonsex as ‘male control of women’s sexual labour’ as Vrousalis does, neglects the relational aspect of sex, and indeed of domination. Pornographic norms also inhibit the sexual development of men. Under such ideology, male sexuality is constructed along the lines
of experiencing pleasure in domination (Dworkin 1981). This is, in some real sense, harmful to the dominating party *qua* dominator. One sense in which it is harmful is because it is limiting; being engaged in a culturally enforced domination relation precludes meaningful connection. Further, the general corruption of human relationships through noxious sexual politics will harm men as participants too.

Wider conceptions of patriarchy can account for this psychic damage to men. Similarly, Vrousalis' account of patriarchy could be extended to account for oppressive gender socialisation more broadly. In a footnote, Vrousalis notes that “patriarchy is the systematic disempowerment of women with respect to their ability to self-direct their free purposiveness” (99). It would seem that under patriarchal socialisation men suffer a similar such fate: the internalisation of harmful norms is poised to undermine their ability to self-direct free purposiveness. If not, then Vrousalis cannot account for the manner in which internalised patriarchal ideology harms men through domestic and sexual norms. Either Vrousalis needs a discriminating reason as to why only women’s purposiveness is undermined by internalised gender norms, or he must concede that men’s purposiveness is similarly undermined. If this is so—that a dominating relationship seems to have such an effect on both the dominant and subordinate party—what does this tell us about the nature and wrongfulness of domination? While this harm to men might not constitute domination, it may still be reasonably considered under the purview of justice, requiring redress.

This discussion suggests that the relational nature of domination is important to understanding its wrongfulness. If it is correct that domination also harms the dominator, an account of why domination is bad as a relation and bad for the dominator seems essential to understanding and arguing for its wrongfulness. This is so far absent from the three reasons for the Proviso that Vrousalis supplies. Being the dominant party in a patriarchy thwarts development and undermines self-actualisation, as all

---

4 Following Dworkin (1981), I am concerned here primarily with heterosexual relations, but norms surrounding masculinity and femininity and their corresponding demands of domination and submission are not the preserve of heterosexuality.

5 To be clear, I do not intend to argue that because men’s purposiveness is similarly undermined that they are also *dominated* by patriarchy. Rather, patriarchy can be further fleshed out as relative group-based advantage and disadvantage, positing men as the dominant party. However, insofar as Vrousalis’ definition of domination focuses on control of purposiveness as the feature that maintains the relation, it needs to reckon with the idea that men’s purposiveness is also undermined. Thus, a theory of domination might also have to account for the dominating party’s undermined purposiveness and the wrongfulness therein.
bear the demands of the ideology’s norms. It is unclear whether this applies to the dominating capitalist (is your boss’ opportunity for self-actualisation similarly thwarted because they dominate you...?)

which attests to the aforementioned difficulty in providing a definition of domination that applies universally or to diverging phenomena. At the very least, however, it seems that patriarchy is a key illustrative instance of domination for Vrousalis and does some important conceptual work. If he wants to retain patriarchy as an illuminating case of domination, then something needs to be said about domination’s wrongful relational harm on both sides of the relation.

This more expansive conceptualisation of domination’s wrongfulness might be illuminated by Aimé Césaire’s (1950) contention that colonisation harms the coloniser:

Colonization works to decivilize the colonizer, to brutalize him in the true sense of the word, to degrade him, to awaken him to buried instincts, to covetousness, violence, race hatred, and moral relativism; and we must show that each time a head is cut off or an eye put out in Vietnam and in France they accept the fact, each time a little girl is raped and in France they accept the fact, each time a Madagascan is tortured and in France they accept the fact, civilization acquires another dead weight, a universal regression takes place, a gangrene sets in, a center of infection begins to spread; and that at the end of all these treaties that have been violated, all these lies that have been propagated, all these punitive expeditions that have been tolerated, all these prisoners who have been tied up and ‘interrogated’, all these patriots who have been tortured, at the end of all the racial pride that has been encouraged, all the boastfulness that has been displayed, a poison has been distilled into the veins of Europe and, slowly but surely, the continent proceeds toward savagery. (35)

Domination instils norms of comportment upon the dominator. Consistently treating others as means to one’s own ends plausibly has some reflexive effect. Lauding power over the powerless creates patterns of thought and behaviour. If not the wholesale brutalisation described by Césaire, it seems plausible that at least some thwarting of the dominator’s self-actualisation occurs. This effect requires consideration in an account of why domination is wrongful. Indeed, it may turn out to be focal to domination’s wrongfulness, but we will not know that until we take the
argument where it goes. I hope to have shown that this aspect of domination demands attention, much more so than I can give it here.

### III. Surrogacy and Exploitation

Commercial surrogacy arrangements⁶ are often analysed through the lens of exploitation. Dominant philosophical analyses of commercial surrogacy have focused on the practice in abstraction, pointing to the potential for exploitation under certain circumstances where these ‘certain circumstances’ have been hypothetical. In considering whether commercial surrogacy should be banned if deemed exploitative, Alan Wertheimer (1992) questions whether such a ban would actually have egalitarian consequences, as “we have no empirical evidence to support the claim that the permissibility of surrogacy would reinforce or perpetuate social inequalities” (238). Wertheimer’s statement appears less plausible now that we have a wealth of empirical evidence to suggest that countries in the Global South have become surrogacy hotspots due to their ability to provide a ready supply of surrogates at a significantly lower cost than the West.

For Vrousalis, commercial surrogacy is exploitative because it:

- Involves unilateral alien control over a woman’s reproductive capacity. Since that capacity is assimilable to productive purposiveness, as I have defined it, it comes under the purview of the Non-Servitude Proviso; contract pregnancy thereby violates it. And since the domination of the surrogate mother involves unilateral labour flow, in the form of her gestational care for the embryo, surrogate motherhood is exploitative. There is, finally, a conditional version of this argument: if contract pregnancy is exploitative, then so is organ donation for money. Humans are, in a relevant sense, ‘pregnant’ with their own organs. (81)

In congruence with the rest of his account of exploitation, therefore, surrogacy is exploitative because it extracts a dividend from servitude. Servitude consists in unilateral alien control over purposiveness. To flesh

---

⁶ Commercial surrogacy involves a financial contract between a surrogate and the intended parent(s) to carry a baby to term. In traditional surrogacy, the surrogate’s egg is used, making her the genetic mother. In gestational surrogacy, the surrogate has no genetic relationship with the child, since both gametes come from the commissioning couple or are sourced elsewhere, with assisted reproduction techniques used to implant the resulting embryo.
this out further, let us consider the example provided by Vrousalis earlier on in his discussion:

To take a concrete example, suppose that omelettes are the only means of consumption and that, if I am to nourish myself, I must produce an omelette. As an omelette consumer I have set myself an end; as an omelette producer I must use means to pursue and fulfil that end. Now, as long as you and others own the eggs, I can produce the omelette only by your permission. This makes my ability to set and pursue the end of omelette production—my productive purposiveness—subordinate to your unilateral will. When that subordination is expressed in extraction of unilateral labour flow from me, you exploit me. Omelette redistribution will not solve this problem, insofar as it leaves the mode of omelette production untouched, no matter how much it ameliorates the mode of omelette distribution in my favour. Egalitarian egg predistribution does better, but still subordinates my omelette production to your will, insofar as it does not preclude your ownership or control of the omelette-producing cookshop. And again, when that control eventuates in unilateral labour flow, you exploit me. (38)

This account, however, is in danger of rendering the surrogate the exploitative party, which would be perverse. Consider: as consumers, the infertile couple have set themselves the end of producing offspring. As producers, they must use means to pursue and fulfil that end. The surrogate owns the means of (re)production that the infertile couple do not. As long as the surrogate (and fertile others) own this, while the infertile couple do not, the infertile couple can produce offspring only by the surrogate’s permission. This makes the couple’s ability to set and pursue the end of offspring production—their productive purposiveness—subordinate to the surrogate’s unilateral will.7 When that subordination is expressed in extraction of unilateral labour flow from the couple (such as the thousands paid in surrogacy fees), the surrogate exploits the infertile couple. As long as production is in the surrogate’s hands, and when that control

---

7 It does not matter that there are other individual surrogate women who the couple could approach if they did not want to be subject to that individual surrogate’s unilateral will. Rather, they are subject to the will of ‘fertile women who are willing to act as surrogates’, as a class, just as the individual in Vrousalis’ example is subject to the will of egg owners as a class.
eventuates in unilateral labour flow; i.e., whenever she charges others for access to the means of production, she exploits them.

It might be responded that this does not constitute the extraction of ‘unilateral labour flow’. This term is rather nebulous, but we can see if this example would constitute unilateral labour flow in reference to the open ways in which it is defined by Vrousalis. It is sometimes cashed out in terms of unreciprocated effort (69), or surplus labour (72). The surrogate of course also undertakes labour, in the dual context of specific contracted reproductive labour, and in generally maintaining her organs and reproductive capacities. ⁸ So perhaps effort is not unreciprocated in this case, and the labour (through payment) extracted from the couple is not beyond the labour put in by the surrogate. There is therefore no extraction of surplus labour, and both parties reciprocate effort.

On the chance there is no exploitation, it nevertheless remains that the surrogate controls the means of offspring reproduction, in a manner that the infertile couple do not. It is possible for dominators to violate the Non-Servitude Proviso without exploiting (52), and it seems that this might be what surrogates do. The couple remain dominated by the surrogate, therefore. Clearly something has gone wrong, however, if we conclude that surrogates are the dominators through their control of a means of production that dictates others’ ends. Such an understanding of both surrogacy and domination takes us rather far from the empirical reality of surrogacy, characterised by marginalised women contracting into unfair arrangements. If we want to maintain coherence with this reality, it seems that we need another explanation as to why surrogacy is exploitative.

Let us grant that a surrogate is exploited according to the story Vrousalis’ account provides. This becomes troubling for another reason. To recall, a woman’s reproductive capacity is assimilable to productive purposiveness, and so comes under the purview of Vrousalis’ Non-Servitude Proviso. Surrogate motherhood involves unilateral alien control over this capacity. The domination of the surrogate mother involves unilateral labour flow, in the form of her gestational care for the embryo, thus surrogate motherhood is exploitative.

However, this account is in danger of rendering non-contract, ‘ordinary’ pregnancy, similarly exploitative. It seems that what makes surrogacy contracts exploitative is alien control over reproductive capacity. However, what exactly does alien control consist of here? And how can

---

⁸ According to Vrousalis maintaining your organs requires labour.
control over one’s reproductive capacity be non-alienated? Is the idea that exploitative surrogacy involves unilateral labour flow in the form of ‘gestational care for the embryo’? Pregnancy, in so far as it involves women’s reproductive capacity and thus their productive purposiveness, cannot avoid being exploitative on this account. Pregnancy will necessarily involve some alien control over women’s purposiveness: if she is engaged in a relationship with a co-parent, they will bear the fruits of her labour without the same level of engagement or effort. Indeed, plausibly without even a commensurable level of engagement, as their body and thus purposiveness is simply not engaged in the same manner. The body is a finite and terminal resource, and pregnancy, childbirth, and nursing will have some unavoidable lasting impact. Even if a woman is single and her reproductive capacities are not drawn upon by a partner or another adult, her gestational care for the embryo seemingly constitutes unreciprocated effort: Can the foetus be said to pay its way?

Under what conditions would a woman’s reproductive labour cease to be under alien control? No answer is supplied, but Vrousalis does provide an answer for when a worker’s labour would not violate the Non-Servitude proviso. Alien control over a worker’s productive capacity would cease under: “A system of worker control coupled with a strongly predistributive form of public ownership” (184). Similarly:

The only way the individual worker will fail to be exploited is if she can bind any other worker in the economy, whether in her own firm or in any other, to performing as much and as good labour as he can bind her. (181)

Can this help us in regard to women’s reproductive labour? Women clearly cannot bind others into performing as much reproductive labour as they undertake. They might be able to enter arrangements with partners that ensure their partner remunerates them for their reproductive labour during pregnancy and maternity leave, but this arguably does not cover the totality of the bodily involvement. Vrousalis is clear to note that bodily transactions necessarily involve labour extraction:

The reproduction of our blood and kidneys requires labour. To think otherwise is to misunderstand the invariably effortful, often arduous, and sometimes painful, relationship to our own bodies and the bodies of others. (80)
Pregnancy arguably requires more such labour than the standard maintenance of our organs. How can women be counterfactually compensated for the way in which their body’s labour has been terminally extracted through their reproductive capacities?

It might be objected that pregnancy need not undermine purposiveness, as women, under conditions of gender and class equality could choose to engage in this reproductive labour on their own terms. However, ‘their own terms’ will arguably be modulated by the consideration of other interests: of the foetus, the co-parent, the community at large. To set one’s own terms of engaging in reproductive labour, in their entirety, and to receive commensurable reciprocated effort and retain the surplus labour, seems like a very long shot. And as aforementioned, it is not clear how one would adequately compensate for pregnancy and childbirth’s lasting physical impacts.

Perhaps Vrousalis would not be concerned with the admission that all pregnancy is exploitative insofar as it precludes the ability to retain control over your purposiveness. Indeed, perhaps this reinforces his earlier claims that women’s reproductive labour is coercively controlled by men under patriarchy. However, the problem here extends beyond patriarchy. Socially imposed disadvantage aside, this points to a fundamental form of exploitation that would persist even under equal gender and class relations, precisely because this labour is sui generis in the manner it undermines purposiveness and cannot be directly compensated for. The means of reproduction cannot change. Extraction seems inevitable. Here, Vrousalis would be in the company of Sophie Lewis (2019), who argued a similar such case with the core claim ‘reproduction is productive work’ in Full Surrogacy Now. All gestation is work because of the immense physical and emotional labour it requires—it is an ‘extreme sport’. Her ‘full surrogacy’ approach involves re-examining our [nuclear] modes of kinship to see where we can alleviate this strain. The idea is that we should live, care, and reproduce in larger systems of care as opposed to discrete private units. It remains unclear of course how such an alternative could adequately compensate women for their reproductive labour and cause their purposiveness to be non-alienated.

In any case, it seems that the application of Vrousalis’ domination-based account does not quite get to the heart of what is exploitative about surrogacy arrangements. In one sense, it might lead to the unpalatable

---

9 Setting aside the possibility, or indeed, desirability, of ectogenesis.
conclusion that surrogates dominate the infertile through their ownership of the means of reproduction. And on the intended interpretation of why surrogacy is exploitative, it might turn out that all pregnancy is exploitative on this view, which may or may not be an appetising bullet to bite.

IV. CONCLUDING REMARKS

It has been suggested that the wrongfulness of domination needs to be understood as a relational harm, on both sides of the relation. For example, a more expansive definition of patriarchy that encompasses oppressive gender socialisation highlights the way in which the dominant can be harmed qua dominators. A recognition of this wrongful relational aspect is particularly necessary as the concept of patriarchy is being used to furnish and provide grounds for accepting this domination account. Next, I examined the application of Vrousalis' account to commercial surrogacy, a practice that is often viewed as a paradigmatic case of exploitation. I found that his domination-based account of exploitation was in danger of rendering surrogates the dominating party in such arrangements. Vrousalis' theory of exploitation certainly provides a rich and convincing story of labour relations under capitalism. It seems clear and unobjectionable that workers are dominated, both before and upon entering the workplace. Alien control over a worker's purposiveness does seem to constitute a core element of what is wrong under capitalism. But definitions struggle to be applied in totality, and something seems to have gone amiss when we try and understand paradigmatic cases of exploitation such as surrogacy in this manner. Further, Vrousalis' understanding of women's reproductive capacity as being assimilable to productive purposiveness points to an unavoidable form of exploitation: pregnancy itself.

REFERENCES


Gulzaar Barn is an Assistant Professor in the philosophy department at the University of Amsterdam. Her main research interests lie in moral and political philosophy. She is particularly interested in theories of exploitation and questions surrounding sexual and reproductive ethics. Contact e-mail: <g.k.barn@uva.nl>
Reponses

NICHOLAS VROUSALIS
Erasmus University Rotterdam

I am grateful to Gulzaar Barn, Suzie Love, and Lucas Stanczyk for their comments on Exploitation as Domination (EaD). They raise questions about theoretical issues, such as the nature of the injustice of exploitation and its Kantian pedigree, as well as practical issues, from patriarchy and colonialism to sex work and contract pregnancy. I will address the injustice and overinclusion questions first, then turn to practical and methodological issues.

Kant, Unrequited Love, Overinclusion

Love and Stanczyk take issue with my account of exploitation on the grounds that it needs to be supplemented by a Kantian theory of freedom (Love) or that its explanatory ambitions are of "real but limited significance" (Stanczyk, 53). I discuss Love first, then Stanczyk.

Love argues that my account of exploitation would do well to adopt a Kantian conception of domination, based on the right to freedom. This includes a right to "innate equality, that is, independence from being bound by others to more than one can in turn bind them" (38). In EaD I remain non-committal as to the exact theory of domination; it suffices for my account that there exists one such sound theory. What I do instead is broach generic republican, Kantian, and recognitionist theories of domination and then argue that exploitation is domination-enabled self-enrichment on any of these accounts. But suppose I were to go Kantian, along Love’s lines: What would the upshot be?

Love writes:

Demonstrating that a particular relationship is a relationship of domination does not suffice to establish that this relationship is unjust. To establish the injustice of domination within a state, the domination must be shown to have legal force. (39)
So in Love’s interpretation, interfering with my body in the state of nature would not wrong me. For a Kantian, this claim is puzzling, since it seems to contradict Kant’s emphasis on independence ‘from the constraining choice of another’ as a *natural* right of humanity as such. For example, Kant thinks that a bully bullying you in the state of nature would wrong you. But Love is right that a similarly coercive and exclusionary right to *external* objects of choice requires a state. For Kant, rightful external possession requires some kind of coercive authority claiming to authorize such possession in the name of all—merely *unilateral* appropriation entails domination. So, if private property requires legal authorization and if capitalism requires private property, then capitalism requires law in that sense.

Chapter 4 of *EaD* argues for this claim, but in a roundabout way: as Love recognizes, my main aim is to show that capitalist domination presupposes a capitalist state that creates, interprets, and enforces property law. The capitalist state *claims* the authority to do these things in the name of all, which is part of what makes capitalist transactions between consenting adults possible. Love and I are therefore in agreement that legal institutions are a necessary condition for capitalist domination. And I think we also agree that they are a necessary condition for the *abolition* of capitalist domination, that is, for a system of laws that does not merely *claim* to be speaking in the name of free and equal people, but genuinely does so. But now note that, although all of these ideas are entailed by Kantian right, they do not seem to require it. So the *EaD* argument does not commit you to Kantian right.

Stanczyk takes a different tack. He argues that the account of exploitation I offer is inferior to a ‘reciprocity conception’ (RC) according to which:

*Reciprocity as mutual advantage relative to an appropriate benchmark.* Productive cooperative activity ought to benefit everyone relative to an *appropriate* benchmark, as given by the claims that people have independently of the demand for reciprocity. (45)

Stanczyk then argues that my domination account is either a variant of RC or otherwise overgenerates—it generates false positives. I will first explain in what sense my view might be construed as a version of RC. I will then show that, even if my view is not a variant of RC, it does not overgenerate. Stanczyk’s dilemma, in other words, is not exhaustive.
The domination account, as *EaD* presents it, is effectively a theory of subsumed labour: exploitation is a dividend of servitude, where that dividend is cashed out in terms of unilateral labour flow. Exploitation, in this view, is domination-induced unreciprocated service to others. The nature of this relationship is typified by the:

Non-Servitude Proviso [NSP]—For any agents or groups engaged in mandatory mutually affecting cooperation under a division of labour, and barring any special justification that exempts them, none should possess unilateral control over the labour capacity of any other. (*EaD*, 40)

Does NSP entail RC? A lot hinges on how one's independent ‘claims’ are spelled out. The NSP supposes that each producer is entitled to equal productive independence from the choices of others, unless there is an independent justification for inequality. If the entitlements it designates entail Stanczykian ‘claims’ then the NSP might entail the RC.

How might my account and Stanczyk’s diverge? One way is suggested by Stanczyk:

Consider [...] the exploited, deeply caring person and ask the following question. Is it possible to take advantage of a person’s love and selfless character without ever *making* her do anything? It seems to me that this is eminently possible [...] If this is right, however, then the domination account faces a dilemma. It can admit that the loving labors of an exploited party to an intimate relationship need not be unilaterally directed by her selfish partner. But in that case, it will have failed to capture one of the paradigmatic forms of interpersonal exploitation. Alternatively, the account can be revised to say that *whenever* a person wants and labors for the approval of someone else, their purposive activity should be seen as ‘unilaterally directed’ by an alien will [...] In this case, however, an implication of the account will be that exploitation and servitude are omnipresent in human life, and likely to be permanently so. (52)

This example is a variant on the *Charlotte and Werther* case (*EaD*, 18). Stanczyk’s treatment of that example suggests that my view is either inconsistent—because it deems Charlotte to be undominated—or overgenerates—because it deems too many relationships as dominated. Stanczyk maintains that RC, by contrast, suffers neither of these infir-
mities, nor the dilemma they constitute. I begin by noting that his example is underdescribed: if Werther does not categorically disavow Charlotte’s giving behaviour, he may yet be *making* her do things. As Goethe knew, unrequited love, unless categorically disavowed by its object, might further excite the passions of its subject. So Werther may still be dominating Charlotte, much like the retired pusher who negligently benefits from the addiction of the addict.

More generally, I do not agree with Stanczyk’s intuition (that Werther exploits) if he does not, in some sense, control Charlotte’s action. The sheer fact that Werther “takes advantage of [Charlotte’s] love and selfless character” (52) does not show that he exploits *her*, as opposed to those mere features *of* hers (selflessness and selfless love). In order to exploit *her person*, Werther must, in addition, get Charlotte to do something, namely serve him in some way. But if so, then her actions are up to him, whether he knows/wills/intends it or not. It is relevant here that Stanczyk’s RC itself is more ambitious than the NSP, since the latter only elaborates on one kind of reciprocity-failure, namely exploitation. Non-exploitative unfairness, *EaD* argues, is a distinct way in which Werther might mistreat Charlotte (*EaD*, 88–90).

Moreover, domination is easier and therefore more widespread than Stanczyk seems to allow. Werther can get Charlotte to act, in the domination-relevant sense, not only when she lacks options or when he deprives her of them; Charlotte might have meaningful options and still be dominated. (Lacking options is a sufficient but unnecessary condition for domination—see *EaD* 42–43, 77–80.) So pace Stanczyk unrequited love is exploitative only if its object benefits through—possibly unexercised—power over its subject. Those unconvinced that this accords with ordinary usage can just call the phenomenon I am describing *explominination* (*EaD*, 89).

Barn mentions yet another way in which *EaD* might overgenerate. Adapting my discussion of omelette production (*EaD*, 38–39) to human reproduction, she writes:

As consumers, the infertile couple have set themselves the end of producing offspring. As producers, they must use means to pursue and fulfil that end. The surrogate owns the means of (re)production that the infertile couple do not. As long as the surrogate (and fertile others) own this, while the infertile couple do not, the infertile couple can produce offspring only by the surrogate’s permission. This
makes the couple’s ability to set and pursue the end of offspring production—their productive purposiveness—*subordinate to the surrogate’s unilateral will*. When that subordination is expressed in extraction of unilateral labour flow from the couple (such as the thousands paid in surrogacy fees), the surrogate exploits the infertile couple. As long as production is in the surrogate’s hands, and when that control eventuates in unilateral labour flow, i.e., whenever she charges others for access to the means of production, *she exploits them*. (63–64)

Barn is right that poor women possessed of reproductive capacities who sell them to rich couples lacking them do not exploit the latter. So if *EaD* yields this conclusion, then it overgenerates. My response in the book (*EaD*, 80–81) is to deny Barn’s premiss: *EaD* does not take poor women possessed of scarce reproductive assets to be exploiters, because both the nature of the good they sell and the terms on which they must sell it is not up to them. Consider, first, the nature of the good. Giving someone discretion over the use of your body in return for money is akin to contract slavery; no such thing is given by the buyer. In addition, the mere availability of pregnancies for sale stigmatizes the sellers as ‘baby machines’, where, again no comparable stigmatization obtains for the buyer.¹ Now consider the terms of the transaction: *EaD*’s judgment of the rich couples as exploiters *might* have been different if the trade obtained between transactors of roughly equal bargaining power. But Barn’s examples take the sellers to be vulnerable and relatively poor people, not free and equal people.² So the *EaD* idea—that the rich couples exploit—survives.

**Patriarchy, Colonialism, Racism**

Barn doubts that the *EaD* account is pluralistic enough to capture paradigmatic cases of exploitation. Unlike Stanczyk, who suggests that the *EaD* theory is overinclusive, Barn claims that it is underinclusive. Noting my definition of patriarchy as “the systematic disempowerment of

¹ In her discussion of contract pregnancy, Satz (2010) emphasizes this aspect, as well as how it reinforces sexist stereotypes of “women as the incubators of men’s seeds” (131).

² I would want to resist Barn’s *reductio* even in cases of equal alienable and inalienable assets (including reproductive assets). One way to do this could, for example, appeal to the additivity axiom of *EaD*, 77–78.
women with respect to their ability to self-direct their free purposiveness" (*EaD*, 99), Barn writes:

It would seem that under patriarchal socialisation men suffer a similar such fate: the internalisation of harmful norms is poised to undermine their ability to self-direct free purposiveness. If not, then Vrousalis can not account for the manner in which internalised patriarchal ideology harms women through domestic and sexual norms. Either Vrousalis needs a discriminating reason as to why only women's purposiveness is undermined by internalised gender norms, or he must concede that men's purposiveness is similarly undermined. If this is so—that a dominating relationship seems to have such an effect on both the dominant and subordinate party—what does this tell us about the nature and wrongfulness of domination? While this harm to men might not constitute domination, it may still be reasonably considered under the purview of justice, requiring redress. (60)

I agree with Barn that patriarchal norms are vehicles of domination (*EaD*, 95–98). But it does not follow that the positions of those who enact these norms are symmetric in the way Barn suggests; to think otherwise is to throw out the domination baby with the structural bathwater. Sure, structure constrains both dominated and dominators, but that is built into the hierarchical structure of pits, cages, prisons, slave markets, and so on. And it is incoherent to think that the injustice perpetrated by the slave owner wrongs him *as* a slave owner. Perhaps he is harmed as a person, but the harm is, by definition, not unjust; it might even be good *at the bar of justice* if the interests and ends of the unjust were set back.

To buttress this charge of underinclusion, Barn quotes a passage by Aimé Césaire, which explains how colonial exploitation dehumanizes the colonizer. The explanation is true but ambiguous. It could mean that the colonizer's character is divided by injustice, such that he cannot act as a unified person. That's bad for his character. But it is only bad for justice if it subverts its realization as a character-independent state of affairs. Marilyn Frye's (1983) seminal discussion of the patriarchy is relevant here:
We hear that oppressing is oppressive to those who oppress as well as those they oppress. Some men cite as evidence of their oppression their much-advertised inability to cry. It is tough, we are told, to be masculine. When the stresses and frustrations of being a man are cited as evidence that oppressors are oppressed by their oppressing, the word ‘oppression’ is being stretched to meaninglessness; it is treated as though its scope includes any and all human experience of limitation or suffering, no matter the cause, degree or consequence. (28)

But suppose that Barn is right and Frye is wrong: the unjust unjustly harm themselves. I do not think self-domination is possible, for the same reasons that self-exploitation is not possible (EaD, 134–135). So if the relevant injustice to self was supposed to be domination, then Barn’s suggestion would be incoherent. Barn suggests that the unjust may be doing themselves an injustice without dominating themselves. This is a coherent possibility, which might give us additional reasons to abolish colonial injustice. Even so, it does not follow that colonial injustice makes the colonizer worse off overall. By definition, the colonial exploiter benefits at the expense of the colonized, so a conscientious colonizer must think that the loss his character suffers is worth it—otherwise he would have quit the whale and gone to Catalunya as Orwell did.

I now register a caveat that supports Barn’s concerns, namely that not all exploiters are blameworthy. Some exploitation may be excusable or even blameless. To see this, suppose that a group optimizes by exploiting as a group—not as a group agent—such that no individual exploiter in it is individually blameworthy for exploiting (see EaD, 101–109); only the group as a whole is blameworthy. For example, it is plausible to think that only capitalists as a whole are (morally) responsible for the exploitation of workers as a whole (EaD 42–43, 105–108). This suffices to let individual capitalists off the hook. In fact, this is what structural exploitation, briefly discussed in chapter 4 of EaD, is supposed to be about: exploiters are collectively, but not individually, blameworthy if each performs her share of an injustice that only the group as a whole can perform and if that performance is not, considered independently of the group’s action, necessarily blameworthy or unjust.

Barn further suggests that the application of my account to colonial and liberal imperialism (EaD, chapter 6) may be underinclusive because
it has little to say about racial subordination. As Barn notes, the reason for this is that I take imperialism to be broader than colonialism, where only the latter is necessarily racist. Moreover, quite independently of any Fanon-type account of the psychology of colonialism, the injustice of colonialism, together with slavery, counts among the most extreme forms of subjection of purposiveness—one’s ability to set and pursue ends. So colonialism and slavery are, on the EaD view, among the most extreme forms of exploitation.

REFERENCES

**Nicholas Vrousalis** is an Associate Professor in Practical Philosophy at Erasmus University Rotterdam.
Contact e-mail: <vrousalis@esphil.eur.nl>
Reflections on the 2021 Nobel Memorial Prize Awarded to David Card, Joshua Angrist, and Guido Imbens

LENNART B. ACKERMANS
Erasmus University Rotterdam

The 2021 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel was awarded in two halves. One half was awarded to David Card “for his empirical contributions to labour economics” (The Royal Swedish Academy of Sciences 2021, 1). The other half was awarded jointly to Joshua D. Angrist and Guido W. Imbens “for their methodological contributions to the analysis of causal relationships” (1). In this article, I (a philosopher of science interested in causal inference in economics) reflect on the second half of the 2021 Nobel Prize, awarded to Angrist and Imbens.

Two beautiful examples of causal inference in economics are Angrist (1990) and Angrist and Krueger (1991), published shortly after Joshua Angrist obtained his PhD in 1989. (His co-laureates David Card and Guido Imbens are his contemporaries, obtaining their PhDs in 1983 and 1991, respectively.) The 1990 study estimates the causal effect of veteran status on earnings 30 years later. It finds that white U.S. veterans from the Vietnam War have approximately 15% lower earnings as a result of military service. The 1991 study finds that, in the U.S., having an additional year of compulsory schooling has a large effect on earnings later in life (Angrist and Krueger 1991).

At the time, other studies into both of these subjects struggled to find causal effects, as opposed to mere correlations. For example, when one finds a negative correlation between veteran status and earnings, it is unclear whether this is because veterans had a lower earning potential prior to being enlisted, or because lower earnings result from serving in the military. Angrist (1990) solved this problem using an instrumental variable approach. During the Vietnam War, men were drafted based on a lottery that assigned numbers between 1 and 365 based on birth date. Only men below a certain lottery number were drafted. Since the draft lottery
number is randomly assigned, the causal effect of the lottery number on earnings can be identified from the observed data. Since an observational situation like this is similar to an experiment, such as a randomised controlled trial (RCT), it is called a natural experiment.

However, we are ultimately not interested in the effect of lottery number on earnings, but in the effect of military service. Lottery numbers are not a perfect substitute for enlistment because some people with low lottery numbers did not enlist and some people with high numbers volunteered. To estimate the effect of military service itself, Angrist used the lottery number as an instrumental variable. (An instrumental variable has some effect on the explanatory variable of interest. It is used for its better inference properties—particularly unconfoundedness, discussed in section 2 below.) An instrumental variable approach first estimates the effect of lottery number on earnings and the effect of lottery number on military enlistment. With some additional assumptions that might be controversial, the two results can be combined to derive the effect of military service on earnings.

The draft lottery study has become a classic example of successful causal inference, but in 1990 Angrist was not yet convinced of its persuasiveness. As he recalls in his Prize Lecture: “Guido and I soon began asking each other: What really do we learn from the draft eligibility and quarter of birth natural experiments?” (Angrist 2021). It was only in the years that followed that Angrist and Imbens were able to answer this question with their ground-breaking methodological work, which includes classic papers such as “Identification and Estimation of Local Average Treatment Effects” (Imbens and Angrist 1994) and “Identification of Causal Effects Using Instrumental Variables” (Angrist, Imbens, and Rubin 1996). Later work added additional causal methods such as regression discontinuity approaches (Imbens and Lemieux 2008). It is in methodological studies such as these where the most important contributions of Angrist and Imbens lie.

An uncontroversial way to characterise these contributions is to say that Angrist and Imbens developed clever methods of causal inference—such as instrumental variable approaches and regression discontinuity—which have allowed economists to produce more successful causal studies, such as the draft lottery and compulsory schooling studies. However, I shall argue that their contribution is more important. In the 1980s, econometrics was in a state of crisis. Edward Leamer put it well in his 1983
article titled, “Let’s Take the Con Out of Econometrics”, in which he pronounced that econometric practice of the time was “decidedly unscientific” (37). Even worse, everyone knew it: “Hardly anyone takes data analyses seriously. Or perhaps more accurately, hardly anyone takes anyone else’s data analyses seriously” (37). (See also LaLonde 1986 and discussions of the credibility crisis by Angrist and Pischke 2010; Imbens 2022.) Angrist and Imbens led the way out of this crisis by shifting the field’s attention toward causal research design. Three decades later, Joshua Angrist and Jörn- Steffen Pischke (2010) declared that econometrics had made significant progress since Leamer’s critique.

I agree. Advances in causal methodology made by Angrist, Imbens, and others have been especially important because they mark the first step out of the dark ages of econometrics. However, this process is still incomplete, and the legacy of Angrist and Imbens will grow larger still—or at least, it should grow larger. This brings me to my second thesis. At present, the econometric methodology of the crisis era is still prevalent. Moreover, economists are divided on how to proceed, with heated debates over the causal framework to adopt. The framework championed by Angrist and Imbens, the Rubi n Causal Model, has limitations that I believe might hamper a more widespread adoption in the field. I will argue that the profession needs to prioritise the resolution of these problems so that it can put causal inference at the forefront of economic inquiry.

I. THE IDENTIFICATION PROBLEM

While it may not have been clear in 1991, the draft lottery and compulsory schooling studies were important achievements that marked the way out of the credibility crisis. But why was there a crisis in the first place, why were these instrumental variable studies so successful, and how did they contribute to resolving the crisis? In order to answer these questions, let me introduce the identification problem.

In simple terms, the identification problem is that statistics is not causal inference. More precisely, exercises in pure statistics are never enough to discover a causal relation. (There are a variety of other definitions of the identification problem in economics. Sometimes it refers to the more general problem that a parameter in a model cannot be estimated from observations.)

By pure statistics I mean mathematically describing data in the following way. It is assumed that the data in one’s sample is drawn from a larger set of data called the population—which can be hypothetical and infinitely
large—either randomly or using some known or unknown procedure. The pure statistician uses the sample data to make inferences about the population data. Typically, one is interested in a number of mathematical parameters describing the population data, such as means, conditional means, and correlation coefficients. Often, the statistical parameters of interest figure as Greek letters in regression equations describing the population data, such as:

\[ Y_i = \alpha A_i + \beta B_i + \epsilon_i. \]  

(1)

Here \((Y_i, A_i, B_i)\) is the \(i\)th data point and \(\alpha\) and \(\beta\) are population parameters that the statistician aims to estimate from the sample data, typically using a regression method such as ordinary least squares (OLS).

The error term \(\epsilon_i\) is the deviation of \(Y_i\) from its expected value given \((A_i, B_i)\), which is estimated by the regression residual. Equations such as (1) can be great for describing data statistically. In that case, they may also be valuable for predicting the values of new samples drawn from the same population.

An entirely different interpretation of equation (1) is that it is a model that describes an underlying causal structure, in which case it is called a structural equation, structural model, or causal model. If (1) is a causal model, \(\alpha\) is interpreted as a causal effect of \(A\) on \(Y\), \(\beta\) as a causal effect of \(B\) on \(Y\), and \(\epsilon_i\) as a combination of unobserved variables having a separate causal effect on \(Y\). These are now causal parameters instead of statistical parameters. On this interpretation, the equation describes how the \(i\)th individual’s outcome \(Y_i\) was causally determined based on the values of \(A_i, B_i,\) and \(\epsilon_i\). It implies that individual \(i\)’s value of \(Y_i\) would have been as described by (1) if \(i\)’s values of \(A_i, B_i,\) and \(\epsilon_i\) had been given any different value. Such counterfactual knowledge importantly allows the researcher to predict what will happen if the studied population changes, for example, due to policy interventions.

If (1) is a causal model, the researcher can still try to estimate the (causal) parameters with the methods of pure statistics, such as OLS. However, such an endeavour is usually unwarranted. A major problem is that the same data can always be generated by multiple different mechanisms. Even if the data is well described by (1), there are many different models that could have generated the exact same data. If the pure statistician would have estimated the parameters of another model, entirely
different causal parameters would have been ‘discovered’. Hence, statistical estimates of $\alpha$ and $\beta$ can be interpreted as causal effects only if we have good reasons to assume that the underlying causal structure is given by (1).

This suffices to show that causal inference requires more than pure statistics. That said, it should be mentioned as an aside that the use of advanced statistical methods is extremely important for causal inference. Some important contributions that Angrist and Imbens have made to causal methodology are best classified as pure statistics. For example, Angrist, Imbens, and Krueger (1999) offer a solution to a problem of bias that occurs in two-stage-least-squares (2SLS) estimation, a statistical method essential for causal methods using instrumental variables. This solution, called Jackknife Instrumental Variables Estimation (JIVE), is then used to re-analyse the 1991 Angrist and Krueger study.

There have been economists throughout the 20th century who understood the identification problem and the difficulties with causal inference (e.g., Haavelmo 1943). Nevertheless, econometrics textbooks to this day—while great at teaching pure statistics—are creating more confusion than clarity when it comes to causal inference, as several authors in the field now recognise (Heckman and Pinto 2022b; Angrist, Imbens, and Rubin 1996).\footnote{See footnote 4 for the relevant quotes from these authors.}

When a typical textbook in econometrics introduces the OLS regression model, it informs the reader of a crucial assumption (e.g., Wooldridge 2010; Greene 2018). The error term $\epsilon_i$, as it appears for example in equation (1), must be uncorrelated with the other variables appearing on the right-hand side. This assumption is known as econometric exogeneity (or just exogeneity). If the error term does correlate with a regressor, called econometric endogeneity, then the regression estimates of the parameters $\alpha$ and $\beta$ are said to be biased or non-causal. Once exogeneity is assumed, econometricians are able to apply powerful tools from pure statistics. However, exogeneity is an assumption about structure (causal or otherwise) and needs a defence that goes beyond pure statistics.

This is where the textbooks fall short. The exogeneity assumption as it typically appears in works of econometrics is “either meaningless or false”, as Pratt and Schlaifer (1984, 11) summarise it. Econometricians tend to define error terms as the combined effect of unobserved variables. However, the concept of ‘omitted variables’—without giving it a more precise definition—is so vague that the exogeneity assumption has no real
content. In any given regression an infinite number of variables have been omitted. Without specifying which unobserved variables are meant, the correlation between $\epsilon_i$ and the regressors is not defined. Depending on how $\epsilon_i$ is interpreted, the correlation with other variables can have any value.

To see why this is so, consider the model $Y = \alpha X + \epsilon_1$, where $X$ is exogenous, i.e., $\text{Cov}(X, \epsilon_1) = 0$ and $X$ and $\epsilon_1$ have mean 0. Now consider another variable, $\epsilon_2 = \beta X + \gamma \epsilon_1$. Then the model $Y = \alpha X + \epsilon_2$ could describe the data just as well. However, in this model, $X$ correlates with $\epsilon_2$. Without specifying what a variable $\epsilon$ is, it could be $\epsilon_1$, $\epsilon_2$, or many other things. Which variable one should choose depends on the causal effect one wants to measure. For example, if the causal effect of $X$ on $Y$ is $\alpha$, then the error term is $\epsilon_2$, and $X$ is endogenous—so the causal effect cannot be identified with OLS. If the causal effect is $\beta$, then $X$ is exogenous and the causal effect identifiable.

Hence, for the exogeneity assumption as it is typically invoked in econometric studies, it is impossible to check whether it is true or false. That said, if one assumes a particular causal structure, it is possible to give the error term a definition for which its correlation with other variables is defined (Pearl 2009, chapter 5). However, textbook econometrics is devoid of such causal assumptions. Under these circumstances, exogeneity is not a meaningful assumption. (These problems have long been understood. See the classic papers by Haavelmo 1943 and Pratt and Schlaifer 1984.)

From my understanding, what was wrong with econometrics as practised in the 1980s was that researchers did not have a clear understanding of the above issues (this diagnosis is similar to Pearl 2009, chapter 5; Imbens 2022). In particular, the difference between pure statistics and causal inference was often obscured, with regression equations like (1) not having a clear interpretation as either describing data or causal structure. As Imbens (2022) observes, the term ‘causality’ was rarely used in econometrics between the 1960s and 1980s, until it was revived in the 1990s—despite the fact that econometricians were often concerned with clearly causal questions.

This lack of causal terminology can even be found in the very articles that identified the credibility crisis in the 1980s. Leamer (1983), while seemingly concerned that conventional regression estimates in econometrics do not match causal parameters, does not mention causality in the
Another interesting paper is LaLonde (1986), which compares experimental and non-experimental methods using the same data. The data comes from an RCT designed to estimate the effect on trainee earnings of an employment program. Putting aside the control group, LaLonde applied state-of-the-art econometric techniques for use with observational data—and he was unable to replicate the results from the RCT. Like Leamer, LaLonde did not use words like ‘causality’. However, LaLonde did identify ‘model misspecification’ in observational methods as a problem. If a ‘model’ is a causal model, this was going in the right direction. However, to properly analyse and resolve the problems that econometricians became aware of in the 1980s, a more principled understanding of causation and methods of causal inference was needed.

Given the clarity of causal reasoning found in the draft lottery and compulsory schooling papers (Angrist 1990; Angrist and Krueger 1991), I imagine that the authors had a clear understanding of the difficulties associated with causal inference. The semi-experimental methods they used were uncommon in economics at the time and would later lead the way out of the credibility crisis. However, the causal reasoning in these papers is not principled in the sense of being based on well-studied formal principles of causal inference. Around 1990, causal reasoning in economics relied on intuitions rather than theory and was thus more of an art than a science. But this was about to change.

II. THE RUBIN CAUSAL MODEL

Causal inference, like statistics, must be done with the help of formal frameworks that assist the scientist in reasoning correctly and precisely. The problem in the 1980s was that economists had mastered well-developed and sophisticated tools of pure statistics, while their tools for causal inference were lagging behind. Fortunately, statisticians had already developed a framework for causal inference, known as the Rubin Causal Model (Rubin 1974) named after Donald Rubin by Holland (1986), but going back to Neyman ([1923] 1990) and Cox (1958). This section introduces the Rubin Causal Model (RCM) and illustrates how it improved econometricians’ understanding of causal methods, using the example of instrumental variables.

The strategy of RCM is to use the RCT as a foundation on which to build a framework which extends well beyond RCTs. As in an experiment, we imagine that each individual can be given the treatment \( T = 1 \) or no treatment \( T = 0 \). An individual’s outcome if treated is denoted \( Y_i(0) \), and
an individual’s outcome if not treated is denoted $Y_i(1)$. These are called potential outcomes, of which at least one is counterfactual. The individual treatment effect for $i$ is given by $Y_i(1) - Y_i(0)$.

It is a virtue of the RCM that it relates a causal effect so clearly to a counterfactual: the effect of $i$’s treatment is the difference between $i$’s outcome if $i$ were treated and if $i$ were not treated. Unfortunately, only one of these outcomes can be observed. Hence, we need clever strategies in order to learn something about causal effects without ever being able to observe them directly.

As it turns out, various types of average treatment effects (ATEs) can sometimes be derived from statistical data. This is the case, for example, for an RCT with perfect compliance. (Perfect compliance means that all participants get the treatment if and only if they are assigned the treatment). In the perfect RCT, due to the random assignment of treatment, the average observed difference between the treatment and control group is an estimate of the average counterfactual difference for all individuals. More precisely, one can show that

$$E[Y_i(1) - Y_i(0)] = E[Y_i | T = 1] - E[Y_i | T = 0],$$

where $Y_i$ is $i$’s observed outcome. The expectation on the left is called the average treatment effect—which is an average of causal effects that is not directly observable. The expression on the right, on the other hand, can be estimated from the observed data with the techniques of pure statistics.

Unfortunately, observational data is never like an RCT with perfect compliance. But fortunately, there is now a large literature with methods to identify treatment effects with weaker assumptions, including assumptions that are sometimes satisfied in observational data. The important contributions from Angrist and Imbens lie mostly in this area.

Their most influential achievement is perhaps a method for identifying the local average treatment effect or LATE (Imbens and Angrist 1994; Angrist, Imbens, and Rubin 1996). The LATE is an average treatment effect for a subpopulation of the data: namely, those individuals whose treatment status always matches their assignment, called compliers. The LATE can be estimated in RCTs with imperfect compliance, but its greatest success stems from the fact that it can sometimes be estimated from
purely observational data. This is the case when the data contains an instrumental variable—call it $Z_i$—which has the properties of treatment assignment in an imperfect RCT.

To illustrate the kind of assumptions required for causal inference within RCM, I will give a somewhat technical discussion of the LATE using RCM terminology. This will pay off in the next section, which compares the merits and problems of RCM with other frameworks.

An instrumental variable $Z_i$ is an observable variable that has some causal influence on individuals’ treatment $T_i(z)$ with $z \in \{0,1\}$. Here $T_i(z)$ is the treatment that $i$ would have if it were the case that $Z_i = z$. The LATE is defined as

$$E[ Y_i(1) - Y_i(0) \mid T_i(1) = 1, T_i(0) = 0 ].$$

Angrist and Imbens showed that the LATE can be identified if three important assumptions are satisfied (as well as some others). First, the potential outcomes $Y_i(t)$ are unaffected by $Z_i$. More precisely, if $Y_i(z, t)$ is $i$’s potential outcome given $Z_i = z$ and $T_i = t$, then we have $Y_i(z, t) = Y_i(t)$, for all $z, t \in \{0,1\}$. (A more intuitive formulation of this assumption may be that $Y_i$ is unaffected by $Z_i$ if the treatment $T_i$ is held fixed.) This is called the exclusion restriction. Second, $Z_i$ must have the properties of random assignment. In RCM terminology, this assumption states that $Z_i$ is probabilistically independent of the potential outcomes $(Y_i(0), Y_i(1), T_i(0), T_i(1))$. (That is, it is jointly independent of these four variables. I will explore this assumption in greater detail in the next section.) This assumption is usually called unconfoundedness. Third, assignment to the treatment must make treatment more likely for each individual. More precisely, there should be no defiers, individuals who do the opposite of their treatment assignments. Defiers are individuals such that $T_i(1) = 0$ and $T_i(0) = 1$.

The LATE method showcases how RCM can be used to prove mathematically that a causal effect can be identified from the data given these assumptions. This subsequently makes it possible for applied researchers to increase the credibility of their studies, provided that they can make it plausible that these assumptions are indeed satisfied. The assumptions contained in LATE and other RCM-based methods are certainly easier to defend than econometric exogeneity, by virtue of their rigorous explication. However, they are still not quite easy to defend—which brings me to one of RCM’s foremost shortcomings (see also Pearl 2009, 98–102).
Let us look at what these assumptions mean in the draft lottery study. In this study, the instrumental variable $Z_i$ is $Z_i = 1$ if the individual has a low lottery number, such that he is eligible for the draft, $Z_i = 0$ otherwise. For the treatment we have $T_i(Z_i) = 1$ if the individual is enlisted, $T_i(Z_i) = 0$ if not. Finally, $Y_i = Y_i(T_i)$ is $i$’s observed income 30 years after the draft.

The exclusion and no-defiers assumptions are relatively straightforward to defend. The no-defiers assumption says that there are no individuals that would have volunteered for military service with a high lottery number but would not enlist with a low lottery number. It seems safe to assume that such individuals are rare enough to ignore.

On the other hand, assessing unconfoundedness is a mental nightmare. Unconfoundedness says that lottery number $Z_i$ is jointly independent of all the potential outcomes $(Y_i(0), Y_i(1), T_i(0), T_i(1))$. Assessing this assumption requires one to imagine population data which contains not only individuals’ actual treatment and outcome values but also the treatment and outcomes that they would have under counterfactual conditions. Without additional guidance, this assumption is very hard to assess. Unfortunately, the causal framework RCM itself does not provide much help in assessing whether unconfoundedness is satisfied.

Methods to test indirectly whether unconfoundedness is satisfied, based on RCM, do exist (see e.g., Imbens and Rubin 2015, chapter 21). The problem is that assessing unconfoundedness requires much more than some mathematical methods which a researcher can simply ‘run’. More importantly, it requires an informal understanding of the underlying causal structure and a way to translate this understanding into formal assumptions of probabilistic independencies. (Even the tests in Imbens and Rubin 2015 require informal input based on the researcher’s intuitions and theoretical knowledge.)

It is my position that a causal framework is supposed to help the researcher with this translation step from structural causal knowledge to methodological assumptions. RCM, however, is unsuited for this task by construction. Causal connections are not expressed in RCM, which instead focuses on independencies in imaginary population data that includes potential outcomes. The result is that all assumptions in RCM are expressed in terms of imaginary data, without using any causal terms. To assess the assumptions, however, one needs to consult one’s causal knowledge. For example, knowledge of whether $Z_i$ has common causes with $T_i$ or $Y_i$ should be used to assess unconfoundedness. Such knowledge comes in terms of causal connections, not imaginary population data.
However, proponents of RCM insist that they find these assumptions quite intuitive. For example, in response to a similar concern voiced by Pearl, Imbens (2020, 1164) replies: “I think that statement [from Pearl] misses the point. This setting, where the critical assumption is ignorability or unconfoundedness, is so common and well studied that merely referring to its label is probably sufficient for researchers to understand what is being assumed”.

Irrespective of its potential problems, RCM has been extremely important for the development of causal methods such as the LATE. Both Angrist and Imbens mention this importance in their prize lectures (Angrist 2021; Imbens 2021). Nevertheless, RCM has not managed to replace the textbook approach to econometrics in most econometric research. Part of the problem is that there are several contenders aiming to replace textbook econometrics as a framework for causal reasoning.

III. CONTENDING CAUSAL FRAMEWORKS

Separately from the RCM developed by statisticians, computer scientists and philosophers developed another causal framework, which I will call the Pearl Causal Model (PCM) after its primary author Judea Pearl (Pearl and Verma 1991; Spirtes, Glymour, and Scheines 1993; Pearl 1995, 2009). The PCM makes extensive use of directed acyclic graphs (DAGs) to formulate assumptions about causal structure. James Heckman has defended another causal framework which he claims is closer to the traditional econometric framework (Heckman 2000, 2005; Heckman and Pinto 2015). Let us call this the Heckman Causal Model (HCM). In this section, I summarise these frameworks and show how they can be used to shed light on the assumption of unconfoundedness.

Both Heckman and Pearl are influenced by earlier economists’ work on structural equation modelling such as Frisch ([1938] 1995) and Haavelmo (1943, 1944). Moreover, in its most recent explication, HCM makes heavy use of DAGs to express structural causal assumptions graphically, as well as other tools from the PCM literature (Heckman and Pinto 2015, 2022b). Hence, the two approaches are spiritually and practically similar. RCM, on the other hand, eschews the use of structural equations.

In both approaches, the foundation of causal inference is causal modelling. Before one can reliably estimate causal effect sizes, one typically needs to have knowledge about causal structure—that is, knowledge
about how the variables in a system are causally connected. Causal models in HCM and PCM summarise such information using equations and graphs. For example, figure 1 represents the research design of an instrumental variable setup like the draft lottery study. The nodes in this graph represent causal variables, and the arcs represent causal connections. For example, \( Z_i \rightarrow T_i \) means that \( Z_i \) is a cause of \( T_i \). The unobserved variable \( U_i \) is responsible for individual differences in their response to the treatment assignment, as represented by the arc \( U_i \rightarrow Y_i \).

HCM and PCM can express unconfoundedness both graphically and probabilistically. In what follows I illustrate how these frameworks give the researchers additional tools in understanding and assessing unconfoundedness.

Consider again the model of an instrumental variable setup in figure 1. The graphical unconfoundedness assumption states that \( Z_i \) is causally connected with \( Y_i \) only by causing it (via \( T_i \), as in figure 1). In other words, there must not be a common cause \( C_i \) of \( Z_i \) and \( Y_i \), that is, a path

\[
Z_i \leftarrow C_i \rightarrow Y_i.
\]

Graphical unconfoundedness implies probabilistic unconfoundedness under an assumption called the *Causal Markov Condition*. The Causal Markov Condition states that a causal variable is independent of its non-descendents conditional on its parents, supposing that the DAG is a sufficiently accurate representation of reality. For instance, in figure 1, the Causal Markov Condition implies that \( Y_i \) is independent of \( Z_i \) given \((T_i, U_i)\). The Causal Markov Condition is a well-studied principle that is plausible in most circumstances, although objections exist (e.g., Cartwright 1999).

Based on the Causal Markov Condition and a variety of rules for manipulating conditional independence relations (from Dawid 1979), a researcher can quickly derive all conditional independencies implied by a DAG. To make a connection with the probabilistic unconfoundedness assumption in the previous section, one can create ‘hypothetical versions’ of a graph in which treatment variables are replaced by counterfactual variables. In HCM, one creates a hypothetical model given counterfactual assignments of \( Z_i \) as follows. First, add a counterfactual treatment assignment variable \( \hat{Z}_i \) to the graph. Then remove all outgoing arrows from \( Z_i \) and instead assign them as outgoing arrows from \( \hat{Z}_i \). The resulting graph, depicted in figure 2, represents the causal model given counterfactual assignments \( \hat{Z}_i \). (PCM uses a slightly different procedure to create counterfactual models.) By reading off independencies from the hypothetical...
graph, the researcher can quickly observe that $Z_i$ is independent of $T_i$ given counterfactual assignments $\hat{Z}_i$. Similarly, figure 3 gives the hypothetical graph given counterfactual assignments of $T_i$. From this graph, the researcher can observe that $Z_i$ is independent of $Y_i$ given counterfactual assignments of $\hat{T}_i$. These results in turn can be shown to imply the probabilistic unconfoundedness assumption from the previous section.\footnote{See Heckman and Pinto (2022b) for a more detailed analysis.}

The above illustrates how causal graphs can be used by researchers to use their theoretical knowledge of causal structure, as expressed in a DAG, to assess whether assumptions for causal methods are satisfied.

---

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{Causal graph of an instrumental variable setup.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2.png}
\caption{Causal graph for counterfactual assignments of $Z_i$.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3.png}
\caption{Causal graph for counterfactual assignments of $T_i$.}
\end{figure}
While above I went through the reasoning from structure to independence assumptions explicitly, researchers do not typically need to do so themselves. Pearl’s book, and Heckman and Pinto’s recent articles, describe many causal methods based on assumptions that are expressed in graphical terms, such as Pearl’s back-door criterion and front-door criterion (Pearl 2009). This allows researchers to immediately apply these methods once they have identified an accurate causal structure.

The graphical approach of PCM and HCM shifts the researcher’s attention to an important precondition of causal inference: the identification of causal structure. As illustrated above, one needs detailed knowledge of causal structure before the assumptions required for causal methods can be verified. Graphical causal frameworks not only make it easier to express structural causal knowledge but also come with a rich literature that helps researchers to discover causal structure from data, including algorithms that search for causal relations in data (Spirtes, Glymour, and Scheines 2001).

Hence, the graphical frameworks PCM and HCM supply the researchers with a more complete set of tools, including tools for estimating causal effect sizes, verifying structural assumptions, and discovering causal structure. All tools are part of the same graphical framework, allowing scientists to combine them easily.

IV. THE RECENT DEBATE: HECKMAN AND IMBENS

The previous section showcases some of the benefits of adopting a graphical approach to causal inference. Given these benefits, there is a good case to be made that economists should adopt HCM or PCM instead of RCM. However, not everyone agrees. Both Angrist and Imbens are vocal proponents of RCM and have written textbooks that exclusively rely on RCM (Angrist and Pischke 2009; Imbens and Rubin 2015). In an article published a year before he won the Nobel Prize, Guido Imbens criticises PCM (and indirectly, HCM), claiming that RCM is better suitable for empirical practice in economics—while acknowledging that the graphical approach “has not had as much impact in economics as it should have” (2020, 1130). Given the similarities between HCM and PCM, many of Imbens’ criticisms of PCM apply to HCM as well. On the other hand, Heckman and Pinto (2022b) argue that HCM is a more suitable framework for economists than both RCM and PCM. This section summarises the debate and argues that the latest advances in Heckman’s camp give the graphical approach an edge over the others.
The 2020 article by Guido Imbens is a great overview of the arguments in support of RCM. First, contrary to what I have argued above, Imbens claims that the formulation of key assumptions is, in fact, more intuitive in RCM than in graphical frameworks. According to Imbens, the RCM formulations “capture the way researchers think of causal relationships” (2020, 1130). Second, RCM is claimed to connect more easily to traditional economic models such as the supply and demand model. Interestingly, Heckman and Pinto make the exact opposite claim, arguing that RCM as well as PCM “have significant limitations when applied to the wide variety of problems that economists face” (2022b, 894). Third, while Imbens acknowledges that PCM is advantageous for complex models with many variables, he claims that such models “are not particularly popular in empirical economics” (2020, 1155). Fourth, RCM is useful for dealing with the problem of treatment effect heterogeneity. Fifth, RCM is claimed to connect better with many practical questions of causal study design and the inference of causal effects.

The sixth and most forceful reason for preferring RCM (in my opinion) is that it is better capable of capturing the assumptions required for some causal methods. By reasoning about probabilistic independencies directly—bypassing considerations of structure—RCM has undoubtedly allowed methodologists to discover methods that would otherwise be overlooked because they seem improbable if you have a graphical perspective. Instrumental variable methods—the LATE in particular—are an example of this. With causal structures as in figure 1, the effect of $T_1$ on $Y_1$ can only be identified given additional non-graphical assumptions such as the no-defiers assumption. This is recognised by the others in the debate as well (Pearl 2009, 90; Heckman and Pinto 2022b, 913).

However, James Heckman and Rodrigo Pinto’s (2022b) recent work demonstrates that HCM is in fact extremely versatile. It is capable of formulating the assumptions needed for instrumental variable methods such as LATE, as well as those needed for methods from PCM, such as front-door and back-door adjustment. Hence, the most apparent advantage of RCM, that it has a natural way of explicating assumptions needed for instrumental variable methods, may no longer be a relative advantage compared to HCM. At the same time, HCM has all the advantages of PCM by virtue of incorporating graphical models, as I illustrated in the previous section.
V. CONCLUSION: **ALL ECONOMETRICIANS SHOULD ADOPT CAUSAL FRAMEWORKS**

Hence, based on the most recent developments, it seems to me that HCM has an edge over the other frameworks. It is versatile, suitable for many empirical methods in economics, and deeply rooted in economic tradition. However, Heckman and Pinto may go a bit too far when they say that the use of RCM and PCM by economists has been detrimental:

> Many econometricians and applied economists now emulate what they read in statistics or computer science journals. They have forgotten or never learned their own field’s foundational work to the detriment of rigorous causal policy analysis.\(^3\) (Heckman and Pinto 2022a)

The above claim is somewhat misleading, given the serious problems with the econometric approach as taught in textbooks and still practised today. This tradition is *responsible* for the problems in econometrics that became apparent in the 1980s. Causal frameworks such as RCM, on the other hand, have greatly contributed to the development of sound causal methods in econometrics. Heckman and Pinto may mean that the field’s founders from which they draw inspiration, such as Haavelmo and Frisch, had a better (and causal) understanding of structural equation models than what is found in textbooks. They are right about that, but this older tradition was forgotten or corrupted in the later 20th century (see Pearl 2009, section 5.1.2). Moreover, authors within PCM and RCM also claim to be inspired by Haavelmo’s work. It may be more accurate to say that all present-day causal frameworks draw on early 20th-century work, while none of the current causal frameworks can claim to stand in a continuous tradition from then until the present.

Both the RCM and HCM sides of the debate now seem to agree that the textbook definition of econometric exogeneity is inadequate, preferring alternative concepts from the newer causal frameworks.\(^4\) While Heckman’s earlier causal framework still relies on econometric exogeneity (Heckman 2005), Heckman and Pinto’s recent version no longer makes

---

\(^3\) The published version Heckman and Pinto (2022b) makes a similar point.

\(^4\) Angrist, Imbens, and Rubin say about exogeneity: “Typically the researcher does not have a firm idea what these disturbances [error terms] really represent, and therefore it is difficult to draw realistic conclusions or communicate results based on their properties” (1996, 446). Imbens calls econometric exogeneity “inadequate” (1997, 93). Heckman and Pinto say: “Econometrics textbooks often discuss causality as a property of an estimator, usually ordinary least squares (OLS). This approach reverses the logic of causality. It also generates confusion, since the OLS model is described by statistical assumptions that are void of causality” (2022b, 896).
any references to econometric exogeneity (Heckman and Pinto 2015, 2022b). This is a clear way in which all frameworks depart from the econometric tradition. It is also a good thing. Although econometric exogeneity when defined precisely and in structural terms can be a helpful concept—as argued by Pearl (2009, 169–170)—economists on both sides are abandoning the ambiguous textbook definition of econometric exogeneity and replacing it with clearly defined causal assumptions.

All of the causal frameworks on offer are a significant improvement to the field of econometrics. What is troubling, however, is that the textbook approach to econometrics is largely unchanged. The typical econometrics textbook has as its foundation the OLS regression model and the econometric exogeneity assumption, while RCM might be discussed much later as an afterthought (see e.g., Wooldridge 2010; Greene 2018). These textbooks have one important improvement compared to earlier days: they recognise that causal identification is the fundamental problem that economists are concerned with. For example, the first sentence of the introduction in Wooldridge reads: “The goal of most empirical studies in economics and other social sciences is to determine whether a change in one variable, say \( w \), causes a change in another variable, say \( y \)” (2010, 3). Hence, it is surprising that these textbooks take an approach that the leading experts on causal inference in the field—including Heckman, Imbens, and Angrist—recognise as inadequate.

The textbook approach has consequences for econometric practice. For the world’s star economists, causal frameworks might not be absolutely essential. After all, Angrist was able to produce interesting and credible causal studies in the early 1990s without relying on RCM. However, he was doing so at a time in which econometric research was widely believed to be incredible by its own practitioners. Causal frameworks are essential for the standardisation of credible causal methods and for bringing these methods to a larger group of researchers.

Moreover, it can be shown that practising economists make mistakes as a direct result of the confusion created by the concept of econometric exogeneity, as I do in Ackermans (2022, appendix A). There I discuss a complicated type of sensitivity analysis invented by economists to estimate the size of causal bias. However, the method is incapable of improving the estimate of causal bias already assumed by the researcher’s choice of parameters. These kinds of useless mathematical exercises can be avoided if modern causal frameworks are used as the foundation of training and practice in econometrics.
Why is progress in econometric education so slow? Perhaps the field simply needs more time. But one factor must be that economists cannot agree on which causal framework should be adopted. Without a consensus on this matter, textbook authors have little incentive to overthrow the approach they have taken for decades and which is currently used more widely than any of the modern causal frameworks.

Like many in the debate, I have strong views on the respective merits of the different frameworks. However, what is more important than which causal framework to adopt is that a causal framework is adopted—since PCM, HCM, and RCM are all big improvements over textbook econometrics. The profession should resolve the dispute about causal frameworks and update its graduate teaching. That is the only way to solidify the advances in causal methodology made by Angrist, Imbens, and others and assist future generations of economists in further advancing their work.

REFERENCES


**Lennart B. Ackermans** is a PhD candidate at Erasmus University Rotterdam. He works on philosophy of science, epistemology, and ethics.

Contact e-mail: <ackermans@esphil.eur.nl>
Reflections on the 2022 Nobel Memorial Prize Awarded to Ben Bernanke, Douglas Diamond, and Philip Dybvig

JENS VAN ‘T KLOOSTER
University of Amsterdam

Our early 21st-century lifestyles will have bleak consequences for humanity, with future generations facing ever more deadly heatwaves, droughts, floods, extreme precipitation, species extinction, and the collapse of ecosystems (Pörtner et al. 2022). As large parts of the planet become less habitable, knock-on effects include mass migration, famines, civil unrest, wars, and societal collapse. As António Guterres, the UN secretary general, described the IPCC’s ongoing sixth assessment cycle, it offers “an atlas of human suffering and a damning indictment of failed climate leadership” (UN 2022). To avert the worst consequences of climatic disruption and environmental degradation, the global economy will need to go through a dramatic transformation. However, if such a transition is forthcoming, we are certainly only in its earliest stages.

Around the world, financial policymakers have made progress in meeting this epochal challenge by acknowledging that climate change is real and, hence, creates new risks for the stability of the banking system (NGFS 2019; BCBS 2021; NGFS-INSPIRE 2022). Banks and other financial institutions are dramatically misaligned with planetary boundaries, which exposes them to trillions in losses (ECB 2021; RAN 2022; Semieniuk et al. 2022). In 2015, Mark Carney’s Tragedy of the Horizon speech set out a project of developing voluntary disclosure standards that allows investors to better screen risk. A Taskforce for Climate-Related Financial Disclosures (TCFD) was set up that year. In 2017, the recognition that climate and environment-related risks are crucial for prudential policy became the founding commitment of a new Network for Greening the Financial System (NGFS 2019). By reorienting policy efforts towards aligning the financial system with climate and environmental policy, policymakers now pursue two objectives. They contribute to financial stability while

**AUTHOR’S NOTE:** I would like to thank Benjamin Folit-Weinberg, Anne Kervers, Tobias Pforr, Bart Stellinga, Matthias Thiemann, Koen van den Bosch, Erica Yu, and the participants in the EIPE Research Seminar of 13 March 2023 for their helpful comments.
also supporting governments in pushing forward a major economic transition of the 21st century. These policymakers did not receive the 2022 Nobel Prize for research on banks and financial crises, nor, for that matter, is any of their work mentioned.

Established in 1969, the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel (henceforth ‘the Prize’) has turned into the most prestigious prize awarded for human progress in the realm of social thought. Initially a project to promote the Riksbank’s vision for managing the Swedish economy, the Prize lends societal legitimacy and prestige to a specific “economic” way of addressing contemporary policy challenges. Who receives the Prize, and who does not, has a profound impact on research and teaching (Offer and Söderberg 2016).

Awarded this year to Ben Bernanke, Douglas Diamond, and Philip Dybvig, commentators have already noted that the choice and justifications put forward by the Committee for the Prize in Economic Science (CPES) goes against the grain of financial policy since the 2007–2008 financial crisis. Adam Tooze wrote that this year’s award:

Has the effrontery actually to celebrate one of the weakest dimensions of modern macroeconomic thinking—its extraordinarily limited ability to grasp the macrofinancial instability of modern capitalism. Rather than challenging the dogged refusal of the economics mainstream to take seriously thinkers who face the essential important of finance and its dangers for the modern world head on, it flaunts the tendency of the mainstream to ignore them. (Tooze 2022; cf. Mehrling 2022)

I argue in the following that the Prize Committee has largely failed to take account of the more profound lessons learned after the crisis. In its scientific background study “Financial Intermediation and the Economy”, the CPES promotes an account of the financial system and the sort of policies that it needs, which is out of sync with recent progress in financial policy. The Committee finds in the work of the laureates a vision of an economy where individual lenders promote prosperity and economic growth. Although the financial system is mostly assumed to achieve an efficient allocation of resources, policymakers intervene to prevent self-enforcing bank runs. Climate change, and in fact most societal problems, can be solved through Pigouvian taxes and redistributive transfers. In re-
cent years, however, policymakers have started to doubt this narrow con-
ception of the objectives and instruments of financial policy, increasingly
moving beyond the pivotal role assigned to individual investors by the
CPES.

After the 2007 and 2008 bank crisis, financial policymakers have ac-
quired a ‘macroprudential’ outlook that is less sanguine about the self-
regulation of financial markets. The core assumption of this outlook is
that individual investors are often unable to accurately price financial risk
because the probabilities of such risks materializing are unknown. This
is particularly the case for events that are infrequent or new, but highly
impactful. From 2015 onwards, policymakers have come to study Climate
and Environment-Related risk (C&E risk) against the background of these
ideas. It has become increasingly clear that C&E risk resists conventional
prudential approaches. Not only are widely used techniques for estimat-
ing financial risk oriented towards the past and, at best, effective over a
short time horizon, the information that is needed for these techniques
to work is inherently unavailable on the level of individual investors and
firms. In the context of a rapid environmental transition, the vision set
out by the Committee fails to account for the profound epistemic obsta-
cles that individual lenders face in navigating these types of risk. Where
policymakers have increasingly recognized that the Anthropocene raises
profound challenges for financial regulation, the 2022 Prize continues to
promote moribund ideas.

My reflections are structured as follows. First, I review the existing
literature on the political role of the prize in promoting academics who
favour market-based solutions to societal problems. Next, I turn to the
Committee’s scientific background study to identify the vision of the fi-
nancial system that informs the 2022 prize. As I illustrate, however, these
ideas have become largely obsolete. This sets the stage for discussing how
financial policy has also been shaped by a profound rethinking of human-
nature relations in the ‘Anthropocene’. First, I set out how the academic
disciplines of economics and finance continue to lag behind recent devel-
opments in the broader social sciences. Then, I discuss how regulators
have rethought their approach to the environment in recent years.

**The Politics of the Nobel Memorial Prize**

In line with Alfred Nobel’s testament, Nobel Prizes are awarded “to those
who, during the preceding year, have conferred the greatest benefit to
humankind” (NobelPrize.org 2023a). Prizes for physics, chemistry, and
medicine are awarded based on scientific progress (the “biggest discovery”), the Literature Prize goes to “the most outstanding work in an idealistic direction”, and the Peace Prize goes to the person who does the most to advance peace between nations (NobelPrize.org 2023a). The Prize for economics, a bit more opaquely stated, is awarded to “a person who has written a work on economic sciences of the eminent significance expressed in the will of Alfred Nobel drawn up on 27 November 1895” (NobelPrize.org 2023b).

The Prize awards specific ways of conceptualizing the economy and often singles out policy implications that increase the value of the researcher's work. Considering policy relevance is not specific to this prize, as recent prizes for medical advances and climate science demonstrate. What is more striking about the prize is that recipients often take policy positions that are controversial within the scientific field itself. Science prizes tend to go either to topics with a narrow epistemic interest (this year's prize for medicine went to sequencing the genome of Neanderthals), or to findings with relatively uncontroversial uses (this year's physics prize went to findings relevant for producing quantum computers, quantum networks, and cryptography).

What the other science prizes do much less frequently (although the prizes awarded for nuclear science come to mind), is reward findings whose practical relevance and societal value is itself a topic of scientific contestation. Climate policy may not be entirely uncontroversial within society, but it is so within earth sciences. Economics, in contrast, features pervasive disagreement over competing notions of efficiency, equity, private autonomy, and well-being. Economists, rightly, take positions on these issues, because that is a requirement to make their ideas policy relevant. However, it also comes with unavoidable partisanship, which any prize awarded to economic findings inherits by conferring prestige and authority to work premised on such an outlook. It is for this reason that Friedrich Hayek (1974) famously noted on accepting it that the Prize:

Confers on an individual an authority which in economics no man ought to possess. This does not matter in the natural sciences. Here the influence exercised by an individual is chiefly an influence on his fellow experts; and they will soon cut him down to size if he exceeds his competence. But the influence of the economist that mainly matters is an influence over laymen: politicians, journalists, civil servants and the public generally.
In their 2016 book *The Nobel Factor: The Prize in Economics, Social Democracy, and the Market Turn*, Avner Offer and Gabriel Söderberg seek to characterize the political role of the Riksbank’s prize. As they argue, the prize was set up by the central bank of Sweden as part of its efforts to promote more market-oriented economic policies domestically. By endowing a specific way of doing academic research with the scientific prestige of the Nobel Prize, the central bank raises the profile of economics relative to other social sciences. Offer and Söderberg approvingly cite Alfred Nobel’s great-great nephew, who notes, “The Economics prize […] is awarded as if it were a Nobel Prize. But it is a PR coup by economists to improve their reputation” (Offer and Söderberg 2016, 103). In promoting a specific subset of the social sciences, the central bank of Sweden seeks to promote a pro-market vision of its domestic economy that has, until then, failed to garner favour with the ruling Social Democratic Party.

Beyond the specific domestic context of Sweden, Offer and Söderberg situate the politics of the Prize in its promotion of the price mechanism as a way of solving societal problems. In their survey of the policy implications of the laureates' research programmes, they show that roughly two-thirds of the prizes have been awarded to economists whose methodological approach assumes that market allocations are socially desirable (Offer and Söderberg 2016, 121–24). Some of the recipients, most famously Hayek and Milton Friedman, use a theoretical perspective according to which “actual market outcomes could be regarded for practical purposes as being close enough to the optimal” (Offer and Söderberg 2016, 121). A second type of pro-market approach is that of “neoclassical Keynesian economists”, broadly capturing the orientation of this year's laureates:

These economists generally embrace the axioms of neoclassical economics which specify rational agents operating with complete knowledge. They also seek an efficient market equilibrium, but they identify circumstances in which markets can fail. (Offer and Söderberg 2016, 122)

This research programme is typically associated with US coastal universities, and indeed Bernanke earned a PhD from MIT in 1979, having been
supervised by the central banker Stanley Fischer, while Diamond and Dybvig were graduate students at Yale in the late 70s and both were supervised by the financial economist Stephen Ross.

Although the academy is only one site of reflection on economic policy, recipients of this prize always hail from an academic career. Notably, even Ben Bernanke, former Chairman of the Fed, received the Prize for his academic work. Moreover, amongst academics, the Prize has favoured US- and UK-based research by male academics in traditional economics departments, with the prize to Elinor Ostrom being the first time it was awarded to a political scientist. Deep understanding of the economy, this year’s Prize again demonstrates, does not consist in the ability to make good policy, but rather in setting out a formal model that starts with the behaviour of individual agents:

When the work of Diamond and Dybvig appeared in the early 1980s, the role of financial intermediaries had been discussed for a long time. [...] Diamond and Dybvig’s research provided logically consistent mathematical models, where the existence and structure of banks were derived rather than assumed. By providing formal models based on microeconomic foundations, the key assumptions and economic mechanisms were laid bare. (CPES 2022, 6)

In light of its societal stakes, I focus on the political significance of this year’s Prize by considering what vision of the economy the Prize Committee puts forward as worthy of science’s highest esteem. I have little to say about the narrow scientific contribution of this year’s laureates, or, in fact, about their broader work and achievements. Their academic work may have been original and contributed to shaping financial economics in the 1980s. Still, the political significance of this prize suggests that we ought to ask, first and foremost, whether their ideas should inform research and teaching today.

**The Prize’s Outdated Vision of the Financial System**

The Committee awarded this year’s Prize for what it considers three foundational insights into the social purpose of banks: maturity transformation, delegated monitoring, and the social cost of banking crises. The vision that the Committee promotes, by focusing on these three achievements, is profoundly at odds with the macroprudential turn in financial policy of the past decade.
First, the Committee credits Diamond and Dybvig (1983) for discovering the unique contribution that banks make to the economy in terms of maturity transformation. The Committee finds in their work a vision, illustrated in the accompanying cartoon, where banks collect deposits from customers and channel them towards productive investments (see Figure 1). In this way, the CPES suggests banks serve an important social purpose, promoting efficiency and economic growth (cf. Zingales 2015; A mischaracterization that policymakers have in the past years turned away from (Braun 2016). There is nothing wrong with the cartoon itself, which more or less accurately represents the vision set out by the Committee.

---

1 This attribution is contestable as a matter of intellectual history. To name just one example, James Tobin, who won the prize in 1981, sets out these ideas as the “new view” in 1963: “the essential function of financial intermediaries, including commercial banks, is to satisfy simultaneously the portfolio preferences of two types of individuals or firms. One on side are borrowers, who wish to expand their holdings of real assets—inventories, residential real estate, productive plant and equipment—beyond the limits of their own net worth. On the other side are lenders, who wish to hold part or all of their net worth in asset of stable money with negligible risk of default” (5f). Of course, Tobin is also wrong to call this view new since even academically it goes back at least to 18th century monetary thought (Arnon 2010). More fundamentally, it is hard to imagine a society with banks but without this basic insight into their function.

2
Bezemer et al. 2021). That role, however, is conceptualised without assigning a role to banks in creating money. Instead, the defining feature of banks is maturity transformation.

The role of banks is to aggregate the savings of investors and invest in long-term projects. Only a fraction of investors will actually need to exercise their option to withdraw their savings early, since only a fraction of investors will be subject to short-term liquidity needs. This makes it possible for the bank to meet the liquidity needs of short-term investors, while investing their savings in productive long-term projects. (CPES 2022, 19)

As the Committee takes Diamond and Dybvig to show, the bank’s role in providing customers with the ability to make payments on demand also comes with a risk, namely that all customers withdraw their savings at the same time. This dynamic can set into motion a panic:

The more agents that are patient start withdrawing their deposits at $TT = 1$, the less long-term investment will be left, and since the bank does not have enough resources at that time to pay everyone, nothing is eventually left for agents who wait until $TT = 2$: the belief is self-fulfilling, so it is in everyone's interest to withdraw early. (CPES 2022, 24)³

On this account, irrespective of how prudently they lend and how rational their customers, banks are at risk of self-fulfilling beliefs bringing about their demise.

Second, the Prize Committee credits Diamond (1984) with setting out a model that explains the social purpose of banks in terms of delegated monitoring. This second model introduces a production cost for gathering information about future cashflows of borrowers ($K$) and monitoring their behaviour ($D$). All lenders can produce $K$ and $D$, but this is relatively costly for lenders who invest smaller sums. Lenders can also delegate the production of $K$ and $D$ to banks, who then invest on their behalf. By pooling funds, banks benefit from economies of scale, which allow them to do risk screening on behalf of lenders while also diversifying risk by lending to a large number of firms. Again, the Committee emphasizes the specific

---

³ See Ricks (2016) for a critical discussion of the model.
vision of the economy implicit in the model. It has no direct policy implications, but rather sets out a financial system that is self-organizing, where banks produce value by gathering information. Market discipline ultimately incentivizes banks to do risk screening well:

As long as the intermediary performs its monitoring properly, it will always be able to cover its monitoring cost and give savers their required return, so deposits become risk-free. And the intermediary has the incentive to honor its obligations in order to avoid costly bankruptcy. (CPES 2022, 31)

The upshot here is that accurate information is crucial for the financial system to work properly. Financial intermediaries should be able to screen firms based on accurate information. The intermediary's customers, in turn, should be able to assess risk and return from their investment. If that information is accurately disclosed, this theoretical perspective suggests that markets can realize an optimal allocation of credit without further involvement from policymakers. Policymakers who want to change the allocation of capital in the economy can do so by imposing taxes and providing subsidies to entrepreneurs directly.

Finally, the Prize Committee credits Bernanke (1983) with further demonstrating the societal importance of banks by bringing out the damage caused by a banking crisis:

Financial intermediaries perform a valuable service by channelling savings to productive investments [...] by disrupting these services, the banking panics in the early 1930s was precisely what generated such a long and deep recession at the time. (CPES 2022, 33)

These insights again have clear policy implications, but the specific twist that the Prize Committee gives to them is almost identical to guiding objectives of the pre-2008 Basel II framework: to balance the economic cost of financial regulation against the potential output loss of a banking crisis (Tarullo 2008).

---

4 See Sullivan (2022) for a critical assessment of Bernanke’s academic contribution.
NEGLECTING THE MACROPRUDENTIAL TURN

The work of the laureates provides insights into the working of financial markets. That progress, however, is primarily one of realism relative to what is assumed as an intellectual baseline throughout the Committee's text: the self-regulation of markets. The possibility of a liquidity crisis requires explanation in terms of rational actors if investors are assumed to be rational. Financial intermediation requires explanation if investors are assumed to be omniscient. In this regard, it is probably true that “their discoveries improved how society deals with financial crises” (NobelPrize.org 2023c, 1). However, as Charles Kindleberger wrote to Bernanke in 1982, “[t]he necessity to demonstrate that financial crisis can be deleterious to production arises only in the scholastic precincts of the Chicago school” (Mehrling 2022). Bernanke more or less conceded that in 1983, writing that Kindleberger and Hyman Minsky had “in several places argued for the internal instability of the financial system,” but justified his own contribution in stating that “in doing so have had to depart from the assumption of rational economic behavior” (Bernanke 1983, 258).

From today’s vantage point, the vision of the economy that emanates from the Prize looks quaint. Its optimistic assumption about investors formed the intellectual background of the Basel II framework, where financial regulators imposed capital requirements to prevent excessive risk-taking, but otherwise assumed the efficiency of a market based-allocation of credit. By simply presenting the ideas as the state of the art in financial economics, the Committee assigns undue prestige to this outdated vision of financial markets. In fact, the sanguine approach to finance of that framework can easily be cast as an important cause of the 2007–2008 collapse (Bezemer 2010; Turner et al. 2010; FCIC 2011) and has since been slowly displaced by a new set of ideas typically described as “macroprudential” (Borio 2011; Baker 2018; Thiemann 2019; Stellinga 2020). Namedropped on page 57 of the CPES (2022) text, that approach is not accurately presented as continuous with the Committee's vision of the financial system, nor is it, unlike Basel II, simply “aimed at striking a balance between the negative consequences of disrupted credit networks emphasized in Bernanke (1983) and the various costs of intervention”.

In a very general outline, the macroprudential turn is about risks to financial stability and the broader economy that result from the inability of individual investors to estimate risk accurately. The perspective is rooted in John Maynard Keynes's writing on radical uncertainty and the
work on financial markets of Minsky and Kindleberger (Keynes 1937; Minsky 2008; Kindleberger and Aliber 2015). Alongside this academic heritage, very similar ideas can also be found among financial market practitioners and financial policymakers, although in these contexts often without these authors’ more pessimistic assessment of market-based finance (Baker 2020).

The core macroprudential belief is that individual investors are often unable to estimate certain types of financial risk, in particular those tied to infrequent or new but highly impactful events. The reason for this is that investors are neither rational nor particularly well-informed. Instead, when estimating risks and rewards, investors take the recent past as their reference point, which means that the likelihood of gains and losses is assumed to persist into the indefinite future. A period of relative financial stability (the pre-2008 ‘great moderation’ providing a case in point) leads to more leverage and increasingly vulnerable investment strategies. Evolving, path-dependent expectations explain not only why investment strategies adjust, but also why new money-like assets can emerge, and regulation becomes less strict. Since there is at best a limited rational basis for financial market valuation, expectations can shift suddenly. A small event can tip the system from mania into panic; at this point the ability to create new forms of money turns out to be a crucial driver of crises. A depression ensues where investor pessimism takes hold, again with only a limited rational basis, but devastating nonetheless.

It is these ideas that have shaped global financial policymaking after 2008. These ideas, as we will now see, also have important implications for the role of the financial system in the transition to net zero.

**CLIMATE CHANGE AND ACADEMIC ECONOMICS**

The term ‘Anthropocene’, coined by the 1995 Nobel Prize for Chemistry laureate Paul Crutzen, marks the recognition of the fact that human civilization, including the past centuries of industrialization, took shape within a period of exceptional climatic stability (Crutzen 2006). Even for advanced capitalist societies, their social and economic processes remain deeply reliant on natural conditions and finite resources. A major insight of 20th century science is that the Earth system can be shifted into new states that are much less favourable to human civilization (Steffen et al. 5)

---

5 Minsky and Kindleberger briefly show up on page 48 but it seems the ideas that the Committee ascribes to them are actually Irving Fisher (1933)’s theory of debt deflation.
Growing greenhouse gas emissions, nitrogen deposition, biodiversity loss, the introduction of novel entities, and other major perturbations are pushing Earth out of its current state, permanently damaging the system’s ability to accommodate human life. The European social sciences evolved taking these environmental conditions as given, theorizing society as developing against the background of unchanging natural conditions. Recognition of these fundamental environmental blind spots of academic research has set into motion a profound rethinking of research and teaching (Bonneuil and Fressoz 2017; Malhi 2017; Charbonnier 2021).

The discipline of economics has been slow to take these broader developments on board and its engagement with the earth sciences and their implications remains superficial. Academic economists have typically theorized environmental impact as an externality and prescribed taxation of greenhouse gas emissions as a sufficient instrument for addressing it. Infamously, William Nordhaus received the 2018 Economics Prize in part for work from 2017, according to which an optimal carbon tax should steer the world towards 3.5°C of warming (CPES 2018, 37; cf. Kelleher 2019). Nicholas Stern, who worked since the mid-2000s to demonstrate the empirical and normative limitations of these ideas, did not receive the prize.

Setting aside these extreme positions, the attitude of the academic discipline to climate change is best characterized as indifferent. Until 2022, the most-cited journal in the field, Quarterly Journal of Economics, had not published a single article on the topic (cf. Stern and Oswald 2019). Similarly, a study of 21 leading finance journals finds that between January 1998 and June 2015:

Only 12 articles (0.06%) are related in some way to climate finance. The three elite finance journals (Journal of Finance, Journal of Financial Economics, and Review of Financial Studies) did not publish a single article related to climate finance over the 17.5-year period. (Diaz-Rainey, Robertson, and Wilson 2017, 243; cf. Kouwenberg and Zheng 2023).

---

6 Although the efficacy of such taxes remain debated, the case for a just transition premised solely on a fixed price per kilo of CO₂ emitted looks shaky (Tank 2020; Känzig 2021; Huwe and Frick 2022) as does the political feasibility of such a policy (Jenkins 2014; Breetz, Mildenberger, and Stokes 2018; Mildenberger et al. 2022).

7 It did publish an article on “Climatic Change and Agricultural Exhaustion as Elements in the Fall of Rome” (Huntington 1917).
These attitudes to climate change reflect a milder form of climate denial in the sense that even those alarmed by impending catastrophe may go about their daily work as if they are irrelevant to their concerns (Jonas 1984; Purdy 2015; Latour 2017).

This year’s Prize again reflects the discipline’s negligent attitude towards climate and the environment, topics which remain undiscussed in the 60 pages of the scientific background study drafted by the Committee (CPES 2022). This insouciant attitude contrasts starkly with the progress made by financial policymakers.

MACROPRUDENTIAL POLICY AND CLIMATE CHANGE

Policymakers, particularly in Europe, have increasingly recognized the pivotal significance of the climate and the environment for their work. In a way that defies justification through simple microeconomic models, at once opaque and pragmatic, financial regulation today combines a prudential concern with a more direct consideration of environmental impact. It is this development, and organizations like the Network for Greening the Financial System promoting it, that should have received a 2022 Nobel Prize awarded ‘for research on banks and financial crises’.

In 2011, the NGO Carbon Tracker published the study “Unburnable Carbon: Are the world’s financial markets carrying a carbon bubble?” (Campanale, Leggett, and Leaton 2011). The report begins by reviewing the climate science concerning the levels of CO₂ emissions that are compatible with warming below 2°C. The report uses this information to calculate a carbon budget for the planet. It then points out that financial institutions and governments owned fossil fuel reserves which, if burned, would amount to over five times the possible emissions. A sudden write-down of these assets would result in huge losses, losses in fact much greater than those which brought down Lehman Brothers. Financial regulators, the report argues, should act to ensure the relevant financial risks are adequately monitored:

The recent financial crisis has shown that capital markets were not-self-regulating and required unprecedented intervention; regulators were not monitoring the biggest systemic risks and so missed key intervention points. (Campanale, Leggett, and Leaton 2011, 3)

---

8 This is particularly surprising as two committee members, John Hassler and Per Strömberg, have published extensively on these topics.
In framing the challenge of climate change in terms of a gradual build-up of risk in the system, the 2012 Carbon Tracker report explicitly links environmental concerns to macroprudential objectives.

The report set into motion a gradual process of institutional learning within the financial policy community (Quorning 2023). In 2015, Mark Carney (2015) distinguished two ways in which climate change and the environment expose the financial system to potential losses not adequately monitored at the time (so-called C&E risks). First, the physical consequences of pushing the earth into a new systemic state might cause large losses to financial institutions due to floods, droughts, extreme weather, and the like. To avert these catastrophic effects, policy action would need to shift the economy dramatically towards a net-zero carbon emissions economy. In turn, such a shift means that financial institutions should anticipate how their investments will be impacted by government intervention as a consequence of climate policy.

The idea that climate change is a source of financial risk introduces the Anthropocene into economic thinking but does so in a way that speaks to regulators raised on Diamond and Dybvig. For one, its starting point is that action to prevent climate change should focus on accurately pricing financial risk and avoiding moral hazard. Banks today, this account suggests, take on risk because their creditors are not sufficiently informed about the risks they are taking. Consequently, basic facts that are well-established in the earth sciences have direct implications for financial policymakers, even within the existing legal framework. Moreover, this framing orients solutions towards improving market functioning. Although it implores financial policymakers to act rather than just wait for a carbon tax to do its work, their interventions are limited to market-correcting methods.

Can investors, properly incentivized to screen risk, ensure that the financial system serves our present needs? If the financial system does indeed work like the Prize Committee suggests, that seems plausible. Banks would be lending to fossil fuel companies because there is either insufficient information or climate policy is simply not credible but banks themselves would not be part of the problem. Despite increasingly ambitious climate targets, improved disclosures, and regulatory guidance, however, banks remain deeply locked into the financing of new fossil fuel infrastructure. Even the banks that are part of Glasgow Financial Alliance
for Net Zero (GFANZ), which implies a commitment to complete decarbonization by 2050, have lent $269 billion to 102 major fossil fuel companies (McCully 2022). As the ECB observed in 2020, the European Union’s (EU) largest banks “do not have the tools to assess the impact of climate-related and environmental risks on their balance sheet” and “only a small number of institutions have fully incorporated climate-related and environmental risks into their risk management framework” (14).

Below the surface of the orthodox account, regulators are increasingly acknowledging the limitations of an approach focused on individual investors and adequate disclosures (Bolton et al. 2020; Ameli, Kothari, and Grubb 2021; Dennis 2022; Oman, Salin, and Svartzman 2022). In this regard, how regulators see their task has been profoundly shaped by the macroprudential perspective (Klooster 2022; Quorning 2023). Against the background of a more realistic understanding of investors as geared to the near past and subject to fads and manias in view, it became clear that the financial system itself can be an obstacle to decarbonization and adaptation. From a macroprudential perspective, there are three reasons to doubt the role of individual investors in producing $K$ and $D$ for C&E risk, even if they are made subject to more market discipline (Dikau et al. 2022).

First, as a category of losses that have not historically occurred, the potential consequences of climate and environmental risk are not of the kind that individual investors can accurately estimate. There is no historical precedent for the climate transition and the catastrophic physical consequences we are facing, placing profound limits to the objective basis for risk estimates (Lenton et al. 2019; Sharpe and Lenton 2021). The regulatory approach in which individual investors could be entrusted to screen risk is premised on essentially backward-looking methods for estimating risk. The risk management techniques that banks and global banking regulation have come to rely on from the 1990s onwards, relies on historical data to estimate probabilities of default (Lockwood 2015). Key obstacles for applying such methods to C&E risks include the complexity and interconnectedness of financial institutions, the evolving economic environment and government policy, as well as the importance of the non-linear impact of tipping points in the climate system itself, as well as the policy transition.

Second, concerning the time horizon of C&E risks, potential losses from investments made today often only materialize years into the future, far beyond the business, credit, and policy cycles that guide behaviour of
individual investors. For this reason, supervisors are trying to move beyond the typical prudential time frame of three to five years and work towards a time frame of at least ten years (e.g., EBA 2020). The reason that prudential supervision focuses on that short time frame, however, is that going far beyond it is simply not possible for individual investors. To align the financial system with such longer time frames, policymakers need to set out the type of future that investors should plan for. It is for this reason that supervisory practice is increasingly gravitating towards stress testing, scenario analysis and, more recently, has resulted in European Union and United Kingdom efforts to introduce transition planning into their prudential framework (Elderson 2021).

This brings up a third challenge, which takes on a much more severe form than the risk-frame suggests, namely that the assessment of C&E risk is hampered by the availability of data (EBA 2020; BCBS 2021; TCFD 2021). On a basic level, simple metrics such as scope 3 emissions and energy performance certificates provide some guidance for risks. The extent of alignment of most investments depends, however, on what other actors do. Sustainability is a feature of an economic system, not of an individual investment or firm (Krahé 2021). For example, decisions made in the allocation of capital within the energy sector have profound implications for other sectors. Similarly, a successful transition in the transportation sector is crucial for production to remain organized on a global scale. Even the most sophisticated models require strong assumptions concerning the speeds of transition, the types of policies enacted, the willingness of consumers to change their lifestyles, and the effectiveness of carbon capture technologies (Monasterolo, Nieto, and Schets 2023).

In the face of these big unknowns, individual investors and firms are unable to produce the information needed to identify financial risk, let alone decide what allocation of investment is optimal. As a proxy for unmeasurable C&E risk, regulators have turned to setting financial policy based on a criterion of double materiality (Boissinot et al. 2022). Policymakers no longer focus solely on often unmeasurable expected losses (single materiality), but also consider the ways investments impact the environment (Smoleńska and van ‘t Klooster 2022).

**CONCLUSION**
The 2022 Prize rewards defunct ideas on the financial system and its regulations. It celebrates a centuries old vision of the financial system where financial intermediaries, nudged forward by gentle regulation, channel
savings towards productive investments. The role that the Committee ascribes to policymakers is limited to preventing self-enforcing bank runs, while allowing investment to be guided by the profit motive. From the perspective of the macroprudential turn, these ideas look outdated. This perspective also allows us to see that individual investors and firms are not able to set out the direction of travel for a fast-paced environmental transition. That is not to say that investors and firms have no role to play or that there are no big open questions concerning where to situate the benefits of market-based coordination. These questions should be at the forefront of academic research, and they quickly lead into the broader topic of economic growth and its pursuit through a market-based allocation of resources. Beyond climate change and the environment, the organisation of a financial system is also crucial for tackling problems like economic inequality, tax evasion, and market concentration, which should come together into what is still nebulously referred to as a rapid and just transition (Wang and Lo 2021; Newell, Geels, and Sovacool 2022). In the face of humanity's towering challenges, the Prize Committee fails to acknowledge most issues that are high on today’s policy agenda. It is hard to see how this year's prize in the memory of Alfred Nobel matches the eminent significance expressed in his will.

REFERENCES


Oman, William, Mathilde Salin, and Romain Svartzman. 2022. “Contending Views on the Role of Central Banks in the Age of Climate Change: A Review of the Literature.” In


Jens van ’t Klooster is Assistant Professor for Political Economy at the Department of Political Science of the University of Amsterdam. He is also currently a visiting fellow at the LSE Grantham Research Institute on Climate Change and the Environment. He holds a PhD in Philosophy (University of Cambridge, 2018) and in Economics (University of Groningen, 2021). His research is multidisciplinary and focused on the governance of financial markets, with a specific focus on how climate change and new macroprudential ideas are reshaping the role of central bankers and banking supervisors.

Contact e-mail: <j.m.vantklooster@uva.nl>
Review of Conrad Heilmann and Julian Reiss’ (eds.)  
*The Routledge Handbook of Philosophy of Economics.*  

D. WADE HANDS  
*University of Puget Sound*

The field of the philosophy of economics has clearly expanded rapidly over the last two decades and Heilmann and Reiss wanted their handbook to reflect these changes. As they say in their introduction:

The sheer number and quality of contributions and contributors we see in conferences, journals, and book publications and the proliferation of first-rate study programs are all testimony to the simple fact that the philosophy of economics is now a field in its own right. Philosophers of economics are everywhere: the leading academic programs and institutes in the field have grown, and there are more of them too. Philosophy of economics has arrived. (2)

But the increased volume of research and the number of contributing scholars are not the only motivations for *The Routledge Handbook of Philosophy of Economics*. It was also a response to the expansion of the set of philosophical ideas at work in the field, as well as changes in both the field’s disciplinary orientation and the demographics of its contributors.

I will elaborate on these motivations and many of the details of the volume below, but I think it is useful here to note that it is a large and wide-ranging collection. It has thirty-six total chapters opening with the editors’ introduction in chapter one. The remaining thirty-five chapters are grouped into eight broad categories (Parts I-VIII)—Rationality, Cooperation and Interaction, Methodology, Values, Causality and Explanation, Experimentation and Simulation, Evidence, and Policy—with each part containing four or five chapters. As one would expect from a volume with this range of topics, there are many chapters presenting material that could easily been assigned to more than one of the eight parts. In general, I think the editors did an excellent job dividing up the chapters into the relevant parts and I doubt it was an easy task.

This volume is one of many handbook-style volumes on the philosophy of economics published in recent years. Others include Kincaid and
Ross (2009), Mäki (2012), and Kincaid and Ross (2021) and it is useful to contrast these volumes with the Heilmann and Reiss volume. While these three handbooks differ in many ways, they were also quite similar in one respect. The vast majority of chapters in these volumes examined topics associated with what economists have traditionally called ‘economic methodology’ and philosophers have typically called ‘philosophy of economics’, which is, broadly, research on the intersection of philosophy of science and economics. Of course, different authors and different editors can, and often do, characterize economics and philosophy of science in quite different ways. This means that there is a lot of variation among books and chapters, but broadly speaking, the recipe for ‘philosophy of economics’ is roughly the same for these other handbooks: combine some part/parts of ‘economics’ with some part/parts of ‘the philosophy of science’ and blend.

Although research on these traditional epistemological and methodological topics still define the subject matter for the majority of the chapters in the Heilman and Reiss volume, there are also quite a few that examine topics quite removed from the traditional philosophy of economics literature. Of course, some of the differences between this volume and the earlier 2009 and 2012 handbooks has to do with changes in economics itself. For example, topics such as behavioral welfare economics and its relationship to public policy (Lecouteux, chap. 4; Grüne-Yanoff, chap. 35) and randomized controlled trials (Favreau, chap. 25; Khosrowi, chap. 27) have only recently begun to receive wide-spread attention. Although changes in economics certainly matter, Heilmann and Reiss emphasize a number of other recent developments that helped guide their choice of the handbook’s topics and contributors.

The first is a significant change in the demographics of the field: “In regard to younger generations and recent developments, our Handbook features, deliberately so, a broad range of scholarly generations” (2). To this, one could also add greater diversity among the authors with respect to a variety of other characteristics in addition to generational differences. Secondly, there have been changes in disciplinary affiliation. In particular there has been “a ‘philosophical turn’ in the relationship between philosophy and economics: currently, the academic discipline of philosophy hosts the main study programs, scholar, journals, and conferences” (3). This recognized, they also go out of their way to note that the increasing one-sidedness is not necessarily a good thing; it “is an empirical fact that we assert, not a value we hold” (3). The third involves a broadening
of the set of philosophical ideas involved in the philosophy of economics. In particular, “value-related research, whether it be on ethical aspects of economics, formal ethics, or economically informed political philosophy, has gained in relevance” (3). Finally, Heilman and Reiss emphasize the movement away from Big-M methodology (economic methodology employing grand philosophy of science such as positivism, Popper, instrumentalism, etc. and generally targeting grand economic theorizing such as neoclassicism, New Classical Macroeconomics, etc.) and toward small-m methodology (applied philosophical inquiry directed at practical concerns and more narrow areas of economic theory and practice). This change was to some extent taking place in the 1990s and was given some attention in the three other volumes discussed above; it is also a change that is very well-documented by the empirical bibliographic and network analysis of Claveau et al. (chap. 11).

Given that the volume has thirty-six chapters involving forty-one different authors, it is unreasonable to give a summary of each chapter in the space of a review; it is also unnecessary since Heilmann and Reiss provide an excellent summary of each chapter in their introduction. That said, it also seems like some discussion of at least some chapters would be useful for this review. Given that I also want to discuss some of the details of the volume with respect to the four changes in the philosophy of economics literature the editors identified, I will begin with some commentary on these changes and then, toward the end of the review, discuss a few chapters that I found to be of particular interest.

It seems clear to me that the demographics of those working in the field of philosophy of economics has changed—there is more interest on the part of younger scholars, but there is also, I would say, more demographic smoothness (for want of a better term). In the 1980s and 1990s there seemed to be a group of junior scholars and also a more senior group, while in recent decades the demographics appear to be more evenly spread out and there is less correlation between the set of topics an author discusses and their years of professional experience. There also seems to be more collaborative research going on as well as more diversity in gender, ethnicity, and regional affiliation. While these latter changes may be a characteristic of academic research in general, it is still an important change that has an impact on the type of research that gets produced in the philosophy of economics. I believe that, all in all, the ed-

1 The terms Big-M and small-m were popularized by McCloskey (1994).
Editors were quite successful in bringing together ‘a broad range of scholarly generations’, as well as fostering diversity along other dimensions, and these are worthy accomplishments.

The disciplinary affiliation of the philosophy of economics has clearly moved from economics to philosophy, primarily but no longer exclusively, philosophy of science. As the editors say, there has been a philosophical turn, but they also emphasize that they are merely reporting, not endorsing, this fact. It is difficult to see how having fewer economists thinking and writing about the epistemological issues associated with economics would be a good thing, but there are both costs and benefits associated with change in the disciplinary affiliation of the field. Consider the decision-making of a student with an interest in an academic career which combines economics and philosophy. On the economics side there are many professional benefits to being an economist. For example, economists generally receive higher compensation as well as “the opportunity to obtain income from consulting fees, private investment and partnerships, and membership on corporate boards” (Fourcade, Ollion, and Algan 2015, 89–90). The problem is that these financial benefits are much more likely to accrue economists specializing in areas like finance or applied microeconomics rather than to those specializing in economic methodology. Similarly, the most prestigious journals in economics almost never publish papers on economic methodology and the economics journals that do publish methodology are not ranked very highly in most economics departments. On the other hand, while most philosophers do not have the compensation or consulting options of economists, research in the philosophy of economics—at least in recent years—is well-received by prestigious philosophy of science journals as well as in most philosophy departments. Thinking about the disciplinary affiliation question in terms of the growth of knowledge in the field rather than in terms of private professional benefits of the scholars involved, there does not seem to be a clear picture of whether having economics or philosophy as the primary disciplinary affiliation is beneficial. One might think that the dominance of philosophers within the field would produce more research of a grand philosophizing sort (Big-M economic methodology) and that the dominance of economists would produce more research that is more practical and cost-benefit oriented (that is, small-m economic methodology), but historically that has not been the case. The recent increase in the influence of philosophers in the field has coincided with a movement away from Big-M and toward small-m research. This said, there are probably far
too many confounders at work during this period of time to know whether disciplinary dominance played a significant role. Perhaps the most reasonable take on the issue of disciplinary affiliation is simply that of the editors: acknowledge the (descriptive) fact that philosophers are now the majority of contributors to the field but remain neutral on the (normative) question of whether that is good or bad (for either the scholars involved or for the growth of knowledge in the field).

Broadening the set of philosophical ideas involved in the philosophy of economics is probably the most important thing that distinguishes this volume from the other three handbooks noted above and, for that matter, distinguishes it from the majority of the philosophy of economics literature in general (at least since the early 1970s).² It should be noted that while the philosophical net is cast much wider in this handbook than what we have grown to expect in the philosophy of economics, the widening is almost exclusively in terms of what the editors call ‘value-related’ research: specifically ideas traditionally associated with ethics and political philosophy. Although it is important to recognize that the philosophical net is wider, it may not be wide enough to satisfy some philosophers and/or heterodox economists. This said, I understand the editors’ decision to expand the set of philosophical ideas by highlighting the contact points that economics has with ethics and political philosophy. There are a number of reasons for this: it is clear there has been a lot more of this type of research going on in the field in recent years, it increases the relevance of the philosophy of economics for real world social and economic policy, and it reconnects contemporary philosophy of economics with ethical and political links that go back to Adam Smith and other classical economists, links that became severed during the heyday of positivist ideas.

The fourth change that Heilman and Reiss emphasize is the change from so-called Big-M methodological literature to small-m methodological literature. There is no doubt this change has taken place, it is a substantive change, and it has been underway for many years. Of course, there has been a turn away from grand theorizing in almost every academic discipline and this includes philosophy of science where universal general theories of scientific knowledge (for example, versions of positivism, Popper, Lakatos, van Fraassen, and so on) have generally been displaced by more narrowly defined and more naturalistic and pragmatic inquiries into

² See Hausman (2017) for a nice discussion of the history of economic methodology/philosophy of economics since the 1970s.
specific aspects of scientific knowledge. So too for economics where there is much less grand theorizing of the Walrasian sort (or on the heterodox side the Marxian sort) and much more emphasis on economics which is more empirical, more practically focused, and more context dependent. Behavioral economics has opened the door to a flood of economic research where outcomes are deeply context dependent and game theory has made strategy and multiple equilibria a standard feature of economic modeling. While these movements away from grand philosophizing and grand economic theorizing have undoubtedly had an impact on the way that contemporary philosophy of economics investigates economic theory and practice, small-m methodology is more than just the result of these two forces.

Heilmann and Reiss do not introduce the Big/small terminology until their discussion of Part III of the volume, the part specifically on “Methodology”, but it seems to me that it is relevant to most chapters in the other parts of the volume as well. The emphasis on Big-X is an emphasis on grand sweeping and/or universal ideas and topics while small-x is an emphasis on ideas and topics that are more applied, pragmatic, practical, and context sensitive, and these seem to be distinctions that can apply not only to different epistemological or methodological approaches, but to the more value-related research as well.

To get a feel for how small-m methodology and small-v (value related ethics and political philosophy) works in contemporary practice it is useful to discuss a few examples to see what such philosophy of economics research looks like. Fortunately, Heilmann and Reiss’ volume has plenty of nice examples. I will discuss three of these. The first two focus on methodological-epistemological issues while the third explicitly considers both epistemic and ethical concerns. Obviously three out of thirty-six chapters clearly neglects many chapters that could easily have served as exemplars just as well as the three I will discuss.

Guilhem Lecouteux’s chapter 4 on “Behavioral Welfare Economics and Consumer Sovereignty” explains the dilemma that led to the current debate in behavioral welfare economics. Traditional Paretian welfare theory uses individual preference satisfaction as the normative standard for what does and does not increase welfare. Of course, behavioral economics has demonstrated that individuals do not necessarily, or even systematically, act in a way that is consistent with acting optimally on well-ordered preferences. Thus the “aim of behavioral welfare economics (BWE) consists, then, of looking for strategies to recover a normatively satisfactory
notion of ‘economic welfare’ from the possibly incoherent choices of the agents” (57). This undermines both the more abstract welfare theorizing of the middle of the 20th century such as the First Fundamental Theorem as well as the foundational basis for basically all microeconomics-driven social policy. Lecouteux argues there is no straightforward solution to the dilemma of reconciling traditional Paretian welfare economics and behavioral economics. The path out need not abandon the importance of preference to welfare, but instead to recognize that the “normative problem does not lie in the preferences but in the process of preference formation”; that BWE should “acknowledge both the individual and the social determinants of our preferences and behaviors, rather than blaming the defective psychology of individual agent taken in isolation” (63). Not everyone will agree with this assessment of course, but it is a critical assessment based on a close examination of the specific details of the tension that exists at the intersection of behavioral and welfare economics and provides a constructive small-m suggestion about how the tension might be defused.

Another example is Judith Favereau’s chapter 25 on randomized controlled trials (randomized field experiments: RFEs). She reviews the rise of RFEs, particularly in development economics, as an experimental and policy-oriented area of economics that uses randomization to “test local development policy and do not rely on any theoretical framework” (343). She argues that the structure of RFEs basically transforms field experiments into laboratory experiments, which results in strong internal validity but weak external validity. Favereau provides a historical overview of field experiments in economics, discusses a number of specific examples in development economics and other areas, and notes many of the criticisms that have been raised about RFEs in the recent literature. She closes with a discussion of qualitative studies as a possible way to get around RFEs’ external validity problem, while preserving the benefits of their internal validity. Again, not everyone will agree with either the critique or the proposed work-around, but the chapter provides a nice, sharply focused discussion of the history and methodological problems of RFEs and provides a small-m suggestion about how the epistemic value and practical usefulness of the approach could be improved.

---

3 Favereau doesn’t call attention to the fact that advocates of RFEs frequently say that no theoretical framework is involved, but they must really mean that ‘no theoretical framework from economics is involved’. Obviously, one doing randomized controlled trials must employ theories of randomization, various statistical theories, and many other theoretical frameworks.
Grüne-Yanoff’s chapter 35 examines an important debate in contemporary philosophy of economics and does so by employing both small-m and small-v resources. Since it concerns behavioral public policy (BPP), there is a bit of overlap with chapter 4, but it is not significant since Grüne-Yanoff is much more directed at the policy side of the BPP debate: his goal is to provide a framework for improved evaluation of both the “effectiveness and ethical acceptability” (480) of BPP.\footnote{The expression ‘effectiveness and ethical acceptability’ is first introduced on page 480 but is used throughout the chapter.} He begins by discussing the diversity of ways to talk about and implement BPP—nudges and boosts in particular—and argues that such “diversity is problematic because it makes BPPs’ effectiveness and ethical evaluation context sensitive”, that is, “the same intervention is effective and ethically acceptable in one context, but not in another” (482). The problem is that since context matters, different policies affect “behavior through different causal pathways” (483), that is, the different causal mechanisms by which they operate. It should be noted that while he is advocating a better understanding of various causal mechanisms at work in any particular BPP, he is not employing the traditional ontic conception of mechanisms, but rather treating mechanisms as abstracting models (484). Returning to the two main BPP policies, nudges and boosts, he concludes “effectiveness and ethical acceptability of a BPP [...] can be derived from this mechanistic distinction between nudges and boosts” (491). While some finesse may be necessary to pass safely through the philosophical underbrush of non-ontic mechanisms, Grüne-Yanoff provides a detailed discussion of both the practical and philosophical issues that currently swirl around BPP and offers an improvement that works on both the effectiveness (epistemic) and ethical side of the debate.

Finally, in addition to the chapters specifically noted above, I would like to briefly say a few words about a sample of the additional chapters that I thought were particularly valuable contributions. They may be valuable because they addressed an important but relatively neglected topic, stood out as notably clear and coherent, handled the balance between surveying the relevant literature and making an original contribution particularly well, or frankly, just because I personally found them to be interesting contributions. They are, though, just a sample.

Moscati (chap. 1) is more historical than any other chapter but does an excellent job summarizing the long and many-faceted history of utility theory. Vredenburgh (chap. 5) is an excellent chapter on “the puzzling,
seeming co-existence of psychological and non-psychological interpretations of ‘preference’ in economics” (67); the chapter is primarily philosophy of economics but also does an excellent job with the historical discussion. Claveau et al.’s chapter 11 is not like any other chapter in the volume. It is an empirical bibliometric study of the economic methodology literature over three decades; it is a novel approach, and the results should be extremely interesting to anyone in the field. Chao (chap. 13) is a nice discussion of the literature on scientific representation with particular attention to economics; it ends by supporting “a pluralistic view of economic representation” (196). Chapter 14 by Vergara-Fernández and de Bruin is a philosophical examination of finance and financial economics—important fields of economics which receive little attention within the philosophy of economics literature. White’s chapter 17 surveys “the various ways that approaches other than utilitarianism have been incorporated into theory, policy, and practice” (237) and does so in a particularly clear and concise way. Chapter 22 by Verreault-Julien and chapter 23 by Jhun both address explanation in economics and, in particular, the relationship between how-actually explanations (HAEs) and how-possibly explanations (HPEs). This is an extremely difficult and contentious, but also important, question about the exact and/or proper relationship between idealized economic models and their associated explanatory targets. Neither author finds an easy or obvious solution—and nor do I—but both have suggestions for moving the discussion in a more constructive direction. For Verreault-Julien there should be more attention to the fact that “adoption of a specific account of explanation may also constrain what economics ends up being about” (308) and thinking more about the fit between explanatory account and economic explanandum may be necessary. For Jhun, there is “no algorithmic route from the model that provides a how-possibly explanation to the model that provides a how-actually explanation” (322) and she argues for the narrative approach of Mary Morgan and others to bridge the gap. Nagatsu’s chapter 24 introduces two typologies of economic experiments—standard and extended which “reveal the different epistemic and nonepistemic interests underlying experiments in economics” (329). He uses Francesco Guala’s distinction between ‘the tester’s tradition’ (individual level) and ‘the builder’s tradition’ (aggregate and interactive systems) and makes the case that psychology is associated with the former and economics the latter. Chapter 26 by Lehtinen and Kuorikoski addresses computer simulation in economics. It
seems that simulation should be an important tool and yet it is less important than one might expect. They examine three approaches to economic simulation and suggest that methods “that can be conceptualized simply in terms of computational solutions [...] are readily accepted in the mainstream methodological palette, whereas simulations that are more about directly mimicking dynamic processes [...] face methodological resistance” (365). Hausman’s chapter 31 examines the methodological issues connected with measuring the health improvements from medical treatments. The main points of his argument are that health cannot be directly measured, but it can be valued, and one of the primary ways to do that is to assign quantitative quality states. He discusses eight problems associated with these quantitative measures and “concludes that the quantification of health is a precarious endeavor” (434).

Overall, The Routledge Handbook of Philosophy of Economics is well-organized, timely, and an important contribution to the literature. There is always some unevenness among the chapters when a volume involves so many authors and so many topics, but that is inevitably the case, and the editors did a very good job keeping it as cohesive as possible. Although I appreciate the editors’ desire to include chapters that were philosophically more value-oriented (ethics and political philosophy in particular) as well as the more traditional philosophy of science-based contributions, it does make the organization more difficult than it would be for a volume with a narrower philosophical focus. Readers coming from the epistemology-methodology side may not have the requisite background for the value-oriented chapters, and of course vice versa. Nonetheless, it seems to work out, and the final product is an extremely valuable contribution to the literature. My experience has been that editing a handbook, or any other anthology, is a very difficult job and one that is best thought of in satisficing, rather than optimizing, terms.

REFERENCES

**D. Wade Hands** is Professor Emeritus in the Economics Department at the University of Puget Sound. He is author of *Reflection Without Rules: Economic Methodology and Contemporary Science Theory* (2001), editor (with John Davis) of *The Elgar Companion to Recent Economic Methodology* (2011) and (with Philip Mirowski) *Agreement on Demand: Consumer Theory in the Twentieth Century* (2006). He is a past president of the History of Economics Society and served for fifteen years as co-editor (with John Davis) of *The Journal of Economic Methodology*. His research crosses the fields of economic methodology, the philosophy of economics, and the history of modern economic thought. Contact e-mail: <hands@pugetsound.edu>

ellas needler
*Erasmus University Rotterdam*

Maria João Pimenta
*Erasmus University Rotterdam*

Robert Skidelsky’s 2020 book, *What’s Wrong With Economics?: A Primer for the Perplexed*, is clearly a guide for the curious student. More subtly, however, it can be read as a call to action for economic methodologists and philosophers of economics to expand their research agendas. To expand our research agendas along Skidelsky’s line would not only be for the benefit of economists-to-be but would improve the economics profession as a whole which spans far and wide—from academia to the public sphere to the private sector.

Skidelsky is best known for his three-volume series\(^1\) on the life and work of John Maynard Keynes. Nevertheless, aside from a brief mention in the introduction and the two concluding chapters, the book rarely relies on Skidelsky’s knowledge of Keynes, instead flexing his broad knowledge of the history of economics and economic methodology. The book is a broad and pluralist exploration of the many facets of economics for the curious, albeit concerned, economics student.

The book undoubtedly takes a broad approach to illuminating the potential for a pluralist and interdisciplinary economics. This is a strength of the book, especially for students, as each reader can find at least one, if not multiple, topic(s) to pique their interest and lay a foundation for further research and exploration. But, when reading through the whole book, this broad approach can create a slow and monotonous overview for those who are more familiar with these topics, likely the readers of the *EJPE*. Nonetheless, philosophers of economics and economic methodologists can find some value in chapters 13 and 14 where Skidelsky lays out the ontological and epistemological basis for his vision of economics.

\(^1\) This was first published as Skidelsky (1983, 1992, 2000), and then compiled and abridged into a one-volume edition as Skidelsky (2003).
This review has a two-fold aim: first, we summarise what curious students and outsiders can expect to gain from this book; and second, we consider what additional value the book holds for philosophers of economics.

I. FOR STUDENTS

Skidelsky shows great empathy to students of economics by (1) allowing them to question the heavyweight concepts of mainstream economics—its mascots, so to say—which they have not had the chance to challenge, and (2) unravelling the hopes and dreams they may hold for the discipline they chose to dedicate their academic life to.

Cleverly, the pace and flow of the book mimics the trajectory of learning an economics' student likely faces in their studies. In chapters 2–5, Skidelsky walks the reader through the basics of economics but diverges from the standard textbook approach by also presenting normative ideas from methodological and historical perspectives. Skidelsky starts chapter 2 by describing supply and demand through wants and means, continues to economic growth in chapter 3, and equilibrium and shocks in chapter 4. In chapter 5, Skidelsky finishes this primer by exploring modelling with a particular emphasis on (i) the role of ideology, similar to Backhouse (2010), (ii) the place of rhetoric in economics, by exploring McCloskey (1998), and (iii) addressing the age-old question: Is economics a science? Skidelsky concludes that it is exactly because economics is not a science that it needs to learn from other disciplines. This sets him up for the next part of the book.

In chapters 6–12, Skidelsky explores the added value for economics outside the discipline itself. Like chapters 2–5, which mimic an economics curriculum, Skidelsky starts in chapter 6 with psychology, and thus behavioural economics, which is likely familiar to young students in economics. From chapter 7 onwards, Skidelsky ventures further out. He takes the reader off the island of economics completely, and with each chapter, ventures from psychology to sociology (chapter 7), institutionalism (chapter 8), political science (chapter 9), the history of economic thought (chapter 10), economic history (chapter 11), and even philosophy through ethics (chapter 12). In each, the reader is told what valuable resources they can learn from each discipline to take back to economics. For example, from psychology, they learn to challenge rationality; from political science, the need to understand the role of power; from sociology, how to question individualism with an introduction to methodological holism.
Along with introducing new topics, Skidelsky’s venture to each different disciplinary island challenges the common mascots of economics. By mascot we mean the ideas which represent the core of mainstream economics. Rather than take these mascots at face value, Skidelsky helps the reader understand the core of the ideas and how they can be challenged. These mascots of the mainstream include concepts such as shocks, models, and rationality. Skidelsky walks the student through each of these mascots using insights from other disciplines. Skidelsky allows and encourages the reader to question these concepts and opens points of curiosity that a mainstream economics teaching may not allow.

Aside from recognizing students’ curiosity around concepts in mainstream economics, Skidelsky strikes a chord with the unlikely but ever-relevant emotional tie students may feel when studying economics. This may seem counterintuitive. Why are students emotional about a field that is largely reliant on rationality? The reason is many students were sold on the idea that economics can do something for the world (something Skidelsky remarks on in the opening line of the preface): That it is a discipline which has the power and the tools to help people and society. Many students come to economics and become invested in the discipline with the hope of solving the problems our generation currently faces, such as climate change, inequality, stagnation, and automation, to name a few (147). These are problems and concerns that mainstream economics has not been the best at providing solutions for. Skidelsky captures the sentiment well when he calls upon the reader saying, “current dissenters from established opinion need not feel lonely” (147). At 83 years old, Skidelsky recognizes these emotions in young students better than many younger, mid-generation economists who fail to see the problems of the paradigm they have inherited.

II. FOR PHILOSOPHERS OF ECONOMICS AND ECONOMIC METHODOLOGISTS
First off, readers of the EJPE will be pleased by Skidelsky’s ability to put his motto of interdisciplinarity into practice, which he explores in chapters 6–12. He engages with the giants, on which philosophers and methodologists of economics stand, including Rawls, Nozick, Sen, List, Arrow, Lakatos, and Kuhn. Skidelsky even engages with current and core practitioners in our field, including Hodgson, McCloskey, Angner, Blaug, Cartwright, and Davis. As mentioned, most interesting to the EJPE reader are the last two chapters, where Skidelsky brings together his overarching critiques of the economics discipline:
The two main problems we have identified in this book are related: insufficient generality of premises (epistemology) and lack of institutional mapping (ontology). We need a science which is more modest in its epistemology and richer in its ontology. (187)

In the penultimate chapter, Skidelsky identifies these two threads that have been running through the whole book. If most of the book is a great primer for students and outsiders, chapter thirteen—titled Retreat from Omniscience—is the part that offers most concrete insights for the philosophers of economics.

The first—epistemological—thread asks us to consider the insufficient generality of the foundational premises of mainstream economics, drawing out in particular how economics prizes utility maximisation. This premise constitutes the starting point of most, if not all, economists in the mainstream. To substantiate this epistemological problem, Skidelsky shows us his Keynesian blood and his decades of academic work on the venerable economist. The epistemological modesty Skidelsky asks of economics lies in the distinction between risk and uncertainty, something Keynes explores in chapter 12 of his General Theory (1936). Incorporating this Keynesian (as opposed to Knightian) distinction would entail that, rather than focus on utility maximisation under certain conditions, economics would instead model human behaviour under different conditions of knowledge, which may have varying levels of uncertainty. This move towards a Keynesian theory of rationality gives us a way to accept that which we cannot know and also that which we cannot represent formally.

The ontological—and dare we say deeper, though less obvious—thread points to economics’ lack of institutional mapping. Here Skidelsky says that a return to Keynes’ “underdeveloped” ontology would be insufficient (185). Mainstream economics’ ‘map’ of the world is populated only by individuals; no groups or institutions make it on the map in any substantial way. In its methodologically individualist ways, economics does not capture influential forces in the economy: those of ideas, technology, community-membership, history, power, etc. Currently, the map shows the micro-level of individuals, and perhaps firms, and the macro-level of government. However, it misses a whole level, namely the meso-economie level, which is composed of many intermediary institutions linked by intricate social networks.
Taking the book’s subtitle—a primer for the perplexed—seriously, we ought to ask, what next? We do not know if Skidelsky took inspiration from E.F. Schumacher’s (1995) *A Guide for the Perplexed*, but the titles and the sentiment of dissatisfaction with the ‘maps’ the mainstream provides us with bear an eerie resemblance. There is no doubt that when Skidelsky takes his reader off the island of economics to gather insights from other disciplines, he is not only asking the reader to bring the insights back to the lonely island but to use them to move the island closer to the mainland of the social sciences. Skidelsky asks us to consider the epistemological and ontological threads together in the following way: to move economics to the ‘goldilocks zone’, which is achieved by moving towards holism and inductivism at the same time (see figure 8, reproduced below).

![Figure 8: Different Approaches to Understanding (Skidelsky 2020, 188)](image)

Herein lies the crux for philosophers of economics. These two dimensions should be understood as paths for change. Moving along the quadrant may lead to more resistance than Skidelsky seems to give room for. In
fact, it raises the question of whether student dissatisfaction is enough to move economics towards this ‘goldilocks zone’. In the worst-case scenario, we could see a large, disruptive crisis to the real economic system leading the economics profession to the path of reflection it sorely needs. But is there a better way? Perhaps, if philosophers and methodologists of economics rise up to the challenge presented in this book and set the research agenda of their field to transform the discipline of economics. If they do not, then what is the point of the philosophy of economics?

**III. CONCLUDING REMARKS**

For students, this book is a staple for trying to quench curiosity and a great bridge from a bachelor’s in standard economics to economic methodology and the philosophy of economics. For those unaware of the philosophy of economics and economic methodology, Skidelsky eases the reader into the main concepts.

For existing practitioners in economic methodology and the philosophy of economics, while the book may prove slow at times, it is nonetheless pertinent. In the final chapters, we read Skidelsky’s book as laying down the challenges that economists and philosophers face to shift the mapping of economics. Skidelsky is asking for a research agenda around the philosophy of *pluralist* economics where some argue current philosophy of economics can be seen as the philosophy of *mainstream* economics (Hands 2015; Mireles-Flores and Nagatsu 2022). If economic methodologists are to take Skidelsky’s call to action seriously, they need to expand their research agenda outside the bounds of mainstream economics to truly challenge the existing paradigm.

**REFERENCES**


**Ella Needler** is a research master student at the Erasmus Institute for Philosophy and Economics and volunteer with the Women in Economics Initiative. Her research interests include the history of economic thought and economic methodology in the topics of macroeconomics, central banking, and gender. She holds a BSBA in finance and economics from Washington University in St. Louis.
Contact e-mail: <ella.needler@wustl.edu>

**Maria João Pimenta** is a research master student at the Erasmus Institute for Philosophy and Economics. Her research interests are in the history of economic thought and the methodology of economics. In particular, she interested in exploring the impact of vision on the practice of economics across time. She holds a Scottish Masters in Economics from the University of Edinburgh, United Kingdom.
Contact e-mail: <mjoaopimenta99@gmail.com>

STAVROS A. DRAKOPoulos
National and Kapodistrian University of Athens

The notion of self-interest has a very long presence in economic discourse and has played a central role in the development of mainstream microeconomics. Contemporary mainstream economic theorists do not refer to self-interest as much, but they assume human agents are rational in the Homo Economicus sense, and proceed with their theoretical or empirical investigations. The standard notion of rationality employed includes a number of assumptions concerning the behavior of economic agents which are not always spelled out. The modern meaning of economic rationality contains the assumptions that agents maximize their own utility only, or that preferences are atomistic or independent, or that utility depends on one’s own consumption only. In other words, self-interest is used as a basic, or at least baseline, behavioral assumption. As Wade Hands (2021) states in the introductory chapter of this volume:

This commitment to self-interest by economists not only started early in the discipline’s history, it has: persisted over the next two hundred-plus years of economic theorizing, continued to condition the way that economists characterize individual and institutional decision-making, and influenced the way that economists think about public policy questions. (2)

Apart from their central role in decision-making, self-interested agents are present in other fields of orthodox economic theory. The case of the representative agent in modern macroeconomics is indicative. For instance, Joseph Stiglitz identifies the place and role of the representative agent in most dynamic stochastic general equilibrium (DSGE) modeling, as follows: “DSGE models seem to take it as a religious tenet that consumption should be explained by a model of a representative agent maximizing his utility over an infinite lifetime without borrowing constraints.”
A GENEALOGY OF SELF-INTEREST IN ECONOMICS / BOOK REVIEW


Equally, there is a long tradition of criticism of self-interested Homo Economicus ranging from the German Historical school to Veblen and Keynes. Many contemporary prominent economists have also pointed to the serious weaknesses of the assumption of self-interested agents (e.g., Sen 2002). Further, notions that challenge self-interest (such as social preferences, reciprocity, fairness, or altruistic punishment) are increasingly employed by behavioral economists. In fact, one of the basic characteristics of this school of thought is to challenge the established model of economic rationality (e.g., Thaler 2017). As a result, the substantial rise of popularity of the new behavioral economics has revitalized the attention to the meaning of rationality in economics and by extension, to the role of self-interest.

The present collection of papers can be seen as a timely contribution to the above debate. It contributes to the discussion of the origins of Homo Economicus, and more specifically of the notion and role of self-interest in economic thought. There have been some attempts in the past to collect papers exploring the concept of self-interest. One which comes to mind is the edited volume by Jane Mansbridge (1990), which was oriented towards a critique of self-interest and argues for alternative formulations by economists and other social scientists. A more recent effort was put forth by Pierre Force (2003), with an emphasis on pre-Smithian economics. However, no volume has been written on a systematic history and role of self-interest in economic discourse, and thus this collection attempts to fill the gap as it covers a much wider time period.

This book comprises 18 chapters including an introduction. In the introductory section, Wade Hands argues that economics has always been concerned with self-interest and demonstrates a variety of different ways that economics and self-interest have been connected in the history of economics. By focusing on the examples of Ricardo and Marx, he also refers to how these relationships manifest themselves in the work of some specific economists (e.g., Daniel Kahneman, Amos Tversky, Richard Thaler, Cass Sunstein). The book is divided into three parts. The first part is entitled Self-Interest in Eighteenth and Nineteenth-Century Economics, and contains seven chapters. The material covered here includes the work of various pre-Smithian writers, including Smith himself, Bentham and Mill, as well as Walras and Wicksteed. The second part is on Self-Interest in Twentieth Century Economics, and concentrates on economists such as
von Mises, Knight, and Hayek, and scholars like Otto Neurath and Karl Polanyi, and includes a chapter on questions about free will, political philosophy, and political science. The third part is entitled *Contemporary Topics in Self-Interest and Economics*, and the subjects studied cover a very broad range, including altruism, game theory, neuroeconomics, and philosophical discussions of the relationship between economics and other disciplines.

The first two chapters of the first part examine the concept of self-interest in pre-classical economic discourse. Beginning with *Self-Interest and French 'Philosophie économique' 1695–1830*, Gilbert Faccarello and Philippe Steiner concentrate on the French economists of the eighteenth century and demonstrate their recognition of the positive results of self-interest as well as its role as a means of government. In *Self-Interest in the Thought of Adam Ferguson*, Craig Smith shows that Ferguson—a prominent member of Scottish Enlightenment—had developed a complex and nuanced understanding of the place of self-interest in moral and political life. The next three chapters focus on classical political economy starting with *Adam Smith on Self-Interest* by Shinji Nohara. Nohara reexamines the well-known Adam Smith problem which refers to the apparent differences concerning the nature of human behavior in Smith’s two major works: self-interest in the *Wealth of Nations* and mutual sympathy in the *Theory of Moral Sentiments*. Nohara maintains that the relationship among Smith’s works is the key to the understanding of the different human descriptions of human nature. In *Bentham on Self-Interest: Institutional Control of Self-Interest*, Hiroaki Itai argues that in Bentham’s system, the issue of whether human beings are selfish or altruistic is not relevant, since a utilitarian system of governance transforms people into agents of appropriate utilitarian accounting and maximizes social well-being. The chapter on *John Stuart Mill on Self-Interest: Focusing on His Political Economy and the Principles* by Yoshifumi Ozawa, is next. After elaborating Mill’s ideas about self-interest, Ozawa maintains that Mill endeavored to design political institutions which would reconcile the self-interested actions of individuals to the public benefit.

The next two chapters move on to the marginal revolution and to the early neoclassical economics by focusing on Walras and Wicksteed. In *Léon Walras on Human Nature and His Social Reform Plan*, Satoshi Takahashi attempts to clarify the relationship between Léon Walras’ notion of human nature and his reform policies toward a unique form of
socialism. The author also develops the interesting view that Walras expects egoistic self-love, along with the private landownership and monopoly to disappear when his reform policies are instituted. This part closes with *P. H. Wicksteed on Self-Interest: Resource Allocation and Social Inclusion*, by Yoshio Inoue. Drawing from Wicksteed’s research on Dante, it is argued that Wicksteed’s economic philosophy reminds us that the original and real significance of the market economy lies in the inclusion of all of the people. In my view, most of the chapters in this section are extremely interesting especially for historians of economic thought, given that they contain new material and approach the issue of self-interest from novel angles.

The starting chapter of the second part entitled *Otto Neurath’s Theory of Felicitology and the Will to Socialization*, is authored by Manabu Kuwata. It investigates Neurath’s felicitology as “the other welfare economics” and its relation to his ideas on socialization and planning (149). This chapter is basically a methodological piece given that it also presents a critical perspective on the conventional interpretation of the relationship between social sciences and logical empiricism. In *Selfish, Therefore Reciprocal: The Second Marginal Revolution of Mises*, Akihiko Murai focuses on the integration of selfishness with societal benefit in Mises’s thought. There is also an interesting comparison to Oscar Morgenstern’s views on this issue. The next chapter is entitled *Frank Knight on Self-Interest*. Masanobu Sato explains Knights’ analysis of the market economy in terms of the character and function of self-interest. Of particular interest is the comparison of Knights’ views with those of the liberal institutionalists and of the old and new Chicago economists (i.e., Viner, Coase, Friedman, Stigler). In *Karl Polanyi’s Motive of Economy and Institution*, Takato Kasai investigates Polanyi’s critique of approaches focused on self-interest exchange. He proceeds to analyze Polanyi’s agenda to inquire into a social institution that can restrict individual and national self-interests on a global scale. Next is *Contributions of Science of Free Will to Neuroeconomics and Quantum Decision Theory* by Taiki Takahashi. It is essentially a paper on political science in which current findings of the science of free will are discussed given the potential implications of these findings for future research in political science. *An Institution to Reconcile Self-Interests: Hayek on the Notion of the Economic Agent and the Establishment of Appropriate Rules* is authored by Masanori Taishido. The aim here is to clarify Hayek’s advocacy of his own institutional design. It is argued that Hayek’s theory of rules is designed to protect the diverse interests and
freedoms of individuals from the threat of collective and uniform interests. In my view, the sequence of the chapters should have been different in order to preserve continuity. The chapters on Mises, Knight, and Hayek should have been first to be followed by the rest of the chapters, which are more of a methodological nature.

The first chapter of the third part, entitled *Creating an Algorithm Based on the Theory of Moral Sentiments*, is written by Susumu Egashira. The paper reevaluates Adam Smith’s argument in *The Theory of Moral Sentiments* based on modern cognitive science, specifically within an agent-based simulation. The next chapter is on *The “Self” and the “Others”: From Game Theory to Behavioral and Neuroeconomics*. Christian Schmidt starts with the idea that the ‘self’ is to be understood by reference to the ‘others’, and connects it with game theory and recent contributions of neurosciences. The following chapter, *Why Is Behavioral Game Theory a Game for Economists? The Concept of Beliefs in Equilibrium* is by Michiru Nagatsu and Chiara Lisciandra. Their main argument is that the interdisciplinary exchange between economists and psychologists has so far been more active and fruitful in the modifications of expected utility theory than in those of game theory. The final chapter is on *Homo Economicus Under Multiple Pressures* by Uskali Mäki. After a brief but efficient review of the concept of Homo Economicus, Mäki provides an assessment and argues that economic theories assuming Homo Economicus find support from such trends in society, while common sense and other disciplines largely speak against it. I think that this paper is a very useful overview of some important current methodological issues pertaining to Homo Economicus.

My first observation concerns the three conceptual parts in which the papers are placed. I think that a more appropriate title for the first part should have been: *Self-Interest in the History of Economic Thought*, and should have included the papers on Mises, Knight, Hayek, and Polanyi (I would suggest that these chapters should be read in that order). The rest of the papers exhibit a more methodological focus and might have been better placed under a corresponding section discussing the epistemological aspects of selfish behavior. Even with this re-arrangement, there are a few chapters that are difficult to classify, such as the one on Neuroeconomics and Quantum Decision Theory, on Behavioral Game Theory, or on Game Theory and Behavioral and Neuroeconomics. These are quite interesting papers, but I suspect they would have fitted better in a different volume(s).
The second observation has to do with the views of some major authors on self-interest who are not included in this edited book. Keeping in mind that the term ‘genealogy’ clearly creates expectations of a systematic treatment of the origins, the development, and the role of this notion in economics, there are some important missing parts. For instance, I think that Thomas Hobbes’s approach to self-interest would have been very useful in the understanding of subsequent developments in economics. The same holds true for Menger and Knies for the Austrian school and the German Historical school respectively. Another omission is the work of Edgeworth whose ‘catallactis’ played an important role in the establishment of self-interest agents in late marginalism and early Neoclassical economics. In Edgeworth’s words: “The first principle of Economics is that every agent is actuated only by self-interest” (1881, 16). One could also add Pareto whose approach to self-interest exhibits some interesting common points to Wicksteed’s views which are covered in this volume. It has to be mentioned though, that this comment originates from a historian of economic thought who might have different expectations from the average reader.

Nevertheless, my overall opinion of this edited volume is positive. The majority of the papers in this book are very interesting and appealing for most historians of economics. There are also papers that will be of interest to economic methodologists and philosophers of economics. Some of the chapters can certainly initiate further research on the topics that they examine. Indicative but not exclusive examples are the chapters by Itai, Takahashi, Kuwata, Sato, Egashira, and Schmidt. In view of the current debate concerning the role of self-interest by economists and other social scientists, this book is a valuable and necessary addition.

REFERENCES
Stavros A. Drakopoulos is Professor of Economics in the Department of Philosophy and History of Science of the National and Kapodistrian University of Athens. He is author of *Values and Economic Theory: The Case of Hedonism* (1991), *Comparisons in Economic Thought: Interdependency Reconsidered* (2016), and *Economic Policy and the History of Economic Thought* (2023), editor (with Ioannis Katselidis). He has 54 scientific publications in refereed journals, and many chapters in edited books published by Routledge, Edward Elgar, and Springer. His research interests include History of Economic Thought and Economic Methodology, Labor and Health Economics, and the Economics of Subjective Well-Being. Contact e-mail: <sdrakop@phs.uoa.gr>

**FERNANDO VARELA LEVY**
*Erasmus University Rotterdam*

It is common to find causal statements in economics journals. In a recent article, for example, Babina et al. claim that “negative federal funding shocks reduce high-tech entrepreneurship and publications but increase patenting” (2023, 1). Unfortunately, the actual meaning of ‘cause’ is not always explicitly spelled out. What does ‘cause’ mean in these statements? Do they connote the same meaning? More concretely, what do economists mean when they say $X$ causes $Y$? These questions motivate Mariusz Maziarz’s intriguing book on causality. Based on a previous systematic literature analysis, Maziarz explores different research methods that economists use to advance causal claims. This allows him to show how different research methods implicitly endorse one of the main five philosophical accounts of causality: (1) the regularity approach, (2) the probabilistic approach, (3) the counterfactual approach, (4) the mechanistic approach, and (5) the manipulationist approach. In following his analysis, he claims: Economists are moderate causal pluralists. The methods they use to support causal claims reveal distinct relations that are consistent with different philosophical definitions of causality.

It is important to note that Maziarz’s study remains agnostic on what causality is for economists as an ontological stance. He focuses on an epistemological description and aims to disentangle what kind of evidence economists use to infer causal relations. Maziarz’s analysis does not stop here, though. The epistemological descriptive analysis allows him to advance some normative claims about policymaking as well. He focuses on how policymakers ought to use causal evidence, and contends

---


2 Maziarz labels his position as moderate in order to differentiate it from radical pluralism. The latter argues that each use of causal language carries an underlying different meaning of causality (see 201, footnote 1).
that different causal claims have different policy implications. In particular, he claims that policymakers should not use causal evidence that implicitly endorses a certain philosophical view to recommend policies that require a different account of causality. For example, a causal claim based on evidence assuming the regularity view of causation should not be used to justify a policy requiring a manipulationist approach. Otherwise, the intervention might result in unforeseen policy outcomes. Maziarz accompanies the discussion with a clear and succinct explanation of the type of policy the research methods can promote.

Maziarz describes in chapter 1 the framework he employs for studying economists’ causal claims. He analyzes these claims through the lens of referential semantics, which posits that the meaning of words and sentences is derived from their referents (i.e., the objects they refer to). Since the economy is only accessible via research methods, Maziarz considers these methods as the objects the causal claims refer to. In other words, Maziarz identifies the meaning of causal claims with the causal relations the research method can reveal. For example, if an economist conducts an experiment to see the effects of $X$ on $Y$, and claims that $X$ causes $Y$, the causal claim indicates the realization of the experiment. Maziarz breaks down the epistemic concept of causation that can be uncovered by the method and aligns it to one definition of causality in the philosophical literature.

From chapter 2 to chapter 6, Maziarz devotes each chapter to one of the five main philosophical views of causation. He follows a clear and well-organized structure. First, Maziarz introduces the respective philosophical approach and its reception in economics; he then follows a discussion on economic research methods that implicitly endorse the corresponding philosophical view. This is accompanied by examples from published papers in top economic journals. Lastly, Maziarz explores the policy implications that follow from the philosophical approach to causality. Chapter 7 serves as a conclusion for the entire book and recapitulates the main argument. Due to this structure, the chapters appear to be self-contained, allowing interested readers to easily delve into any chapter without necessarily having to read the ones preceding it.

In general, his descriptive claim about causal pluralism and his normative claim about how evidence should be used for policymaking are articulated as follows.

Maziarz maps economists’ research methods to philosophical approaches on causality in order to demonstrate their causal pluralism. The
*regularity account* suggests that empirical regularities, or constant event conjunctions, are indications of causal relations. Maziarz argues that this is implicit in theory-driven econometrics and data-driven cliometrics methods. The *probabilistic view* holds that causes increase the likelihood of effects. In this case, Maziarz contends that economists implicitly assume this view when using atheoretical econometric methods and when they differentiate cause from effect based on time precedence. This is the case, for example, when economists use vector autoregression models (VAR) to advance causal claims based on Granger causality tests.\(^3\)

The *mechanistic view* on causality is implicit in theoretical models without empirical support and DSGE models. In particular, Maziarz asserts that economists implicitly accept the mechanistic view in accordance with the New Mechanistic philosophy and Marchionni’s (2017) definition of economic mechanisms: the causal claim is produced by the interaction of entities. Finally, the *manipulationist view* of causality is implicit in experimental and quasi-experimental methods. By analyzing Instrumental Variables (IV), Regression Discontinuity Designs (RDD), randomized experiments and laboratory experiments, Maziarz argues that these methods are exemplary of Menzies and Price’s (1993) manipulationist approach in which human interventions raise the probability of certain effects materializing.

Maziarz argues that methods not based on the manipulationist view fail to guarantee a causal relation to be invariant under intervention. They do not necessitate that an intervention will bring about the effect. Data-driven methods assuming a regularity or probabilistic account can result from common-cause fallacies: The founded causal relation can be the result of a third variable causing both. With regards to theory-based methods which implicitly assume the regularity account, Maziarz argues that an intervention might change the causal structure previously assumed by the theory, even if the theory correctly depicted it. His argument is based on differentiating causal structures with and without interventions. Since they differ, a founded causal relation based on the latter cannot guarantee an effect on the former.

Regarding mechanistic evidence, he claims that this kind of evidence is unable to assure invariance because most models represent single mechanisms. In the economy, several mechanisms operate at the same time. The problem with the counterfactual account is that these two kinds

---

\(^3\) X Granger-causes \(Y_{t+1}\) if \(P(Y_{t+1} \mid \Omega, X) \neq P(Y_{t+1} \mid \Omega)\), where \(\Omega\) represent a vector of recent and lagged relevant variables.
of research methods can only establish token-level causal claims and thus their results cannot be extrapolated to guide type-level policy implications. Doing so would require additional knowledge.

Maziarz is not pessimistic, though. He claims that casual evidence not demonstrating invariance is still useful and can justify certain kinds of policies. On the one hand, studies presupposing probabilistic or regularity views on causality are sufficient for policies that do not modify the relata of causal claims (the cause). Policymakers should only use this kind of evidence to implement policy if the action does not interfere with the causal structure producing the phenomena of interest. On the other hand, studies assuming a mechanistic view can justify institutional reforms. This kind of policy aims at promoting certain kinds of mechanisms previously represented in economic models. Institutional reforms modify factors that shape preferences and decisions of economic agents, aiming to promote the emergence of mechanisms represented by models.

Even though Maziarz posits that methods presupposing the manipulationist view as the only way to assert invariance, he emphasizes the extrapolation problem. The causal claim is true only for the sample of the population under study, it does not guarantee the same effect on the target. Thus, Maziarz demands to make sure that a certain level of similarity exists between the sample population and the population target. This can be accomplished by means of pilot policies or pilot interventions on the target.

In brief, this book guides the reader through the philosophical literature of causality and explores how some economic methods presuppose different views of it. By doing this, the author also provides some normative assertions on how policymakers should use causal evidence. While I think that Maziarz’s arguments are valuable, I could not ignore some inaccuracies in how economists approach causal inference. I focus on a few of them.

Maziarz describes regression methods and correlation as being similar because they are symmetrical, and thus they cannot establish causal direction on their own (see 18, 71, 73 and 167).\footnote{Marzjarz appears to be using regression coefficients and partial correlation coefficients interchangeably. For example, he refers to Mian and Sufi (2012) regression coefficients as partial correlations (95). They are not necessarily equivalent.} While it is true that regression itself does not allow the economist to infer the causal direction
of two variables, a shared symmetrical characteristic with correlation coefficients does not correctly describe the problem of establishing causal direction.

Correlation and partial correlation coefficients are indeed symmetric. They are symmetric in the sense that it does not matter whether you calculated it between $y$ and $x$ or between $x$ and $y$. Either way, you will get the same result. Regression coefficients, on the other hand, are not a symmetrical relation in this sense. The regression coefficient of $y$ on $x$, does not equal the regression coefficient of $x$ on $y$. They do not even equal their respective algebraic inverse. Regression implicitly assumes causal direction from the right-hand to the left-hand side of the equation (Hoover 2008, 2012). It is true that there is a relation between regression coefficients and correlation. In a simple two variables case, for instance, they represent different normalisations of the correlations between the variables, but this does not mean they are symmetric.

The problem for inferring causal direction with regression can be illustrated by a simulation example inspired by Hoover (2001, 194). Figure 1 shows a figure in which data was generated as $x$ being a cause of $y$. Two regression lines are marked. One of $y$ on $x$ and another of $x$ on $y$. As previously argued, it can be seen how they are not symmetric in the same sense as correlation. What is unfortunate, however, is that nothing tells us which one of these two regression lines should be preferred. They are observationally equivalent. Given that in the real world, we do not know the real casual structure (i.e., the data generating process) as data can support two different casual structures. Regression alone cannot solve this problem. The problem arises because it is a statistical method. And as a statistical method, it cannot differentiate the conditional distributions of the variables. There is nothing in the joint statistical properties of $x$ and $y$ to tell us that the former causes the latter or vice versa. For this, we need causal assumptions: assumptions about the true data generating process.

![Figure 1: The problem of observational equivalence](image)
One way to deal with the needed causal assumptions is through the *potential outcomes framework*. In a nutshell, the potential outcomes framework, closely developed as part of randomized experiments, sees the problem of causal inference as a comparison between two potential outcomes: the outcome if individual $i$ is treated, $Y^T_i$, and the outcome if individual $i$ is not treated, $Y^C_i$. Unfortunately, Maziarz does not provide an accurate representation of this approach.

Maziarz contrasts economics with epidemiology, and asserts that counterfactual causality is part of the latter but not part of the former because epidemiology uses the potential outcomes framework (88). This assertion is not accurate. A significant part of economic research leans on the potential outcome framework. The Nobel Prize in Economics in 2021 recognized Guido Imbens and Joshua D. Angrist for their efforts in introducing this framework into economics (see Angrist 2022; Imbens 2022). Their work, in some sense, shaped how economists think about research designs and the type of causal relations they can uncover. A clear example of this is Doyle’s (2007) instrumental variable (IV) paper which Maziarz discusses in chapter 6. In this paper, the author explicitly references Angrist, Imbens, and Rubin (1996) and Imbens and Angrist (1994) when discussing his empirical strategy. Both papers demonstrate how the IV method can be interpreted within the potential outcome framework. It is unclear why Maziarz neglects the potential outcome framework as an important part of economic research.

Maziarz not only minimizes the impact of the potential outcomes framework in economics but also misrepresents it. He analyzes Mian and Sufi (2012) and claims that the study is based on the potential outcomes approach (88, 96). There is no use of formal potential outcomes notation in the paper and so it should not be taken as an example of this framework.

In summary, Maziarz’s book does an excellent job by emphasizing the causal pluralism that is part of economics research. The analysis is also valuable because he studies the implications on policymaking. Nevertheless, if philosophers take this book as a first approach to causal inference

---

5 The potential outcomes framework is also discussed in the literature as the Rubin Causal Model (e.g. in Weinberger 2002) or the Neyman-Rubin Framework (e.g. in Heckman and Pinto 2022.)

6 They shared the prize with David Card for his empirical contributions to labour economics.
in economics, they should carefully evaluate the explanations that are part of the book.

REFERENCES


**Fernando Varela Levy** is a research master student at the Erasmus Institute for Philosophy and Economics. His research interests include the methodology of economics, complexity economics and the philosophy of money. He holds a BA in Economics from the Center for Research and Teaching in Economics (CIDE), Mexico.
Contact e-mail: <fdo.varelaevy@gmail.com>

DAVID HOLLANDERS  
*University of Amsterdam*

This book characterizes the EU in more than one way. The EU allegedly exemplifies *inter alia* economic constitutionalism, benign despotism, welfare corporatism, executive federalism, technocratic managerialism, and authoritarian populism. This might suggest that the author prioritizes flowery description over analytical precision. Not so, as the central claim that the EU should first and foremost be characterized by the first two words of the title, ‘authoritarian liberalism’, remains clear throughout. As such, the argument has two components. First, the name of the game in the EU—and previously in the European Economic Community—is liberalism. Second, this liberalism is authoritarian.

Wilkinson’s aim is to “open the space to develop critical theories of European integration, not as some form of external constraint, or experiment gone awry” (viii), but as a project destined to become what the title of the book claims it is. He contrasts himself with academics who view the EU as a ‘sacred object’ that is not without flaws but also not inherently authoritarian.

The book develops the argument in four parts. In the first part, *Interwar*, the inherent tension in interwar liberalism—that is, how to reconcile political equality with economic inequality—is addressed with the work of Hermann Heller, who coined the term ‘authoritarian liberalism’. The second part, *Post-War Europe*, understands the early European project as a ‘new political constitutionalism’ that seeks to integrate through law. The third part—*The road from Maastricht to Lisbon*—interprets the two constitutions as the outcome as well as the accelerator of an increasingly authoritarian pressure for convergence. *Euro crisis*, part four, finally argues that the euro crisis concentrated power in the executive branch of government.

Not many readers will actually have to be convinced of the proposition that the EU is liberal. The European Commission would probably agree,
itself arguing that the single market “by removing technical, legal and bureaucratic barriers, [...] allows citizens to trade and do business freely” (EU, n.d.). This fits with Wilkinson’s understanding of liberalism as a praxis-cum-ideology that “depoliticizes the economy, naturalizes inequalities, and values markets, competition, and private ownership” (3). The author does not spill any beans over other possible interpretations of liberalism. This is not a hiatus, as the crucial element is present: depoliticization of the economy. This foreshadows the second part of the argument, that identifies depoliticization with de-democratization.

Empirical exhibit A of the authoritarian nature of the Union is the shock-therapy of Greece in 2015. Before election day, Wolfgang Schäuble already indicated that “New elections change nothing” (BBC 2014). After having won the elections in January and despite over 61% of the voters saying ‘Όχι’ (No) on July 5, the Syriza-government did indeed capitulate. It accepted all conditions and agreed to refrain from any unilateral action. In return Greece was ‘allowed’ to borrow from the European Stability Mechanism in order to pay off the Troika (ECB, IMF, EU-countries), who in turn lent to Greece in 2010 and 2012 in order to repay the banks who now lent to the ESM.

Wilkinson qualifies the ‘rescue’ of Greece as authoritarian, the latter being understood as the opposite of democratic. Parliament was indeed de facto suspended, when it chose the first of two options, either to set aside campaign promises or financial Armageddon, which would be unleashed by the ECB. Within 24 hours after announcement of the referendum, the ECB had already restricted liquidity to Greek banks on June 28, 2015 necessitating a cap of 60 euros on ATM-withdrawals. It made clear, if it wasn’t already, that the ECB is not so much independent from member-states as it is superior to them. Here and elsewhere Wilkinson tacitly assumes readers to be familiar with events. After brief, almost hurried descriptions he takes the time to fit the histoire événementielle in a broader understanding of the EU. Although the book is organized chronologically the author is then more a political scientist than a historian. In his reading, during the euro crisis the EU mutated from a rule-based order into bureaucratic Caesarism with power even more concentrated in the executive. Euro crisis management was ‘emergency neofunctionalism’ that replaced ‘integration through law’ by ‘integration through fear’.

These characterizations are on point, but one could argue that drastic times call for drastic measures. The book is largely devoted to refuting this position. The authoritarianism on display during 2010–2015 is seen
as neither incident nor accident. It is a built-in feature. The treaties of Paris (1951) and Rome (1957) were meant to finally resolve the dilemma that figured in all Western countries in the 1930s: how to reconcile democracy and capitalism in the presence of a demos that could always turn to expropriation, nationalization, and redistribution. The solution was simple enough and consisted of separating the economy from democracy. The *trente glorieuses* (1945–1975) did not constitute a lasting compromise between capital and labour, but a surrender in slow-motion of the latter. Capitalism was managed but not subjugated, while democracy was choked by being restricted to the non-economic domain. The liberal separation of economy and democracy was buttressed by counter-majoritarian institutions, many of which were European. The EU then was not only meant to keep fascism at bay but was also aimed to restore capitalism and reign in socialism. And that wasn’t without its urgency as in 1951 communist parties seized 37% of the vote in Italy and 29% in France.

The 1960s–1990s paved the road to authoritarianism by the concomitant Europe-wide replacement of mass mobilization by interest representation, and the deradicalization of socialist parties. Of course, left parties let themselves be pacified. Wilkinson briefly mentions the volte-face of Mitterand in 1983, characterizing it as the collapse of national Keynesianism for fear of currency devaluation. The historical description is overtly thin here; Wilkinson does not flesh out the historical trajectory of the capitulation. He also does not discuss whether this crucial episode epitomizes the 1980s ‘soft authoritarianism’ of the EU, interstate rivalry, economic civil war, or political opportunism of social(ist) democratic parties for whom European integration became an empty signifier promising much but delivering little.

Colonial wars waged by member-states are also not integrated into the analysis, and the Algerian war is mentioned once. Instead, the book fully focuses on the internal dynamics of the EU eco-system, in which “state sovereignty is restrained, but without being scaled up to the supranational level” (74). This leaves member states “with politics but no policies” and the EU “with policies but no politics” (206). The key is that not the demos but ‘the economic’ serves “as the principal ground of authority for the ‘constitution’ of the polity as a whole” (118). While acknowledging the permissive consensus in the pre-Maastricht era, the European project was, and is, a revolution from above, resulting in juristocracy and the rise of the executive. This all constitutes the soft authoritarian prehistory of
the post-2010 subjugation of parliaments and national governments in Ireland, Cyprus, Italy, Spain, Portugal, and Greece by the ECB-led troika.

There are discontinuities as well. First, in the 1950s the European market was “a means to connect consumers and producers beyond national borders within established productive structures”; yet, in the 1980s it became “a tool to modify the productive structures themselves, in order to improve their competitiveness” (130). Second, and most relevant here, the last 12 years saw a deconstitutionalization with the ad hoc formation of informal, extra-legal entities as the ESM, the Eurogroup, and the Troika. This followed years of (over)constitutionalization, but it served the very same aim: to de-politicize the economic sphere. As this authoritarian element already figures in the definition of liberalism, authoritarian liberalism is essentially a pleonasm in this book. The EU being authoritarian is the final verdict, and it is a convincing one.

REFERENCES


David Hollanders is lecturer in economics at the department of European Studies, University of Amsterdam. Contact e-mail: <d.a.hollanders@uva.nl>

ANDRES LAZZARINI

*Goldsmiths, University of London*

Two crucial methodological questions underlie the construction of formal economic models: (i) is it legitimate to build economic models on the assumption of a purely individualistic ‘economic agent’, often referred to as *homo oeconomicus*? And (ii) would adopting alternative conceptions of individuals other than *homo oeconomicus* allow for the creation of more realistic models, perhaps closer to the original aims of the economics discipline? Anna Horodecka’s *Human Nature in Modern Economics: Structure, Change and Perspectives* (2022) is a very well-researched and comprehensive attempt at addressing these two problematic yet central issues in economic methodology. It does so by investigating the salient concept of human nature in economics, both from a purely methodological perspective and from the history of economic thought. One of the main theses of this book is that how pure theoretical economics (i.e., the search for those scientific relations and mechanisms that are at play in economies) is practised fundamentally depends on which concept of human nature is adopted in different contexts. For Horodecka, human nature is a “key construct” (2) that is needed in order to achieve a thorough comprehension of economic reality, and the changing conception of human nature across the discipline reflects how the economy must make adjustments to adapt itself to a changing environment. Furthermore, differing views of human nature could be the fundamental dividing line between orthodoxy and heterodoxy because opposing conceptions of human nature have determined the evolution and transformation of economics. In this way, the author argues, different conceptions of human nature have established not only different assumptions of economic agency in formal models, but they are also responsible for the way economic policies derived from those models have been implemented. Not surprisingly (for reasons which will become apparent below), most of the changes in the conception of human nature took place within the myriad of heterodox economic approaches since the 1980s.
The book is composed of an introduction, six main chapters, and the conclusion. Chapter 1 introduces different conceptions of human nature and how to study them. It establishes the terminology and structure of the meta-model (or meta-concept) of human nature that will serve as the basis of comparison across different strands of economics throughout the book. Chapter 2 shows how the concept of human nature determines how economics is done, and persuasively argues how the relationship between human nature and the foundations of economics caused the division between orthodox and heterodox economics. Chapter 3 discusses the evolution of human nature as seen in orthodox economics while, at the same time, showing how *homo oeconomicus* became the starting point of conceptualising alternative concepts of human nature within heterodox economics. Chapter 4 presents a thorough description and analysis of different alternative concepts of human nature for a wide range of heterodox approaches. Chapter 5 emphasises the importance of the environment of the human being and how the economist must take this into account in their characterisation of the economic agent. Finally, chapter 6 discusses the perspectives which are opened to the economics discipline in light of the changing and transforming concepts of human nature.

The topics covered in this book are highly abstract, and they will certainly require the keen attention of the reader. Anna Horodecka, however, does a great job at examining these highly conceptual notions underpinning her main theses. Chapter 1 lays out the specific terminology related to different concepts of human nature. Drawing heavily on philosophical anthropology, the general model developed by the author (which she refers to as a ‘meta-model’ or ‘meta-concept’ of human nature) adopts three dimensions that capture the nature of the human economic agent: (i) the ontological dimension, whose job is to explain relations between the human being, nature, and the metaphysical world; (ii) the social dimension, which examines relations among people; and (iii) the individual dimension, which portrays the human being as an individual. Associated with these dimensions are three “vital levels of human nature” (3): action, motivation, and reason, which the author links to the anthropological levels of body, soul, and mind, respectively. The general framework in the meta-model of human nature (31–39) allows for comparisons across both orthodox and heterodox schools of economic thought.

From the standpoint of the methodology of science, anyone trained in orthodox economics may argue that economics performs three well-
known functions: to better grasp the working of market economies (descriptive function), draw accurate diagnoses (explanatory function), and propose economic policies (predictive function). Yet, as the author convincingly argues in chapter 2, these functions are shaped by the conception of human nature adopted by economics, which plays a key role in the context of discovery and justification. The context of discovery refers to the realm of seeking knowledge. Within this context the view of human nature depends on knowledge, cultural embeddedness, ideas, and philosophical approaches that shape theorizing about the economic research at hand. The context of justification, in turn, refers to how views of human nature are represented by assumptions adopted by economic theory which are formed through abstraction and idealisation. At this point of the discussion, the author persuasively leads the reader to ask themselves about the problem of precedence concerning human nature in economics:

If the objective and the method of economics should be set first, followed by a search for a suitable concept of human nature, one might ask about how humans established that specific rather than any other, objective of economics and those, and not any other, methods for studying economic agents, phenomena and processes. [...] To some extent, the concept of human nature precedes that process by definition. It is a cognitive construct, like spectacles allowing a clearer view of the reality under examination. Due to their limited perception capabilities, humans see reality through cognitive heuristics (assumptions about human nature and the world). (69; emphasis added)

Indeed, the relationship between human nature and the foundations of economics is central to the division between orthodox and heterodox economic approaches. The criticisms raised by the latter against orthodox approaches are chiefly focused on the concept of human nature present in its theories. Furthermore, the specific assumptions on human nature are not only reflected in the ends pursued by the economic agent, but they also determine the way of formulating the objective, scope, and methods in economics. Precisely because the heterodox approaches originate in the criticism raised against the orthodox view of human nature present in the assumption, it is of paramount importance to understand the evolution of human nature as seen by orthodox economics.

The conception of human nature in orthodox economics is fully examined and discussed in chapter 3. Although this conception is presented
in the book in its most oversimplified version (see, e.g., 86–87, 90–91), the author ingeniously shows how the only purpose of introducing human nature into the orthodox framework is to elucidate the systematic relationship between the perfectly rational individual economic agent and their aim to maximise their utility. In other words, the individual (or this concept of human nature) is equated with utility maximising behaviours, thus becoming the key cognitive heuristics used by orthodox economics. Therefore, any change in the context whereby such an individual was ideally constructed would be avoided or pre-empted in advance, because this disturbance would entail a change in the cognitive heuristics (i.e., in the conception of the human nature itself). It thus runs the risk to change, in turn, the objective, scope, and methods of economics and, therefore, the whole of the economics discipline. Therefore, if the general conditions whereby human beings exchange, produce, consume and distribute change, another conception of human nature will be necessary to reflect such changes.

Indeed, as the author makes the case, the criticism of the individualistic nature of the human conception in orthodox theory has become the starting point for heterodox economics to emerge in the history of economic thought. Thus, by employing the meta-model components mentioned above, heterodox economics starts by criticising the treatment of the mind as a calculating tool to maximise utility—which is solely responsible for explaining human behaviour. Drawing on results from evolutionary economics, behavioural economics, humanistic economics, and neuroeconomics, among other strands of heterodoxy, the author argues that orthodox economics has in fact neglected other sources of human action. Motivations such as moral development, adaptation to social and environmental requirements, reduction of imbalances to boost social well-being are also drivers, among others, of human action. They therefore call for a redefinition of the cognitive heuristics, and thus of the conception of human nature used in economic theory. For Horodecka, heterodox economics aims to create a framework where individual activities combine both individual and systemic goals (see, for example, 96–97). Yet, heterodox economics would need to develop alternative conceptions of human nature for that purpose.

In chapter 4, the author presents a thorough description and analysis of different alternative concepts of human nature for a wide range of heterodox approaches, covering ecological economics, behavioural economics, feminist economics, evolutionary economics, humanistic economics,
Buddhist economics, and religious economics. Horodecka asserts that heterodox economics’ criticism of *homo oeconomicus* actually implies that economics as a science (positive aspect) cannot be separated from economics as an art (normative aspect). Both aspects are necessarily and simultaneously wrapped up in heterodox economics because the aims and general reflections of heterodoxy not only encompass the logic of the economic system (maximum growth, stability, etc.) but also social well-being, moral values, social adaptations, and requirements. Given this broad objective, the concept of human nature adopted entails taking up an *interdisciplinary perspective*. Thus, heterodoxy borrows different elements from several disciplines, such as psychology, linguistics, philosophy, sociology, and the natural sciences to build its views and conceptions of human nature. The apparent flexibility in the construction of the view of human nature in heterodox economics is chiefly rooted in the complex nature of economic phenomena, which cannot be isolated from the historical context and the institutions and norms in which the economy and the individuals running it are embedded. It is the complexity of processes involved in the economic activities and relationships that becomes the main determinant of changes in the concept of human nature in economics. Indeed, the concept of human nature discussed for each of these heterodox approaches is broken down into three core aspects: behaviour (the body), motivation (the soul), and reason and meaning (the mind). One of the main upshots of the chapter is that describing a human being as an ‘economic agent’ indicates that economic activity is indeed one of the various activities carried out by humans; but also that these activities are pursued by a complex human nature. Acknowledging different degrees of complexity will define, ultimately, how far the concept of human nature adopted by heterodox economics departs from the *homo oeconomicus* ideal.

The reader is led, in chapter 5, to acknowledge the importance of the changing environment of the human being and how the economist has to take this into account in their characterisation of the economic agent. Once again, by making use of the suitable meta-model introduced in earlier chapters, the author highlights that human action not only involves the individual dimension of their nature but also the ontological and social ones. Therefore, in examining the relationships between an economic agent with others, both the nature and the system of ontological values (e.g., the underlying visions or principles guiding economic agents) have to be taken on board. However, due to an increasingly complex nature of
the economy, the environment is subject to changes which, once acknowledged by the economist, will entail persistent changes in the concept of human nature adopted. Therefore, changing environments will likely lead to changing conceptions of human nature, which will make it all the more difficult for the orthodox paradigm to maintain its rather unrealistic conception of human nature, which sees the individual economic agent aiming at maximising utility under perfect rationality. To come to grips with such tremendous challenges posed by a complex reality, the author proposes that the economist can draw inspiration and eventually adopt new concepts of human nature based on more colloquial views.

The economy (and the embedded actors, societies, political bodies, markets, institutions, etc.) analysed by both orthodox and heterodox economics is claimed to be a ‘real’ entity. Moreover, human beings share the same structural biology in the same market economy. Yet, in order to deal with such a complex concept as that of human nature, both orthodox and heterodox approaches have highlighted its different, radically diverging aspects for the purpose of building economic models. But can mainstream economics and heterodox approaches communicate with one another? Is it possible to find some commonality between such diverging conceptions of human nature? Such key questions linger in the last chapter of the book (chapter 6), where the author discusses the perspectives which are opened to the discipline of economics in light of the changing and transforming concepts of human nature. The chapter makes the point that different assumptions about human nature will likely bring about both cognitive and practical consequences in economics, and therefore poses the big question of whether orthodox economics will undergo modifications or a paradigm shift; or if economics as a whole will fall into either separatism or a pluralism embracing different approaches. Surely, the sketched range of possibilities may sound quite extreme, and, in fact, it will likely be the case we find the evolution of economics moving somewhere in between. Interestingly, this is the point made by Horodecka in this chapter, where she discusses degree of penetration of assumptions rooted in heterodox traditions into the orthodox paradigm. The author calls this the “absorption” strategy pursued by orthodox economics (229). An example of this can be seen in the late 1930s and 1940s, with Keynesian macroeconomics being incorporated into the orthodox framework (the ‘neoclassical synthesis’). Although the neoclassical synthesis is not examined in this book, the historical analogy may be useful to the reader
to understand the complex process of ‘absorption’ of heterodox elements and assumptions by the orthodox paradigm.

Overall, this is a very-well researched, deeply explored, and widely informed scholarly work, which attempts to understand abstract, general conceptual notions underpinning the different assumptions about human beings (understood as economic agents) in varied economic models. Built on an incredibly huge literature spanning different disciplines, this book painstakingly makes the point that how thinking and theorizing about economics is done is determined by the economist’s understanding of human nature. This is indeed a key contribution to economic methodology, not to be found, to the understanding of the present reviewer, elsewhere in the relevant literature (e.g., Becker 1976; Davis 2006; Mäki 2012).

On the other hand, the book has missed out on two important historical episodes that have shaken the development and consolidation of orthodox economics during the 20th century: the Keynesian revolution of the 1930s and the Cambridge capital theory controversies of the 1960s. Both theoretical conflicts, as acknowledged in the history of economic thought, put the orthodox paradigm under serious risks (see, e.g., Harcourt [1972] 2022; Mirowski [1989] 2011; Pasinetti 2009), and yet it survived both crises and continues to inform most economists in academia, industry, and government. There is no doubt that this would have been a nice addition to the historical sketches about the different heterodox schools discussed in the book, but this omission in no way detracts from its important contribution to the understanding of the epistemological and ontological underpinnings behind assumptions about human nature in economic models. Indeed, whether orthodox or heterodox, economists rarely pose the question of what the ontological, social, and individual dimensions of human nature present in their assumptions about the economic agent are. Yet, assumptions about human nature play a key role in creating the cognitive heuristics that enable different schools of economic thought to be able to study highly complex economic processes. At any rate, as this contribution has shown, it is clear that a single concept of human nature, such as the *homo oeconomicus* guided exclusively by maximising utility, would be ineffective to create new theories that can fit and grasp ever-new problems and issues arising in the modern economy.

REFERENCES


Andres Lazzarini is Lecturer in Economics at Goldsmiths, University of London. His research fields are history of economic thought, methodology and epistemology of economics, and economic development in historical perspective. He has published in Review of Political Economy, Research in the History of Economic Thought and Methodology, Review of Radical Political Economy, Filosofia de la Economia, among other scholarly outlets.

Contact e-mail: <A.Lazzarini@gold.ac.uk>
PHD THESIS SUMMARY:
Gender in the Labor Market

ANNE SOPHIE LASSEN
PhD in Economics, November 2022
Copenhagen Business School

In the novel Orlando: A Biography, Virginia Woolf (1928) tells the story of
the nobleman Orlando, who undergoes a mysterious change of sex and
lives the rest of his life as a woman. Woolf depicts Orlando’s despair after
discovering the limitations and expectations that come with being a
woman:

The comforts of ignorance seemed utterly denied her. She was a
feather blown on the gale. Thus it is no great wonder, as she pitted
one sex against the other, and found each alternately full of the most
deplorable infirmities, and was not sure to which she belonged—it
was no great wonder that she was about to cry out. (77–78)

As economists, we might think of these deplorable infirmities in other
ways than Orlando does, but we are well aware of gender differences in
behavior and inequalities in outcomes. While men and women's roles in
society have converged over the last century, substantial gender gaps re-
main. This dissertation approaches the question of gender equality from
different angles—parenthood, job loss, and occupational outcomes—to
shed new light on how gender norms contribute to persistent gaps.

The notion of gender norms is making its way into economics
(Lundberg 2022), and in this endeavor it is useful to pay attention to in-
sights from the other social sciences. Sociologists Pearse and Connell
(2016) show that economists often—either explicitly or implicitly—ass-
sume a strong consensus in society and fail to acknowledge the within-
society variation regarding gender-related attitudes. Drawing on insights
from the broader social sciences, Pearse and Connell (2016) also argue
that while gender norms can be stated abstractly (e.g., in value surveys),
what matters is the way norms function in social life. Treating norms as
something that inhabits the mind of individuals (and not too different
from the concept of preference), overlooks how norms function as the
pillars of a community or society, and how they are embedded in institutions, whether in the form of organizations, public policies, or collective identities.

Parental leave policies provide a useful example of this. In high- and middle-income countries, parenthood is the main driver of gender inequality in the labor market. To help mothers balance work and family considerations, most countries have implemented some sort of maternity leave system. Chapter 1 (published as Canaan, Lassen, Steingrimsdottir and Rosenbaum 2022) reviews a large literature on the impact of maternity, paternity, and parental leave on women's labor market outcome, maternal and child health, and firms. While parental leave can be used by any parent, mothers are the primary users of these policies and mothers subsequently allocate more time to childcare than fathers.

Chapter 2 provides an improved understanding of the striking gender gaps in time allocation following parenthood. A parental leave reform in Denmark provides a useful setting for understanding how parents distribute leave. The reform economic compensation new parents received while on leave effectively increased the leave duration. At the same time, policymakers removed two weeks of earmarked leave specifically allocated to fathers. First, I show that mothers increase their leave by 5 weeks while the average leave duration of fathers remains unchanged, irrespective of relative earnings. Second, I ask how gender norms at the family level influence women's take-up. The pattern is highly consistent with a model of gender identity: women who had a working mother take a shorter leave than those with a stay-at-home mother. Moreover, I document peer effects among sisters who take a longer leave if exposed to the reform-induced change in leave duration. The chapter utilizes within-society variation in gender norms to understand the mechanisms being the striking gender gaps in time allocation around parenthood and highlights how policies and norms are intertwined.

When the responsibility of childcare falls disproportionately on women, it likely imposes constraints on how women can adjust to labor market shocks. Chapter 3 investigates the effects on women's and men's job loss. Following job loss, women are facing a higher risk of unemployment than their male counterparts. Women are over-represented among workers with little formal education. In addition, both the absolute and the unexplained parts of the gender gap grow in the presence of children. Women’s gain in the labor market is fragile and mothers are facing meaningful barriers.
Chapter 4 focuses on intergenerational transmission of the gender composition of occupations. While female-dominated sectors are growing in size and importance (Petrongolo and Ronchi 2020), men appear reluctant to enter these occupations. To improve the understanding of the gender norms that are influencing men, I use detailed administrative data to construct measures at the family level, school level, and broader municipal environment to capture transmitted norms of the ‘appropriate role’ of men and women. Boys socialized in environments with men who have gender-stereotypical labor market behavior enter male-dominated occupations themselves. In general, women’s labor market behavior has negligible effects on boys. However, when boys grow up around a larger share of mothers who work full-time gender segregation decreases in the next generation. The chapter highlights both the immediate family and the schooling environment as important sites for the construction of gender norms.

Each of the empirical chapters of the thesis are answering independent questions related to gender equality. With detailed and linkable administrative data, I construct measures of gender norms and exploit variation in important spheres. Next to the empirical contributions of this thesis, there is an underlying lesson: gender norms are not monolithic. Paying attention to within-society variation and the sites that enforce gender norms provides meaningful insights into the mechanisms behind persistent gender gaps.

REFERENCES

Anne Sophie Lassen is a postdoc at Copenhagen Business School at the Department of Economics. Her research is primarily concerned with the mechanisms behind and implications of gender differences in labor market outcomes. She received a PhD in Economics from Copenhagen Business School in 2022.
Contact e-mail: <assl.eco@cbs.dk>
PHD THESIS SUMMARY:
Elucidating the Role of Value Judgments in Normative Economics

NESTOR LOVERA NIETO
Doctorate in Economics and PhD in Management, November 2022
Université de Reims Champagne-Ardenne and Neoma Business School

My thesis claims that normative economics should be recognized as a science and that this status should be maintained and strengthened. For this reason, the aim of this thesis is to answer the following question: What is the role of value judgments in the study of normative economics?

I proceed to answer this question in three chapters. In the first chapter, I address the research question from a general perspective by drawing a distinction between the value judgments of economists and those of economic agents. This allows me to analyze, in the second chapter, how economists study and characterize the value of freedom in relation to the value judgments of economic agents (as represented by their preferences). I focus on analyzing the freedom of choice that an individual has over an opportunity set at a given moment. My purpose is to determine how the opportunity or preference-based approaches use the tools of social choice theory to develop the conception of freedom in economics. The study of the value of freedom leads me, in the third chapter, to consider the importance of value judgments in the elaboration and interpretation of normative economic theories, specifically in Arrow's impossibility theorem, given its importance in the history of normative economics. Through the study of these three chapters, I seek to show that economists cannot dispense with value judgments in the study of normative economics.

In the first chapter, I argue that the distinction between the value judgments of economists and those of economic agents has not been explicitly established in welfare economics, which has typically treated economics as a neutral social science. The importance of making this distinction lies in determining whether economists may make value judgments in their professional work and how value judgments may be crucial to justify the economic agents' preferences.
Two approaches can be useful in order to make the distinction: Sen’s (1967) classification of value judgments and Mongin’s (2006) theses about value neutrality in economics. The criteria proposed by Sen offer elements of an answer to the way economic agents justify their preferences. Thus, economic agents justify their preferences (i) if there is a set of evaluative judgments that can play a determining role in their choice; (ii) if there are categorical or hypothetical imperatives to which the agents agree; and (iii) if there is a change in factual circumstances that leads agents to make a value judgment. According to Mongin’s analysis, the weak non-neutrality thesis is the most appropriate to study the role of values of economists in their professional work. This means keeping two things in mind: (i) the importance of the nature of predicates in the statements made by economists, which will be crucial to determine if a statement is logically evaluative or logically factual and (ii) that it is difficult to make a clear distinction between normative and positive economics.

Then, I use the subject of interpersonal comparisons of utility to discuss this distinction, more specifically the analysis of Harsanyi’s (1955, 1977) impartial observer theorem. This starts from the assumption that economists take on this role because they can make judgments of values and judgments of facts about individuals’ well-being with a certain degree of neutrality, while I contend that the value judgments of economic agents justify their preferences when making choices. The analysis of this case study allows me to show that it is not possible to ‘objectify’ subjective attitudes of economic agents, and that economists cannot avoid value judgments when they recommend an action for the benefit of the society.

In the second chapter, I consider three ways in which preferences are conceptualized in the measurement of freedom of choice, i.e., over actual preferences, over future preferences, and over reasonable preferences. In examining the origins of these conceptions I identified three tools of social choice theory that influence the study of preference-based approaches. First, the association between freedom and the individual’s preferences is related to the importance of preferences in social choice theory, where preferences are at the basis of social choices. Second, the important place given to consequences in social choice theory leads me to consider two aspects when assessing freedom: the process aspect of freedom concerned with procedure of autonomous choice and the opportunity aspect of freedom concerned with the real opportunities that individuals have in order to achieve things that they value. Third, the notions of incompleteness and maximality that come from social choice theory.
play an important role in the preference-rankings for the evaluation of freedom. The evaluation of freedom does not lie in being able to rank each set of options vis-à-vis the others, but rather in making use of as much reasoning as possible to rank different alternatives in a very definitive way.

In the third chapter, I inquire how Arrow's own values and cultural context influenced his methodological choice. I argue that the academic context surrounding Arrow, his own interest in mathematical logic and in election systems, and the geopolitical context characterized by the beginning of the Cold War were all essential to his formulation of social choice theory.

However, value judgments did not only influence Arrow's methodological choices, but they have also had an impact on how his impossibility result has been interpreted. I analyze Arrow's impossibility theorem from the perspective of Mongin (1999, 2002, 2006). I take as my starting point the interpretations proposed by Mongin (1999) to study how Arrow considers value judgments in his impossibility theorem. I consider Mongin's (2002) analysis of the impact of Arrow's development of social choice theory in the progress of normative economics. I also examine Mongin's (2006) hypothesis that Arrow can be considered a weak-neutrality theorist, meaning that Arrow considers it logically possible for an economist qua economist to make value judgments, and that there are occasions in which they might, and even should, make value judgments. Mongin's value neutrality classification is useful to determine that Arrow cannot dispense with making value judgments when stating his impossibility theorem and its possible solutions.

I show throughout the three chapters of the thesis that economists cannot dispense with value judgments in the study of normative economics. Value judgments are thus essential not only to choose the methods and criteria that are useful for evaluating the desirable or undesirable character of economic situations, but also when formulating prescriptions for the implementation of these methods and criteria.

REFERENCES


**Nestor Lovera Nieto** will join the Center for the History of Political Economy at Duke University as a Visiting Scholar to begin September 2023. His research combines the philosophy of economics and the history of economic thought to show that economists cannot dispense with value judgments in studying normative economics. He received a double doctoral degree, a Doctorate in Economics from Université de Reims Champagne-Ardenne, and a Ph.D. in Management from Neoma Business School in 2022.

Contact e-mail: <lovera.nestor@gmail.com>
A Model-Based and Mechanistic Approach to Social Coordination

MATTI SARKIA
PhD in Philosophy, September 2022
University of Helsinki

Human beings live in group sizes that outscale any other animal society, with the exception of genetically closely related superorganisms, such as ants and beehives. The average proximate group size (involving regular contact and name recognition) of most humans is on the scale of 150 individuals, thousands of individuals cross each other on a daily basis when commuting, and millions of individuals may occupy the same urban territory. Such large-scale social life demands coordination: in order to avoid hostile confrontations and accidental collisions, individuals need to be able to coordinate their attitudes and behaviors in a manner that is mutually beneficial, or minimally, that avoids unnecessary harm. How do we achieve this remarkable feat?

The human capacity for social coordination, as well as the cognitive and social mechanisms underlying its manifestations in various environments, have become prominent topics of research in many different scientific disciplines during recent years. These disciplines include analytical philosophy, cognitive science, developmental psychology, experimental economics, game theory, and evolutionary anthropology. They engage with human social coordination on multiple spatial and temporal scales, from the phylogenetic time scale of gene-culture coevolutionary processes (consider the evolution of symbolic signaling systems) through the ontogenetic time scale of individual development (for example, learning natural language during infancy) to the synchronic time scale of two physically co-present individuals coordinating their motor behaviors with one another (such as the tendency to adjust one's tone of voice to that of one's conversation partner). However, genuinely interdisciplinary investigations of social coordination have remained rare, and there have been few attempts to integrate different disciplinary perspectives into a more comprehensive, multidisciplinary theory of social coordination. Of
course, institutional incentives may play some role in hindering interdisciplinary collaboration. However, my dissertation is premised on the idea that one important obstacle to more extensive interdisciplinary integration is the absence of shared methodological frameworks for formulating multilevel explanations of complex phenomena that occur at multiple temporal and spatial scales, involve a mix of sub-personal and personal-level cognitive processes, and are influenced in myriad ways by the cultural and social environments that they occur in.

Two such methodological frameworks from contemporary philosophy of science play an especially central role in my dissertation. First, the mechanistic approach to explanation, which was originally developed in the biological and life sciences, enables me to frame questions about how investigations of social coordination in the behavioral, cognitive, and social sciences relate to one another (Sarkia et al. 2020; Sarkia 2016; 2022b). The central idea behind mechanistic explanation is to explain why some phenomenon occurs by describing how it occurs, or in the words of one classic paper, by describing "a structure performing a function by virtue of its component parts and component operations and their organization" (Bechtel and Abrahamsen 2005, 423). While mechanistic explanations might be viewed as ontologically reductive, I argue in my dissertation that they are nevertheless epistemically anti-reductive, because the parts and operations that make a difference are individuated relative to the phenomenon to be explained. This is so especially in the case of complex behavioral phenomena, such as human aggression, altruism, coordination, or cooperation. In these cases, the phenomenal decomposition of the explanandum into different manifestations (consider the distinction between pure coordination games and mixed-motive games in game theory) serves as an important prerequisite for its mechanistic decomposition into entities and activities that may produce these different manifestations (such as psychological mechanisms for altruism or empathetic preferences as explanations of cooperation in public goods and dictator games). Although not all phenomena are mechanistically decomposable, many are, and much cutting-edge research at the intersection of the behavioral, cognitive, and social sciences is premised on the assumption that complex human behaviors do not resist mechanistic explanation in principle, even if mechanistic explanation is a more interdisciplinary and multifaceted endeavor than is often imagined.

The second methodological framework that plays a central role in my dissertation is theoretical modeling, understood as a distinctive approach
to scientific investigation, which is based on the construction and elaboration of hypothetical surrogate systems that are used to stand in for some system in the real world. The central idea behind model-based science is that instead of studying the world directly, by means of observation and experimentation, it is sometimes a more feasible strategy to study the world indirectly, by first constructing a theoretical model, and then using that model to draw inferences about its target. Theoretical models abound in sciences that deal with complex phenomena, which resist experimental isolation and control: consider climate models in environmental science, models of rational action in economics and artificial intelligence, models of predator-prey interactions in ecology and evolutionary biology, and model organisms in medicine. However, even when they engage with the same subject matter, there can be significant differences between the types of models that are developed in different academic disciplines, which in turn makes it difficult to draw comparisons and adjudicate among them. For example, game theorists typically model social coordination as equilibrium behavior that is based on best response-reasoning, where no individual has an incentive to change her behavior provided that the others do not change their behaviors. On the other hand, philosophers have often discussed social coordination under the rubric of collective intentionality (referring to the sharing of goals or other mental states with other individuals), while cognitive scientists have also discussed non-intentional forms of coordination (such as two individuals unconsciously aligning their bodily postures, mannerisms, and facial expressions). Given their contrasting disciplinary conventions and perspectives, the integration of theoretical models from different scientific fields gives rise to numerous conceptual and methodological challenges. My article-based dissertation aims to answer such challenges in a manner that is piecemeal and problem-oriented, in accordance with a naturalistic approach to the philosophy of social science.

The first three articles of my dissertation deal with questions of general philosophical interest, philosophical methodology, and social ontology. In “Modeling Intentional Agency: A Neo-Gricean Framework” (Sarkia 2021), I distinguish three alternative strategies for modeling intentional agency in analytic philosophy of mind and compare them to similar strategies of scientific modeling. The bottom-up strategy of Gricean modeling is compared to the use of computer simulations and agent-based models, while the horizontal strategy of analogical modeling is elucidated by reference to research on analogical inference in cognitive science, and the
A top-down strategy of theoretical modeling is compared to the uses of abstract and domain-general mathematical frameworks in fields like economics and ecology. In “A Family of Models of Shared Intentionality” (under review), I describe different philosophical analyses of shared intentionality (by Bratman, Gilbert, Searle, and others) as alternative elaborations of a common folk psychological framework, which share a set of core assumptions about how different types of mental states (such as goals and intentions) are related to one another, but which make contrasting elaborations and extensions to this framework thus leading to contrasting accounts of shared intentionality. And in “A Model-Based Approach to Social Ontology” (Sarkia 2022a), I provide an in-depth analysis of Raimo Tuomela’s (2013) philosophical account of the we-perspective as a highly idealized theoretical model, which deliberately abstracts away from certain aspects of social reality in order to bring others into sharper relief.

The latter three articles of my dissertation engage with questions of interdisciplinary integration and division of labor. In “Minimalism and Maximalism in the Study of Shared Intentional Action” (Sarkia 2016), I argue that cognitively undemanding accounts of social coordination have been defended by means of two contrasting strategies of argumentation, which I describe as ‘complementarist’ and ‘constitutionalist’. In “Mechanistic Explanation, Interdisciplinary Integration, and Interpersonal Social Coordination” (under review), I apply the distinction between phenomenal decomposition and mechanistic decomposition to carve out a schematic division of labor between behavioral and cognitive scientists studying social coordination. And in “Mechanistic Explanations in the Cognitive Social Sciences: Lessons from Three Case Studies” (Sarkia et al. 2020), I (together with my co-authors) present the mechanistic approach to explanation as a general template for integrating models and theories across the behavioral and social sciences, using cognitively inspired research on ethnicity, social coordination, and transactive memory systems to make our case.

Taken together, the six original articles in my dissertation demonstrate how contemporary approaches related to mechanistic explanation and theoretical modeling in the philosophy of science can play an important role in adjudicating interdisciplinary debates related to social coordination, and indicate how philosophical studies of social coordination interlock with empirical and theoretical studies of social coordination in other scientific disciplines.
REFERENCES


Sarkia, Matti. 2022b. *A Model-Based and Mechanistic Approach to Social Coordination*. Philosophical Studies from the University of Helsinki.


**Matti Sarkia** obtained his PhD from the University of Helsinki in September 2022. He is currently working as a post-doctoral researcher in the Kone Foundation-funded research project: Cognition in Social Interaction: Towards a Mechanistic Integration of the Cognitive and Social Sciences. Contact e-mail: <matti.sarkia@helsinki.fi>
Between political instabilities, the looming effects of climate change, and the insurgence of a viral pandemic, we face great challenges as communities, societies, and species. During uncertain times, the importance of experts in our collective decision-making increases as we want to learn as much as possible about our difficult circumstances in order to implement effective measures to face whatever is going on. Yet, our need for trustworthy information may hide the risk of over-relying on experts. The challenges we face are complex issues, the resolution of which requires both empirical knowledge and value judgements. If our uncertainty covers both but we rely on scientists to find resolutions, we are leaving it to the scientists to make value judgements too. Then, over-relying on experts can lead to shielding policy makers from their responsibilities, and to nourishing anti-science positions perceiving the political content of scientists’ opinions.

This dissertation is an attempt to understand what sort of uncertainty decision-makers face, and to explore the implications of this analysis for private and public decisions: my main thesis is that there is a plurality of types of uncertainty, not all of which can be reduced through acquiring more evidence.

The first chapter is devoted to a theoretical account of uncertainty and of its variability. In mainstream decision theory (specifically, the tradition following loosely von Neumann and Morgenstern 1944 and Savage 1954), uncertainty is expected to vary only in terms of severity, and the standard probabilistic treatment has been expanded accordingly to tackle increasingly severe types of uncertainty. But this approach assumes that uncertainty only concerns how things are, whereas there is a variety of doubts that agents can face in their decisions and that go beyond that. Building on Makins (2021), I propose that we understand uncertainty as a matter of conflicting reasons to hold alternative attitudes, which means that it arises when there is some disagreement and that its resolution
requires the resolution of this disagreement. But given that in some cases this disagreement can persist even under ideal conditions, then in those cases the improvement of these conditions—for instance, with the collection of new evidence—cannot be expected to lead to the resolution of the uncertainty. On this account I construct a typology of uncertainty, identifying the conditions under which it is grounded in radical disagreement and retracing the role that these radical uncertainties can play in decisions.

Among those at play in decisions, one which has received relatively little attention is uncertainty over how the decision is modelled. Typically, decision theories require agents to model their situation identifying a set of alternative options, a set of possible states, and a set of consequences. But what should be included in each of these sets is neither uncontroversial nor inconsequential, given that different models of the same decision may lead to different results. The agent may leave something out because they are unaware of its possibility, but also because they do not deem it to be relevant to their decision. While the concept of unawareness has started to attract some research, the notion of relevance is still poorly understood. My second chapter is devoted to clarifying its importance in decisions and to propose a possible interpretation of its meaning.

The last two chapters try to apply to policy making the analyses developed in the first two. In chapter three I use some of the uncertainties proposed in my typology to identify the boundaries of expertise in policy decisions. Understanding these boundaries is critical to provide an assessment of the so-called ‘Evidence Based Policy’ movement, an important trend in development economics that aims to increase the quality of policy making by focusing not on ideologies but on the testable effectiveness of different measures. While the appeal of the movement rests on an idea of experts as impartial scientists that only provide evidence to decision makers, their practice often pushes experts to provide opinions on the policy decision as a whole, thus overstepping the boundaries of the uncertainties over which they have epistemic authority. I illustrate this mechanism with the example of a scholastic policy implemented in India, which has been first tested and then actively supported by researchers.

Finally, chapter four tackles the recent COVID-19 pandemic, as a case in which severe and complex uncertainties made timely, effective policy making particularly difficult. But rather than taking the uncertainty as a reason to delegate policy-making to experts, I argue that different types
of uncertainty all provide reasons to make our collective decision making more inclusive, especially with respect to marginalised groups.

In conclusion: there is more to uncertainty than ignorance about how things are or will be. Uncertainty is plural, as it can concern a variety of aspects and show different properties. It can concern values and priorities, making it not entirely susceptible to being reduced with empirical methods. But while the focus of the thesis is on this plural uncertainty, there are two other recurring themes that it may be helpful to highlight here.

The first is the concept of relevance: besides having its own dedicated chapter, relevance plays a crucial part in model uncertainty. What is deemed relevant has implications not just for individual, but also for collective, public, and even scientific decision making. Overlooking something important and wasting resources on confounding factors are sins of relevance, and understanding the mechanisms of relevance can only help in treating and avoiding them. The second theme is the role of reasons in decision making. Even though I do not devote a specific chapter to reasons, these are central to my accounts of both uncertainty and relevance. The importance reasons have in deliberation is undeniable—and yet, works bridging the philosophy of reasons and decision theory are still very sparse. Hopefully, this thesis on uncertainty can contribute to shed some light on the connections between relevance, reasons, and decisions as well.

**REFERENCES**


**Malvina Ongaro** is a postdoc researcher at Politecnico di Milano. She obtained her PhD in 2022 from the Università del Piemonte Orientale, where she worked mainly in decision theory and philosophy of economics. Her current interests include models of natural risks, multi-criteria decision making, and cost-benefit analysis.

Contact e-mail: <malvina.ongaro@polimi.it>