BOOK OF ABSTRACTS

17th Biennial INEM Conference

Table of Content

Cultural Influences on Economic Methodology: The Intersection of Indigenous Knowledge and Modern Economics in Africa 8

Abdullahi, Kamal Tasiu and Mansour, Mariam Ahmed Mohamed

Philosophical Perspectives on Resource Curse: Lessons from Nigeria's Oil Economy 10

Abdullahi, Kamal Tasiu and Muhammad, Abdulaziz Muhammad

Why not ends paternalism? 12

Abrahamson, Måns

Precision-Like Mesh: On Economic Measures of Discrimination 14

Adams, Florence E

Social Science Denial 17

Angner, Erik

An Argumentative Device for All Sides: The Supply and Demand Model in the Covid-19 Pandemic 19

Aydinonat, Emrah

Model Transfer and Its Challenges in Economics 21

Bakeev, Murat; Herfeld, Catherine; Peruzzi, Edoardo and Šešelja, Dunja

Trust attitudes and the epistemic mechanisms of smartphone addiction 23

Barnes Rodiguez, Angela

Values and Trustworthiness in Economic Inequality Metrics 25

Basso, Alessandra

On the Diversity of Models in Normative Inquiry 28

Beck, Lukas and Jahn, Marcel

Welfare Assessments with Incomplete Preferences: Context-Dependent Choices and Plural Identities 31

Binder, Constanze

A Bigger Game: The Critiques of Market Design 32

Boldyrev, Ivan

Tax cuts for socially valuable work 35

Brouwer, Huub and Van der Deijl, Willem

Underdetermination in experimental game theory: the anti-holistic response 38

Caamano, Maria

What does the Trust Game say about Trust ? 41

Camilotto, Nicolas

Conventions and the performativity of economics **43**

Campos, Jonas

Why Bloomington is not in Virginia: Contrasting the Social Ontologies ofhe Ostroms and Buchanan **45**

Cervesato, Massimo

What is a structural model? **48**

Clarke, Christopher

Qualitative Economics: Integrating Qualitative Research Methods into Economic Analysis **50**

Conte, Leonardo

Collective Action Problems, Self-defeating Behaviors, and their Alleged Equivalence **52**

Costella, Annalisa

PREFERENCES: MENTALISM, BEHAVIORISM AND FUNCTIONALISM **54**

Cozic, Mikael

Coping with uncertainty **56**

Crespo, Ricardo and Tohmé, Fernando

Revising the Moral Arguments Black Reparations in the US in light of the Economic Arguments **58**

Davis, John

Co-Designing Nudges: Measuring the Impact of Expert and Citizen Collaboration **60**

De La Cruz Solal, Samuel and Lecouteux, Guilhem

Is value a substance? A deconstruction of the institutional approach **62**

De Michel, Giustino

Reconceptualizing Individual Agency for the Ecological Transition **64**

Desmarais-Tremblay, Maxime

Experimental elicitation of the 'beliefs' of large language models with uncertain lotteries **66**

Dodivers, Eléonore and Lecouteux, Guilhem

Algorithms and Autonomy: Liberal Responses to the Preference-Shaping Power of Recommender Systems **68**

Dold, Malte

Reaping the Benefits of Interactive Scientific Pluralism: The Case of Ecological Economics **69**

Düring, Jeremias

The Epistemic Boundaries of Economic Informality: Rethinking Methodological Approaches in Economic Policy **72**

Elgin, Ceyhun

Policy Post-Growth and the Lack of Diversity in the Scenario Framework **75**

El Skaf, Rawad

The ”Mortality Paradox”: History and Methodology of Poverty Measurement **77**

Favereau, Judith

A Defence of Preference-Based Approaches to Welfare Evaluations in Public Policy **80**

Fumagalli, Roberto

Channels of Model Performativity **83**

Genta, Benjamin

Assessing Development Projects through the Lens of the Human Development and Capability Approach: How a More Comprehensive Methodological Framework Can Reshape Development Policies. **86**

Ghirlanda, Pietro and Chiappero-Martinetti, Enrica

Why be a frame-sensitive reasoner? On the epistemic foundations of ultraintensional preference formation **89**

Grayot, James

Health, Death, and Well-Being: Toward an Intermediate Account of Well-Being in Terminally Ill Patients **91**

Grinman, Lior Nissim

The Origin of Money – a Methodologically Informed Appraisal of the Debate **93**

Gschwandtner, Jakob

Taxing Carbon, Framing Responsibility: Who Values the Environment? **96**

Guigourez, Mathieu

”Feminist Political Economy and the Migration Choices of Undocumented Latina Women” **98**

Harrington, Aubrial

The insuffciencies of game theory: complementary perspectives by Simone Weil **100**

Heifetz, Aviad

Scientific objectivity and normative economics: the case of fairness theories **104**

Heilmann, Conrad and Wintein, Stefan

Majority voting on nudges **106**

Henschen, Tobias

Construct validity and mechanistic extrapolation in the psychology of poverty **108**

Hokkanen, Lari

Thick Concept, Thin Consensus: Normative Judgments in Estimating Inequality of Opportunity **110**

Hörl, Moritz

Anti-Reductionist Strategies for Macroeconomists **113**

Hoyningen-Huene, Paul

Between economics and philosophy: a reappraisal of the Rawls-Harsanyi debate **115**

Igersheim, Herrade and Carvajalino, Juan

Rethinking Adam Smith's Philosophy: A New Foundation for Political Economy **117**
Issaoui, Ibtissem

Standardizing mean differences: A cautionary tale from a rational choice perspective **119**
Jäntgen, Ina

Like It or Not - Recommender Systems, Welfarism, and Performativity **122**
Khosrowi, Donal and Beck, Lukas

Model Choice in Economics: Simulation Models, Empirical Evidence and (Normative) Representational Decisions **125**
Kothe, Anna-Katharina

When all bets fail (to reveal beliefs): "The Problem of Two Perspectives" **128**
Kudryavtseva, Vita

Who are experts on the sustainability of economic growth? **130**
Lari, Teemu

The Economics of Coercion: Kolm's Liberal Social Contract and Public Goods **133**
Leportier, Kevin

The Value Judgments of Economists and of Economic Agents: The Case for a Distinction. **135**
Lovera, Nestor

How to be (robustly) fair and just **138**
Luttens, Roland; Uzunova, Marina; Wintein, Stefan and Korte, Ruth

From Theory to Practice (and Back Again): Measuring the Value of Life and the Science of Preferences **140**
Maayan, Yam

On the normative authority of economics **143**
Mäki, Uskali

Carroll's Tortoise Against Meta-Decision Theory **145**
Makins, Nicholas

Which values in modern economics? Analytical and methodological strategies for identifying non-epistemic values in economic research **148**
Małecka, Magdalena

When is a replication attempt successful? **151**
Maziarz, Mariusz

Causal presuppositions in econometrics and the science-policy gap **154**
Mireles-Flores, Luis

Original Positions and Decision Theory **156**
Moehler, Michael

Psychological narratives in decision theory: what they are and what they are good for **159**

Moscati, Ivan

Design principles for the IPCC emission scenario ensemble **161**

Nappo, Francesco

Should we really worry about Arrow? On using the axiomatic method in normative inquiry **164**

Nissan-Rozen, Ittay and Enoch, David

Competing Claims with Uncertainty **166**

Nojoumian, Makan

The Relevance of Internalization for Social Norm Interventions **169**

Nuotio, Tatu

Identity, Social Comparison, and Institutional Categorization: A Tripartite Model of Institutional Dynamics **172**

Okamoto, Noriaki

Cost-Benefit Analysis, Incommensurability, and Imprecise Utilities **175**

Ongaro, Malvina and Zanetti, Luca

Big M 2.0. The art of war for philosophers of economics **177**

Ostapiuk, Aleksander

Of the Vices Behind Biases: Reassessing the Contribution of Nudges as Welfare Policy Tools **180**

Rojas, Daniel Zarama

Modeling in Economic Policy: Rethinking Models' Epistemic Limitations **183**

Ruiz, Nadia

Scratching the Surface: The Use of Participatory Methods in the Lab to Uncover Wicked Problems **186**

Rusconi, Massimo and Gagliardi, Lorenzo

On the Utility of Models: Bridging the Theoretical and Analytical Sides of Model Selection **188**

Rusconi, Massimo and Seri, Raffaello

Welfarist Meritocracy **190**

Salonia, Enrico Mattia

Reality bites: Bringing "policy" back to philosophy of evidence-informed policy **192**

Sequeiros, Sofia Blanco

Diversity and strategic beliefs **194**

Slepowronski, Milosz

Rethinking the Rules of the Game: Synthetic Data for Institutional Economics **196**

Špecián, Petr

Should economics as a science be impartial? **199**

Trojanowska, Hanna

Decision Models and Tainting Beliefs **201**

Ulusoy, Oyku

Social Preferences and Social Welfare in Behavioral Welfare Economics **204**

Vromen, Jack

Senian conceptions of agency and commitment for individual choice and welfare assessment in the environmental domain **206**

Wdowin, Julia

Regret aversion **208**

Wilkinson, Hayden

Manipulation in Collective Decision Mechanism: Solutions from Economics and Computer Science **210**

Wollesen, Bele

Negative Externalities in Attention Economies **212**

Yee, Adrian K.

Predictions, Potential Outcomes, and Performativity **214**

Zezulka, Sebastian and Genin, Konstantin

Microfoundations in measurement and theory **217**

Zhu, Junyi

Cultural Influences on Economic Methodology: The Intersection of Indigenous Knowledge and Modern Economics in Africa

Kamal Tasiu Abdullahi*[1] and Mariam Ahmed Mohamed Mansour

[1]Istanbul University – Turkey

Abstract

Frequently, the cultural and historical circumstances from which economic techniques develop are reflected in the methodologies themselves. When it comes to global economic discourse, Western paradigms tend to dominate, which often results in the marginalization of alternative perspectives, particularly those that are anchored in indigenous knowledge systems. The purpose of this study is to investigate the junction of traditional African knowledge and contemporary economics in Africa, and to argue for the use of methodological pluralism as a means of addressing the specific socio-economic difficulties that are faced by the African continent. The communal and relational aspects of African economic practices are frequently overlooked by Western economic models, which are constructed on ideas such as individualism, the maximum of profits, and the efficiency of the market. The reality of economies in which informal sectors predominate and traditional practices play a significant role is not adequately addressed by these models, despite the fact that they are influential. Researchers have the ability to build approaches that are more culturally relevant and practically effective if they include indigenous knowledge into economic analysis. Through- out Africa, traditional economic practices are investigated in this article, with an emphasis placed on the adaptation and durability of these systems. For millennia, communal land ownership systems in Nigeria have been a source of support for the sustainability of agriculture and the social cohesion of the country. In a similar vein, informal financial institutions, such as ”ajo” (rotational savings programs), offer individuals who are not able to access official banking systems the opportunity to obtain loans. The conventional ideas of property rights and financial systems are called into question by these behaviors, which also provide alternative frameworks for the distribution of resources and the transfer of risks. A further il- lustration of the potential of indigenous knowledge in tackling modern difficulties is provided by case studies from Kenya. Women and rural communities have been given the opportunity to participate in the financial system through the establishment of community-based financial systems, such as ”chama” (self-help groups). In stark contrast to the individualistic orientation of mainstream financial institutions, these systems are based on trust, mutual accountability, and shared aims. All of these elements are essential to their operation. It is possible for governments to improve social fairness and financial inclusion by incorporating these concepts into official economic policies. The theoretical insights that come from the field of philosophical anthropology offer a framework that can be utilized to comprehend the epistemic structures that underpin indigenous knowledge. These foundations, which have their origins in relational and holistic worldviews, place an emphasis on the interconnection of both individuals and communities, as well as the environment. These kinds of perspectives

*Speaker

are especially pertinent when it comes to addressing issues such as the alleviation of poverty and the sustainability of the environment, which are areas in which Western techniques frequently fall short. The incorporation of indigenous knowledge into mainstream economic discourse still confronts considerable hurdles, despite the fact that it has the potential to be beneficial. Non-Western knowledge systems are frequently devalued due to epistemic biases, which have their origins in the history of colonialism. Both the inertia of institutions and the commercialization of knowledge are further obstacles that stand in the way of efforts to incorporate indigenous customs into formal economic programs. For the purpose of overcoming these obstacles, the article proposes for research approaches that incorporate collaborative efforts and actively involve local communities. Not only does this participatory method verify indigenous knowledge, but it also guarantees that policies are acceptable in terms of both culture and context. The importance of implementing legislative changes to safeguard and advance indigenous knowledge systems is another point that is emphasized in the report. Take, for example, the incorporation of traditional agricultural practices that have been demonstrated to be beneficial in the management of resources in a sustainable manner into national development strategies. In a same vein, financial regulations ought to acknowledge and promote informal savings and credit systems, including them into more comprehensive frameworks for economic inclusion. In conclusion, the research emphasizes the significance of cultural epistemology in the process of forming economic methodology. It is possible for scholars and policymakers to construct economic models that are not only theoretically sound but also practically useful in tackling the complex socio-economic realities of Africa if they demonstrate an appreciation for indigenous knowledge and include it into their models. This method, which is pluralistic, has the potential to revolutionize economic research and policies, thereby creating more inclusion and sustainability.

Keywords: Indigenous knowledge, cultural epistemology, economic methodology, Africa, Nigeria

Philosophical Perspectives on Resource Curse: Lessons from Nigeria's Oil Economy

Kamal Tasiu Abdullahi*[1] and Muhammad Abdulaziz Muhammad*†

[1]Istanbul University – Turkey

Abstract

Within the realm of economic theory, the concept of the "resource curse" continues to be a contentious issue, with significant repercussions for countries that are abundant in re- sources. Nigeria, which is one of the major oil producers in Africa, offers a compelling case study that can be used to investigate the philosophical aspects of this phenomena. The resource curse is investigated from a philosophical point of view in this study. It challenges the deterministic interpretations of the resource curse and highlights the socio-economic and political foundations upon which it is based. Traditionally, the resource curse has been thought to be a contradiction in which countries that are abundant in resources endure slower eco- nomic growth and greater political instability than their counterparts that are deficient in resources. On the other hand, the purpose of this study is to show that the resource curse is not a natural consequence of the abundance of resources; rather, it is a construct that is formed by governance failures, weak institutions, and distribution of resources in an unequal manner. The study highlights the need for a more nuanced understanding of the resource curse that takes into consideration contextual socio-political factors. This is accomplished by taking a critical stance against the fundamental assumptions that are underlying ortho- dox economic theories. The study examines the oil-dependent economy of Nigeria through the prism of Africana philosophy and social ontology. The country of Nigeria serves as the main point of the investigation. It investigates the ways in which historical legacies, such as the exploitation of resources by colonial powers, have influenced the economic practices of the present day. For example, colonial extraction strategies gave priority to the exploitation of resources for export, which resulted in a legacy of reliance and underdevelopment. For a better understanding of the issues that Nigeria is currently facing in terms of resource governance, this historical backdrop is necessary. In order to gain a better grasp of the sociopolitical aspects of the resource curse, the region of the Niger Delta can be considered a microcosm. Within the scope of this study, topics such as revenue allocation, corruption, and environmental injustices are investigated, with the goal of shedding light on the intricate relationship that exists between resource richness and socio-economic growth. The damage of the environment that has been caused by oil exploration has had a significant impact on the livelihoods of regional communities, which has contributed to the worsening of poverty and social instability. In the study, existing policy frameworks are criticized, and it is argued that these frameworks frequently place a higher priority on short-term economic advantages than they do on long-term sustainability and social equality. As an illustration, revenue-sharing agreements between the federal government and oil-producing states have not been successful in addressing the fundamental factors that contribute to regional imbalances and disputes

*Speaker

†Corresponding author: optimisticmuhammadam@gmail.com

with one another. In addition to this, the study investigates the ethical aspects of resource management and raises questions about the role that multinational firms play in maintaining economic dependence and causing environmental damage. When it comes to social responsibility, these firms frequently put profits before of social responsibility, which contributes to the exploitation of local communities and natural resources. For the purpose of ensuring that the abundance of resources is beneficial to all stakeholders, the study advocates for increased accountability and transparency in the operations of corporations. This study suggests a philosophical rethinking of resource management policies as a means of addressing the issues that have been presented. In doing so, it promotes participatory government models that put the welfare of local communities at the forefront and encourage equitable distribution of resources. For instance, community-based resource management systems, which incorporate local stakeholders in decision-making processes, have demonstrated that they have the potential to improve accountability and sustainability. Additionally, the study underlines the significance of incorporating ethical issues into economic research, stating that resource management has to strike a balance between economic efficiency and social fairness. Moreover, a consideration is given to the deeper implications that the resource curse has for the theory of global economics. Instead of relying on deterministic models, it is necessary to conduct contextualized studies that take into account the specific sociopolitical and historical realities of countries that are abundant in resources. In conclusion, overcoming the resource curse demands not just the implementation of technical solutions, but also a thorough reevaluation of the principles and priorities that form the basis of resource management practices.

Keywords: Resource curse, Africana philosophy, social ontology, oil economy, Nigeria

Why not ends paternalism?

Måns Abrahamson

Erasmus University Rotterdam – Netherlands

Abstract

Behavioural paternalism (BP) is based on a strong commitment to means paternalism and a rejection of ends paternalism (see, e.g., Conly 2013; Le Grand & New 2015; Bernheim 2016; Thaler & Sunstein 2021). Means and ends paternalism, respectively, are roughly understood as ”paternalism directed at the decisions people make to achieve certain goals or ends, and paternalism directed at the ends themselves” (Le Grand & New 2015: 27). While means paternalism is seen as permissible, three interrelated reasons are standardly invoked against ends paternalism in the BP-literature. (1) The available empirical evidence does not indicate that individuals fail in their selection and specification of ends-e.g., ”(behavioural economic) findings do not support ends paternalism” (Sunstein 2014: 26). (2) There can- not, anyways, be any empirical evidence indicating ends-related failures of individuals since ”the occasional objection to (an individual’s judgements about ends) entails nothing more than a difference in opinion between the analyst and the consumer as to what constitutes a good or fulfilling life” (Bernheim 2016: 18). (3) Ends paternalism would further violate the commitment of a liberal state ”to promote no particular conception of the good” (Le Grand & New 2015: 28n5).

This article challenges the received view in BP and makes a case for a constrained form of ends paternalism. This case is based on a recognition that individuals, plausibly, do not only have ultimate ends that they seek to efficiently promote in their actions, but a hierarchy of ends with multiple types of ends nested in each other (e.g., Raz 1986; MacDonald 1991; Frankfurt 1992; Schmidtz 1994). In individuals’ hierarchies of ends, some ends are ultimate ends (pursued for their own sake), while others are instrumental ends (pursued for the sake of promoting some other end), or constitutive ends (pursued as part of some other end), or some combination of the three. What follows from this is that we can meaningful talk about ‘ends-related failures’ in relation to some ends of individuals, where an individual’s selection and specification of her ends undermines her pursuit of her own conception of the good life. This directly challenges reasons two and three in the received view.

This article focuses on two such ends-related failures. The first type of failure concerns an individual selecting an instrumental or constitutive end that is not efficiently promoting associated (further) ends of the individual. More specifically, in the case of instrumental ends, the selected end is not efficiently promoting the further end(s) for the sake of which it is pursued; in the case of constitutive ends, the selected end (while possibly appropriately constitutive of some other end) scores poorly on so-called ‘subsidiary criteria’ compared to alternative (similarly appropriate) constitutive ends available to the individual (see Schmidtz 1994: 228n4). The second type of failure concerns an individual specifying an end in a self- undermining way, in the sense that she specifies the end in a manner that makes it less likely (compared to alternative specifications) that she will realise the end.

* Speaker
Corresponding author: m.abrahamson@hotmail.com

More specifically, I focus on instances of individuals holding 'vague ends', which are problematic for their lack of specification (see Tenenbaum 2020). I argue that these types of ends-related failure, rather than means-related failures, can be seen to characterise a significant portion of the everyday failures that individuals are subject to in pursuing their own idea of the good life. Ends- related failures are therefore of significant relevance for any paternalistic programme aimed at making individuals 'better off, as judged by themselves'. I further argue that there is, by now, substantial empirical evidence in psychology and behavioural economics that supports this perspective on ends-related failures. This directly challenges reason one in the received view.

This article concludes by presenting the contours of a constrained form of ends paternalism in light of the preceding discussion. This includes a discussion of new knowledge problems associated with making credible attributions of ends-related failure as well as a discussion of appropriate methods of intervention for a paternalism targeting faulty ends.

References

Bernheim, D. B. 2016. "The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics." Journal of Benefit Cost Analysis 7(1): 12–68.

Conly, S. 2013. Against Autonomy: Justifying Coercive Paternalism. New York, NY: Cam- bridge University Press.

Frankfurt, H. 1992. "On the Usefulness of Final Ends." Iyyun: The Jerusalem Philosophical Quarterly 41: 3–19.

Le Grand, J., & B. New. 2015. Government Paternalism: Nanny State or Helpful Friend? Princeton, NJ: Princeton University Press.

MacDonald, S. 1991. "Ultimate Ends in Practical Reasoning: Aquinas's Aristotelian Moral Psychology and Anscombe's Fallacy." The Philosophical Review 100(1): 31–66.

Raz, J. 1986. The Morality of Freedom. New York, NY: Oxford University Press. Schmidtz, D. 1994. "Choosing Ends." Ethics 104(2): 226–251.

Sunstein, C. R. 2014. Why Nudge? The Politics of Libertarian Paternalism. New Haven, CT: Yale University Press.

Tenenbaum, S. 2020. Rational Powers in Action: Instrumental Rationality and Extended Agency. New York, NY: Oxford University Press.
Thaler, R. H., & C. R. Sunstein. 2021. Nudge: The Final Edition. New York, NY: Penguin Books.

Keywords: behavioural welfare economics, paternalism, instrumental rationality, wellbeing, public policy

Precision-Like Mesh: On Economic Measures of Discrimination

Florence E. Adams*

Department of History and Philosophy of Science, University of Cambridge – United Kingdom

Abstract

In August 1969 Joan Vuyanich filed a class action case with the Equal Opportunities Employment Commission against her previous employer, the Republic National Bank in Dallas. Vuyanich v National Bank is most notable for its extensive treatment of statistical issues: almost 80 pages of the 127-page opinion deals with the mathematics of regression analysis. In his opinion, Judge Higginbotham complained of the 'precision-like mesh of numbers' that 'tends to make fits of social problems' (Fienberg 1989: 25). How have economic models of discrimination become so central to legal and political decision-making in the twentieth century? How precise is this 'precision-like mesh'? And what is at stake in this meeting of two cultures?

My starting point is a longstanding debate in constitutional legal theory which concerns the normative foundations of antidiscrimination law. Fiss (1976) distinguishes two conceptions of anti-discrimination as a political project. The dominant, anti-classification principle views discrimination as a procedural wrong and equates 'equality' with neutrality in decision- making. Instead, he advocates for an anti-subordination principle, which recognizes that true equality cannot exist under conditions of extreme social stratification. It therefore shifts focus away from formal procedural fairness towards the substantive effects of laws, allowing for non-neutral decision-making if it mitigates group disadvantage (Siegel and Balkin 2003, Siegel 2003, Sunstein 1994).

My paper has two strands: one historical and one philosophical. In the former, I outline the development of economic models of discrimination in twentieth-century economics, from early rational-choice models (Becker 1957, Arrow 1972, Phelps 1972) to regression models and their entry into discrimination litigation in the 1980s, and up to recent causal and experimental measures. More important than what has changed across these shifts, I argue, is what has remained constant: a conception of discrimination as a narrow political project aimed at decision neutrality (i.e. an anti-classification conception). Discrimination was first operationalized as behaviour arising from faulty or irrational psychological processes, such as bias, during the 'boom' of behavioural sciences in the early Cold War. By the late twentieth century, this theoretical account was operationalized in causal terms: multivariate regression analyses of outcome inequality sought to quantify discrimination by isolating the 'pure' causal effect of social categories, which was taken to approximate decision-maker bias. Drawing on existing sources I argue for the historical and political contingency of these measures, and the influence of sociopolitical pressures on early Cold War social science on their elevation to 'gold-standard' status.

*Speaker

In the latter strand, I argue that exogeneity assumptions in economic models of discrimination directly track normative commitments regarding the scope of anti-discrimination as a political project. To develop this claim, I draw on recent literature on 'representational risk' (RR) in scientific modelling (Winsberg 2018, Winsberg and Harvard 2022) – the risk of making a representational decision that is inadequate for purpose. I argue that the way economists have navigated this risk reflects a market-structured notion of fairness whereby 'fair' decisions are benchmarked against market outcomes. This creates concerns when lawyers and policymakers outsource uncertainty to economists, since this conception of fairness embedded in economics models may be inadequate for many of the purposes towards which it is put, including, for instance, guiding the interpretation of otherwise ambiguous legal provisions.

In developing this claim, I intend to provide an alternative lens with which to interpret both i) contemporary methodological disputes in the economics of discrimination (e.g. Knox, Low and Mummolo's (2020) critique of Fryer (2019) or more general concerns about post- treatment bias) (1)and ii) conceptual critiques of causal measures of inequality grounded in the claim that they make a 'gross category error' about what categories like race and gender reference (Kohler-Hausmann 2018, Hu 2020).(2) These critiques occur respectively at too low and too high a level of abstraction. The former obscures a crucial epistemic limitation faced by social scientists seeking a causal framework for discrimination: observational data undermines causal structure unless supplemented by additional, normative commitments. The latter obscures a crucial normative consideration: conceptual commitments in social science can be resolved only against a background normative framework. My paper fills this gap by developing an explicit account of the normative foundations of causal measures of discrimination, which can highlight deep continuities between technical and theoretical disputes about discrimination.

(1) For a canonical such case, see Knox, Lowe and Mummolo's (2020) critique of Fryer (2019) in which they argue that 'existing empirical work in this area is producing a misleading portrait of evidence as to the severity of racial bias in policing' (620). See also Naryanan (2022).

(2) I take this claim to be implicitly recognized in methodological concerns around post- treatment bias (e.g., Gaebler et al. 2022)

References

Arrow, K. (1973). "The Theory of Discrimination." Pp. 3–33 in Discrimination in Labor Markets, edited by Ashenfelter, O., & Rees, A. Princeton, NJ: Princeton University Press.

Balkin, J. M., & Siegel, R. B. (2003). The American civil rights tradition: Anticlassification or antisubordination. Issues in Legal Scholarship, 2(1).

Becker, G. S. (1957). The economics of discrimination. University of Chicago Press.

Fiss, O. M. (1976). Groups and the equal protection clause. Philosophy & Public Affairs, 107-177.

Fryer Jr, R. G. (2019). An empirical analysis of racial differences in police use of force. Journal of Political Economy, 127(3), 1210-1261.

Gaebler, J., Cai, W., Basse, G., Shroff, R., Goel, S., & Hill, J. (2022). A causal framework for observational studies of discrimination. Statistics and Public Policy, 9(1), 26-48.

Harvard, S., & Winsberg, E. (2022). The epistemic risk in representation. Kennedy Institute of Ethics Journal, 32(1), 1-31.

Hu, L. (2023). What is "race" in algorithmic discrimination on the basis of race?. Journal of Moral Philosophy, 21(1-2), 1-26.

Kohler-Hausmann, I. (2018). Eddie Murphy and the dangers of counterfactual causal thinking about detecting racial discrimination. Nw. UL Rev., 113, 1163.

Knox, D., Lowe, W., & Mummolo, J. (2020). Administrative records mask racially biased policing. American Political Science Review, 114(3), 619-637.

Narayanan, A. (2022). The limits of the quantitative approach to discrimination. James Baldwin lecture (transcript), Princeton University.

Phelps, E. S. (1972). The statistical theory of racism and sexism. The American Economic Review, 62(4), 659-661.

Siegel, R. B. (2003). Equality talk: Antisubordination and anticlassification values inconstitutional struggles over Brown. Harv. L. Rev., 117, 1470.

Sunstein, C. R. (1994). The anticaste principle. Michigan Law Review, 92(8), 2410- 2455.

Winsberg, E., & Harvard, S. (2024). Scientific models and decision making. Cambridge University Press.

Keywords: economic models of discrimination, causal inference, representational risk, fairness, exogeneity assumptions, statistical evidence, regression analysis

Social Science Denial

Erik Angner*

Stockholm University – Sweden

Abstract

”Science denial” has been defined as ”the dismissal of well-established scientific evidence or the scientific method as a means to gather reliable evidence” (1). Relatedly, ”science skepticism” has been defined as ”the systematic and unwarranted rejection of science” (2). Dismissal or rejection can take multiple forms. One way to exhibit science denial or skepticism is to explicitly reject well-established scientific evidence, findings, or methods. Another way is to simply ignore it and pretend it doesn't exist. Much of most of the literature on science denial or skepticism has focused on certain particularly controversial, policy-relevant natural and medical sciences, notably climate science, evolutionary science, and vaccine research. Much less attention has been given to social- science denial or skepticism: the unwarranted dismissal of well-established evidence, findings, or methods.

This paper situates critical attitudes to social and behavioral science – especially of course economics – in the context of the literature on science denial and skeptics. The central thesis is that, given the established meanings of these terms, social-science denial or skepticism is pervasive, even among people who are themselves scientists or who would otherwise consider themselves ”for” science, and that there's no reason to think social-science denial is less harmful than the denial of natural or medical sciences.

An obvious instance of social-science denial is the literature that's come to be known as anti-economics, which denounces its practitioners, attacks its assumptions, rejects its conclusions, and protests its influence (3), (4). A much less obvious instance is literature that simply ignores well-established social-scientific evidence, findings, or methods. The latter category is arguably more insidious, because it doesn't wear its science skepticism on its sleeves.

The paper will review a series of examples of social-science denial or skepticism drawn from the pandemic, explore the harms that such behavior might have caused, and try to draw some more general conclusions.

Reference

K. M. Jylhä, S. K. Stanley, M. Ojala, and E. J. R. Clarke, ”Science Denial,” Eur. Psychol., Dec. 2022, Accessed: Mar. 03, 2025. (Online). Available: https://econtent.hogrefe.com/doi/10.1027/1016-9040/a000487

B. T. Rutjens and R. van der Lee, ”Spiritual skepticism? Heterogeneous science skepticism in the Netherlands,” Public Underst. Sci., vol. 29, no. 3, pp. 335–352, Apr. 2020, doi: 10.1177/0963662520908534.

*Speaker

W. O. Coleman, Economics and its enemies: Two centuries of anti-economics. London: Palgrave Macmillan, 2002. Accessed: Nov. 28, 2019. (Online). Available: https://link.springer.com/book/10.1007/ 1-4039-4148-0

D. Ross, ”Economic theory, anti-economics, and political ideology,” in Philosophy of Eco- nomics, U. Mäki, Ed., Amsterdam: North-Holland, 2012, pp. 241–285. doi: 10.1016/B978- 0-444-51676-3.500105.

Keywords: Knowledge resistance, science denial, economics, behavioral science

An Argumentative Device for All Sides: The Supply and Demand Model in the Covid-19 Pandemic

Emrah Aydinonat*[1]

[1]University of Helsinki – Finland

Abstract

At the onset of the Covid-19 pandemic, there was a shortage of personal protective equipment (PPE), such as face masks, just when they were most needed by healthcare professionals and individuals particularly vulnerable to infection. The PPE shortage was followed by sharp price increases, allegations of 'price gouging', the appearance of fake and low-quality PPE on the market, and various policy decisions that influenced the course of the pandemic. Interestingly, articles and opinion pieces on PPE shortage used the very same supply and demand (SD) model to explain the situation, but the policy recommendations contradicted each other. The actual policy responses were similar. While many countries decided to limit or ban exports, many others reduced or removed tariffs to facilitate imports (Evenett 2020, Hoekman et al. 2020). One problem, two very different policy responses-both justifiable using the SD model!

The SD model is one of the most used models in economic explanations and occupies the central stage in Econ 101 courses. In a typical textbook (e.g., Mankiw, 2009), students are introduced to questions about how policy interventions affect market outcomes, using the SD model, often accompanied by real-world examples. During the pandemic, for example, instructors were encouraged to use it to explain the toilet paper shortages and evaluate government policies (Zhang & Ramse, 2021). Out of the classroom, the SD model is frequently used by policy analysts and researchers to interpret current events and societal problems. For example, during the pandemic, it has been used to explain inflation and to argue that businesses should not be blamed (Dubay, 2022). It has also been used to suggest that a face mask recommendation is superior policy than a universal face mask mandate–since it produces more consumer surplus (Goel and Haruna 2021, p. 580)

The use of simple economic models in policy contexts is, of course, not new. Critics have long drawn attention to their misuses in real-world policy contexts, calling this practice "economism" or "Econ 101ism" (e.g., Kwak, 2017; Smith, 2015, 2016). The underlying criticism is most familiar to philosophers of economics: the SD model employs unrealistic assumptions and its explanatory power and policy relevance are crippled by these assumptions. However, despite the vastness of the literature on idealised models, philosophers of economics did not pay enough attention to the most influential model in economics: the SD model (see, however, Boland, 2017; Hausman, 1990; Jhun, 2018; Kincaid, 2004). The present paper builds on previous research on the topic in an attempt to close this gap and examines the uses (and misuses) of the SD model in the context of the Covid-19 pandemic.

*Speaker

The paper shows that simple idealised models, such as the SD model, can be easily used, and are used, to support policy conclusions that cannot be derived from them. By combining the SD model with additional Econ 101 concepts, such as externalities or the invisible hand, one can argue for or against government intervention in response to a shortage. However, neither of these conclusions can be fully justified with the SD model itself. Of course, this point – that a simple model can be used to support policy conclusions that cannot be directly derived from it – is not surprising. What is surprising is that while the explanatory and ex- ploratory functions of idealised models have been widely discussed, there is little discussion concerning the argumentative functions of idealised models. The present paper examines the argumentative functions of the SD model and the structure of the political arguments using it; shows that simplicity of the SD model and its familiar logic makes it a good argumentative device that lends itself to contradictory uses; and argues that even an explanatory model such as the SD model is not necessarily a good model for settling policy debates. It thereby illustrates of how simple models are used as argumentative devices for policy purposes and proposes a way to expose the gaps and leaps in model-based policy argumentation.

References

Boland, L. A. (2017). Equilibrium models in economics: Purposes and critical limitations Oxford University Press.

Dubay, C. (2022, March 1). Econ 101: Inflation is Caused by Supply and Demand . U.S. Chamber of Commerce.

Goel, R. K., & Haruna, S. (2021). Unmasking the demand for masks: Analytics of man- dating coronavirus masks. Metroeconomica , 72 (3), 580–591.

Evenett, S. J. (2020) "Flawed prescription: Export curbs on medical goods won't tackle shortages" in Baldwin R. & Evenett S. J. (2020) *COVID-19 and Trade Policy: Why Turning Inward Won't Work*, CEPR Press, pp. 49-61.

Hausman, D. M. (1990). Supply and demand explanations and their ceteris paribus clauses. Review of Political Economy , 2 (2), 168–187.

Hoekman, B. et al. (2020) "COVID-19: Export controls and international cooperation" in Baldwin R. & Evenett S. J. (2020) *COVID-19 and Trade Policy: Why Turning Inward Won't Work*, CEPR Press, pp. 77-88.

Jhun, J. S. (2018). What's the Point of Ceteris Paribus? Or, How to Understand Supply and Demand Curves. Philosophy of Science , 85 (2), 271–292.

Kincaid, H. (2004). There are laws in the social sciences. In C. Hitchcock (Ed.), Contemporary Debates in Philosophy of Science (pp. 168–186). Blackwell.

Kwak, J. (2017). Economism: Bad Economics and the Rise of Inequality. Pantheon.

Mankiw, N. G. (2009). Principles of microeconomics (5th ed). Cengage Learning.

Smith, N. (2015). Most of What You Learned in Econ 101 Is Wrong. Bloomberg.Com . https://www.bloomberg.com/opinion/articles/2015-11-24/most-of-what-you-learned-in-econ- 101-is-wrong

Smith, N. (2016). Noahpinion: 101ism. Noahpinion https://noahpinionblog.blogspot.com/2016/01/101ism.htm

Zhang, C., & Ramse, J. (2021). Teaching economics behind the global COVID-19 pandemic. International Review of Economics Education , 36 , 100206.

Keywords: supply, demand model, models, policy, argumentation

Model Transfer and Its Challenges in Economics

Murat Bakeev[*†1], Catherine Herfeld[*1], Edoardo Peruzzi[*1], and Dunja Šešelja[2,3]

[1]Leibniz University Hannover – Germany
[2]Eindhoven University of Technology – Netherlands
[3]Ruhr-Universität Bochum – Germany

Abstract

Recent discussions in philosophy of science have highlighted the importance of cross- domain model transfer as a key characteristic of modern science and a driver of progress. Prominent examples are the introduction of psychological models into economics to model various motives behind human decision-making, the application of game-theoretic models from mathematics into evolutionary biology or models from physics and complexity theory into mathematical finance. While case studies on model transfer have flourished, the philosophical discussion – however - remains in its early stages. The aim of this symposium is to take the debate one step further. By focusing on cases from economics, there are at least two open issues in particular that we will address in our symposium.

One open issue concerns the unit of transfer: in some cases, specific models-such as the Ising model-are transferred. In other cases, what is transferred are highly generic models, like coupled harmonic oscillators, Lotka-Volterra equations, or game-theoretic frameworks. Accordingly, some philosophers argue that the transferred unit is the scientific model it- self, while others suggest that what is transferred is a more abstract structure-or what Paul Humphreys and others call a template (e.g., Houkes and Zwart 2019, Humphreys 2019, Knuuttila and Loettgers 2014, 2023). Such diverging views raise questions such as what templates exactly are, how they relate to scientific models, and what enables their transfer across distinct domains to solve fundamentally different problems.

A second open issue concerns the transfer process itself, which often involves modifications rather than direct re-application of models. Additionally, the impact of model transfer on scientific progress varies across fields, with some transfers facing significant challenges or even failing. Economics exemplifies these challenges, as models imported from physics, sociology, and psychology often encounter substantial barriers to entry or fail to integrate successfully. Even when models are adopted, their transfer does not always yield beneficial epistemic effects.

This symposium not only addresses questions related to the unit of transfer and the characteristics of the transfer process, but also examines the challenges that model transfer faces and the conditions under which it contributes to progress in economics. As such, our contributions will also link key questions in the debate to broader discussions in the philosophy of economic modelling.

---

* Speaker
† Corresponding author: murat.bakeev@philos.uni-hannover.de

References

Houkes, Wybo; Zwart, Sjoerd D. (2019): Transfer and Templates in Scientific Modelling, Studies in History and Philosophy of Science Part A, 77, 93-100.

Humphreys, Paul (2019): Knowledge Transfer Across Scientific Disciplines, Studies in History and Philosophy of Science Part A, 77,112–119.

Knuuttila, Tarja; García-Deister, Vivette (2019): Modelling Gene Regulation: (De)compositional and Template-based Strategies, Studies in History and Philosophy of Science, 77, 101–111. Knuuttila, Tarja;

Loettgers, Andrea (2014): Magnets, Spins, and Neurons: The Dissemination of Model Templates Across Disciplines, The Monist, 97, 280–300.

Knuuttila, Tarja; Loettgers, Andrea (2023): Model Templates: Transdisciplinary Application and Entanglement, Synthese, 201 (6), 200.

Participants/Presenters

Murat Bakeev (murat.bakeev@philos.uni-hannover.de), Leibniz University Hannover, Germany.
Title: ”Methodological Barriers to Model Transfer and Their Mitigation: The Case of Agent-Based Models in Macroeconomics.”

Catherine Herfeld (catherine.herfeld@philos.uni- hannover.de), Leibniz University Hannover, Germany.
Title: ”Challenges to Model Transfer in Science.”

Edoardo Peruzzi (edoardo.peruzzi@philos.uni- hannover.de), Leibniz University Hannover, Germany.
Title: Model Transfer Through Formal Template Accumulation: The Case of Game Theory.”

Keywords: model transfer, template, scientific progress, game theory

Trust attitudes and the epistemic mechanisms of smartphone addiction

Angela Barnes Rodriguez*[1]

[1]Arizona State University – United States

Abstract

So much of technology ethics is about the production of technology. What should we be creating? How can we do so ethically? While these questions are important for technology developers, I focus instead on the consumption of technology. We are currently saddled with a wealth of technologies for which no one asked these important questions, and people will continue to create technologies that don't abide by ethical rules. In the meantime, we have to understand how we can be responsible consumers of manipulative technologies. One of the most prominent places this enters people's lives is through their smartphones - hence the emerging epidemic of smartphone addiction. It is becoming increasingly clear that smart- phone addiction is more than a metaphor. The effects on users' neurological systems, health, and long term capabilities are well documented, but very little philosophical work has been done to explore smartphone addiction.

We have an immense amount of trust in our smartphones. Unfortunately, the trust that we have in our phones reinforces some of the cognitive dysfunctions characteristic of addiction. These mechanisms make it difficult for addicts to stop using, and make it difficult to stay stopped if they succeed in stopping. One cognitive dysfunction that is central to addiction is denial (Picard, 2016). Addicts continue to use despite negative consequences partially because they have not epistemically discovered the consequences of their use. While we might see the negative consequences in our lives, we fail to see them as being direct results of addictive use. If we see the smartphone as being trustworthy, then we will not believe it to be the cause of the negative consequences in our lives. Another cognitive dysfunction characteristic of addiction is judgment shift, which are moments where we recalculate our reasons for belief or action – without encountering new evidence – and come to change our minds (Levy, 2019). Judgment shifts happen when we find escape clauses in our resolutions because of the presence of an addictive substance or device. Through notifications, the smartphone speaks to us to get our attention, and supplies us with an endless stream of reasons to use. Smartphone addiction is particularly resistant to resolutions because of the endless escape clauses (triggers) the smartphone offers.

Recovery from addiction requires that we find ways around the cognitive dysfunctions that come from addiction. We have to stop trusting our minds, and learn how to manage our capacities as an observer. This is why so many recovery methods, secular and non-secular, operate around concepts like "HALT", pausing, meditation, or prayer. In the meantime, addicts in recovery have to reassemble their trust attitudes elsewhere. Interpersonal trust must be built, and the trust in one's capacities must diminish. I argue, in the case of smartphone addiction, that this requires also that we lose our attitudes of trust in the phones themselves.

*Speaker

To break the smartphone addiction epidemic, we have to fundamentally break the trust between people and their smart devices. We should not trust them to offer us new 'goods' or applications. We should not trust them to soothe our children. We should not trust them to guide us or provide us with correct information. There are also ways in which smartphones keep us from building interpersonal trust - the other has become increasingly suspicious, in part because of the way we see each other online.

To lose this trust is epistemically rational and normatively fit as smartphones should not have our trust in the first place. It is well known that smartphones and many of the applications on them are designed to manipulate. Yet still we treat them as though they have our best interest in mind: we let them tell us what we should trust them to do - an attitude usually reserved for our closest familial relationships.

Part of a technology ethics education should be dedicated to being a responsible consumer of technology. Responsible consumption is not done in isolation, as we have to have people who can teach us how to manage our cognitive functions and dysfunctions. This education is immensely personal, and much of it will need to be done outside the classroom. But in the classroom we can teach people how to identify trust that is unfounded, recognize the costs behind manipulative technologies, and to see the value of communities built around responsible consumption.

Keywords: Technology, addiction, education

Values and Trustworthiness in Economic Inequality Metrics

Alessandra Basso*†[1]

[1]University of Cambridge – United Kingdom

Abstract

Are empirical findings in the social sciences always dependent on underlying normative assumptions, or can certain empirical claims command consensus despite value disagree- ments? This paper investigates the relationship between values and trustworthiness in social indicators, using the World Inequality Report 2022 (WIR2022) as a case study. The report's ambition to provide objective claims on economic inequality seems at odds with the prevailing view among philosophers of science that values are pervasive and inescapable in scientific inquiry. This tension raises questions about the trustworthiness of economic inequality metrics in contexts of value pluralism.

The valueladenness of social research raises two major concerns. The first is about autonomy: if scientific practices embed researchers' own values, this may undermine the autonomy of users - such as policymakers and citizens - to form their own judgments. The second is a concern about deliberation: if users disagree about values, reliance on value-laden metrics risks privileging certain perspectives over others. Two influential proposals have recently been advanced to address these concerns.

The first proposal is 'democratic alignment', which demands that scientific research align with values that are democratically endorsed by relevant stakeholders. This approach promises to respect users' autonomy and address value disagreements through deliberation (Alexandrova 2017; Alexandrova and Fabian 2022). The second proposal is 'value pluralism' (Thoma 2024). Thoma objects to democratic alignment, arguing that it fails to enable proper democratic deliberation. In her view, democratic alignment grants epistemic access to trustworthy measurements only to those who share the democratically endorsed values. To foster genuinely democratic deliberation, Thoma advocates for a plurality of measurements that reflect the range of values held by users.

This paper challenges both proposals on the grounds that they assume what I call 'value conditionality': the idea that measurements are trustworthy only for those who endorse the underlying values. This idea clashes with the ambition of the WIR2022 to produce empirical claims that are trustworthy across moral and political divides. While value conditionality may be appropriate for strongly subjective measurements - such as well-being, quality of life, or patient-centred measures - I argue against its broader application to social research.

In the case of measurements with a strong subjective component, like Patient Reported Outcome Measures, the pre-existing values of respondents play a crucial role. These measurements depend on their capacity to capture the subjective evaluations of those being measured.

*Speaker

†Corresponding author: a.basso@lse.ac.uk

Researchers aim to elicit the perspectives of patients, making the trustworthiness of these measurements contingent on their alignment with respondents' subjective values (McClimans 2024; Rodriguez Duque, Tal, and Barbic 2024). Although such measurements are not entirely individualized, it is reasonable to design different instruments for distinct groups, such as active older adults or pediatric patients with celiac disease.

By contrast, the values relevant to inequality indicators cannot be assumed to pre-exist independently of the measurement itself. People typically hold general moral commitments (e.g., egalitarian vs. meritocratic) alongside more immediate interests (e.g., preferences for low taxation vs. inclusive social benefits). However, it is not straightforward how these values map onto specific choices in the construction of inequality indicators - such as which economic resources to measure (income, wealth, consumption) or which parts of the resource distribution matter most (different weighting systems). Moreover, such choices are not immune to empirical evidence: regularities observed in the world can inform and revise value-laden assumptions (Anderson 2004). For example, Anthony Atkinson justified prioritizing income and wealth inequality by pointing to their correlations with social problems like poor social cohesion and ill health.

Furthermore, people often hold conflicting values regarding inequality - such as endorsing egalitarian principles in the abstract while opposing redistributive taxation in practice. It is unclear how such conflicts could be embedded in research practice, or how users with conflicting values could select the measurement that best represents their own values from a plurality of options.

To address these issues, I propose a more suitable framework for understanding values and trust in economic inequality metrics. I advocate for a pluralistic view of social research, where different measurements are constructed for different purposes rather than to capture different sets of pre-existing values. Examples of such purposes include monitoring trends for international cooperation and foreign aid versus designing national policy. These purposes cluster together a narrower range of values that can be embedded in research practice, constrained by pragmatic, epistemic, and methodological considerations.

Consensus on empirical claims within domains of application can be built through normative principles other than value alignment. One such principle is appeal to regularities, where value-laden choices are justified by observed patterns in the world. Another principle is robustness to value disagreement: when multiple methods yield consistent conclusions across different value-laden assumptions, this provides reason to trust the findings regardless of which values one endorses.

These principles allow empirical claims to achieve broad acceptability within specific con- texts of application, reconciling valueladenness with the aspiration to produce trustworthy evidence. Economists already appeal to these principles when evaluating the stability of their measurements by comparing fallible indicators and examining their relations between theoretical variables. This paper argues that, in doing so, they not only stabilize their measurements but also create consensus over the values embedded in methodological choices.

References

Alexandrova, Anna. 2017. *A Philosophy for the Science of Well-Being*. New York, NY: Oxford University Press.

Alexandrova, Anna, and Mark Fabian. 2022. 'Democratising Measurement: Or Why Thick Concepts Call for Coproduction'. *European Journal for Philosophy of Science* 12 (1): 7. https://doi.org/10.1007/s13194-021-00437-7.

McClimans, Leah. 2024. *Patient-Centered Measurement: Ethics, Epistemology, and Dialogue in Contemporary Medicine*. Oxford Scholarship Online. New York, NY: Oxford University Press. https://doi.org/10.1093/oso/9780197572078.001.0001.

Rodriguez Duque, Sebastian, Eran Tal, and Skye Pamela Barbic. 2024. 'The Role of Ethical and Social Values in Psychosocial Measurement'. *Measurement* 225 (February):113993. https://doi.org/10.1016/j.measurement.2023.113993.

Thoma, Johanna. 2024. 'Social Science, Policy and Democracy'. *Philosophy & Public Affairs* 52 (1): 5–41. https://doi.org/10.1111/papa.12250.

Keywords: normative assumptions, values in science, measurement, inequality

On the Diversity of Models in Normative Inquiry

Lukas Beck*[1,2] and Marcel Jahn

[1]London School of Economics (LSE) – United Kingdom
[2]Leibniz University Hannover – Germany

Abstract

Models are an important tool in many scientific disciplines. As idealized representations, they enable scientists to isolate and systematically investigate specific elements of the world. To date, philosophers concerned with the use of models have focused primarily on those that pursue descriptive goals like explanation or prediction. A key theoretical challenge in this context has been to understand how idealized – and thus factually incorrect – representations of target systems can nonetheless succeed in attaining their descriptive aims (Weisberg 2007; Khosrowi 2020). Yet, in recent years, scholars have come to recognize that models are not limited to descriptive purposes but can also serve important normative aims (e.g, Colyvan 2013; Beck and Jahn 2021; Bykvist and Roussos 2025). As these philosophers highlight, various fields dealing with normative questions – e.g., economics, decision theory, computer science, and ethics – frequently employ models as part of normative analyses.

Now, philosophical research on such normative models is still in its infancy. Recently, how- ever, some philosophers have made an important observation about normative models: they can fulfill a wide range of roles in normative inquiry (see Bykvist and Roussos 2025). Yet, virtually all scholars in the field fail to acknowledge that normative models – depending on their precise role – require very different treatment and pose distinct theoretical problems.

In this paper, we argue that the failure to explicitly disaggregate normative models into different types comes at significant theoretical costs. It leaves scholars without the conceptual tools necessary to articulate and evaluate specific philosophical worries tied to particular types of normative models. To remedy this situation, we propose a typology that categorizes normative models according to their aims. We identify four distinct types: analytical, characterizing, testing, and justifying normative models.

Analytical normative models aim at enhancing our understanding of normative concepts through conceptual exploration. We will argue that various models in social choice theory fall into this category. Characterizing normative models seek to correctly characterize a particular normative domain. A prime example here is Rawls' (1999) ideal theory of justice. Testing normative models are tools that are widely used to test verdicts derived from normative theories. A striking example of this model-type are Parfit's (1984) population ethics models. Finally, justifying normative models are designed to reveal and justify what we should do by illustrating what highly idealized agents would do in idealized circumstances. Various models pertinent to urgent policy matters have such a distinctively justifying ambition. For instance, a key aim of so-called Integrated Assessment Models (IAMs) is to inform

*Speaker

us about, and justify, the optimal level of climate change mitigation (Stern and Stiglitz 2021).

To demonstrate the usefulness of our typology, we explicate the unique theoretical challenges associated with each model-type. We then focus on the success conditions of different normative models, i.e. the conditions under which normative models can meet their aims. As we will suggest, the absence of type-specific theorizing has led to notable confusion and unnecessary disagreements in the debate about the success conditions of normative models. To make progress, we illustrate how our typology can remedy this situation.

First, we argue that a recent account offered by Bykvist and Roussous (2025), which we will call the calibration-confirmation account, is best construed as an account of characterizing normative models. As the name suggests, according to this account, assessing the success of a normative model involves two steps: calibration and confirmation. Calibration refers to checking whether the model is error-free and works as expected. Confirmation is about examining whether the model gets (part of) the normative domain right. According to Bykvist and Roussos, confirmation proceeds by testing the model against our normative judgments. To put the issue in a nutshell, according to the calibration-confirmation account, all the model does is generate normative verdicts that need to be checked against an independent source. As a result of this, we will argue that the calibration-confirmation account provides a good account of the success conditions of characterizing normative models but cannot help us to elucidate the success of analytical, testing, and justifying normative models.

Second, we propose that Beck and Jahn's (2021) extending account gains plausibility when recognized as addressing only justifying normative models. This account centrally posits that a given normative model can only be successful if it operates against a body of independently justified normative verdicts that we already accept. Specifically, the model has to entail a sufficiently large subset of these independently justified verdicts. If this is the case, then the model – by providing a sparse set of idealizing assumptions – effectively summarizes these prior verdicts and also reveals a pattern in them. Based on that, the model can then provide (pro-tanto) justification in cases where we lack firm normative judgments. Yet, the extending account remains silent on the success conditions of the other model-types we identify.

The paper proceeds as follows. Section 2 introduces normative models in more detail and highlights that scholars have, for the most part, treated them as a single, undifferentiated phenomenon. Section 3 develops our four-way typology by reviewing key examples of normative models from the literature. It also outlines the specific challenges tied to the different types of normative models. Section 4 highlights that the typology offers valuable guidance in understanding the success conditions of normative models and advances the ongoing debate on the matter.

References

Beck, L., & Jahn, M. (2021). Normative models and their success. Philosophy of the Social Sciences, 51(2), 123-150.

Bykvist, K., & Roussos, J. (2025). A plea for modelling in ethics. Synthese, 205, 42. Colyvan, M. (2013).

Idealisations in normative models. Synthese, 190(8), 1337–1350.

Khosrowi, D. (2020). Getting serious about shared features. The British Journal for the Philosophy of Science, 71(2), 523-546.

Parfit, D. (1984). Reasons and persons. Oxford University Press.

Rawls, J. (1999). A theory of justice: Revised edition. Harvard University Press.

Stern, N., & Stiglitz, J. E. (2021). The social cost of carbon, risk, distribution, market failures: An alternative approach. National Bureau of Economic Research.

Weisberg, M. (2007). Three kinds of idealization. The Journal of Philosophy, 104(12), 639-659.

Keywords: Normative Models, Success Conditions, Scientific Models

Welfare Assessments with Incomplete Preferences: Context-Dependent Choices and Plural Identities

Constanze Binder*†[1]
[1]Erasmus University Rotterdam (EUR) – Netherlands

Abstract

Rationality, defined as the consistency of context-independent all-things-considered preferences, is a standard assumption in wide parts of welfare economics. In recent decades, findings in behavioral economics casted doubts on the consistency and context-independency of people's preferences as revealed in their choice behavior. In the light of these findings, some have advocated to abandon preferences as the basis of welfare economic judgements (Sugden 2018).

In this article we draw on the literature that explores the relationship between rationality and a plurality of motives (Steedman and Krause 1986, Davis 2011, Arlegi and Teschl 2012, Binder 2014), in order to see under which conditions seemingly inconsistent and context- dependent choice behavior can be the result of a plurality of (possibly conflicting) identities that become salient in different spheres of a person's life.

For this purpose, a social choice framework is adopted to explore a person's intra-personal aggregation process: she aggregates her plural motives or identities, reflected by a profile of strict orderings, into an all-things-considered preference ordering over the alternatives. In a first step it is shown how acyclicity of a person's all-things-considered preferences (and the impossibilities obtained in Binder 2014) can be avoided by allowing a weak form of dictatorial power of parts of a person's identity over different sub-domains of alternatives/ spheres of her life. In a second step, it is shown how such acyclic (and possibly incomplete) all-things considered preferences can lead to context-dependent choice behavior, but nevertheless be employed as the basis for preference-based welfare judgements. In a third, and last step, the obtained insights are discussed in the light of the recent literature on behavioral welfare economics beyond (complete) preferences (Bernheim 2021, Sugden 2021). The article concludes with a discussion how the obtained result connects to the psychological literature on (plural) identities and intrapersonal conflict.

Keywords: behavioural welfare economics, economic agent, rationality, all things considered preference, plural identity, social choice, intra personal aggregation

*Speaker
† Corresponding author: binder@esphil.eur.nl

A Bigger Game: The Critiques of Market Design

Ivan Boldyrev*

Radboud University – Netherlands

Abstract

Market design – the engineering field of microeconomics – has gained much prominence in recent decades. Its application has already generated a lot of interesting methodological literature, inaugurated by Guala (2001) and overviewed in Boldyrev (2025). It deals, in particular, with the implicit understanding of models, the idea of progress, and the performativity of economic modeling. Crucially, market designers' epistemic effort includes the creation of a system that needs to be controlled. The paper will demonstrate that this control is predicated on the ability to account for the factors and options facing the agents outside the interactive mechanism that is being modeled. This brings us into a realm quite familiar to economic methodologists (partial equilibrium and ceteris paribus assumptions) and to historians of economics (the shift from general equilibrium to game theory).

This paper takes a closer look at the recent critical discussion of market design models (Spiegler 2024; Sönmez 2024). It shows that in many theoretical contexts in economics, general equilibrium explanations still form a core of an economic argument. Additionally, in economic theory context, the implicit position of a theorist becomes essential. Thus, Spiegler (2024) argues that mechanism design models are biased in this respect toward the designer's point of view. They restrict the model's implied ontology in ways that would be compatible with this point of view, instead of avoiding any additional constraints. The modeling becomes much more precise in specifying the parameters of economic behavior and in guiding the questions a modeler is willing and able to answer. This exercise becomes arguably less universal and thus could be considered as a disadvantage to a theorist. Moreover, Spiegler (2024) complains that the narrowness of the mechanism design approach – in his case, seemingly making the theory more 'applied' – actually led to introducing less realistic assumptions into the model.

I argue, however, that the constraints imposed by the modification in approach are not merely epistemic. Rather, design economists adopt a more constructivist stance, with models serving as blueprints for design interventions (Rilinger 2023). Here, the question of a normative gap (Hitzig 2020) looms large: are we designing mechanisms that really give people what is best for them and/or what they want? A 'detached' analysis of a competitive equilibrium, offering a more neutral and open-ended attitude, extracts economists from the picture and thus does not deal explicitly with their roles and related responsibilities. But this is not the case in market design, as the designers' reflections on their interventions reveal.

These reflections provide the way to understand the differences in analytic attitudes within market design. The more abstract and neutral stance, refusing to adopt any particular position and guided by the idea to understand what is happening in a mechanism, could be contrasted with a more interventionist perspective –

*Speaker

acknowledge the differences in the objectives of various stakeholders, the former stops there, while the latter focuses on certain objectives in order to participate in policy. The aim could be either to create new institutions or to repair existing ones.

In both cases the need to deal with the outside options – the 'general equilibrium' perspective – turns out to be indispensable. Again, the differences are in the relevant stance. While more theoretically inclined economists merely account for these options, interventionist designers seek to change them (or at least understand how this may be done).

I argue that the key to understand the differences between these perspectives lies not in a traditional division between those believing in spontaneous markets and the constructivists/planners. Rather, the demarcation lies between those who, like Spiegler (2024) follow the neutrality of theory and its separation from 'practice', and those who, like Sönmez (2024) and, to a lesser extent, Roth (2002), are actually involved in pragmatist 'theoretical practices.' To use Peirce's language, the idea of design, but also equilibrium and optimality, becomes the idea of its practical effects, and the 'truth' is reconsidered as something that 'works' (Callon 2007). In this ('performative') perspective, an 'unrealistic' assumption in a model could be considered not only as a flaw, but also as an invitation to change.

The challenges faced by design economics still involve democratic participation and transparency (Basshuysen 2023). But this is the pragmatist stance, I argue, that, while involving more interventionist attitude, actually provides more opportunities to stimulate participation and overcome the risks of non-transparent mechanisms. It does so by acknowledging the role of non-academic, historically rooted practices and path-dependencies in institutional design, but also by aligning theoretical constructs with the needs of stakeholders. At the same time, this 'science as practice' attitude allows us to critically reconsider the risks of capturing the designs, so that the 'bigger picture', following the reflexivity precept of the 'strong programme' (Bloor 1976), can always include the design economists themselves.

References

Basshuysen, P. van (2023) Markets, market algorithms, and algorithmic bias, Journal of Economic Methodology 30(4), 310–321.

Bloor, D. (1976). Knowledge and Social Imagery. London: Routledge & Kegan Paul.

Boldyrev, I. (2025) History and Philosophy of Market Design: Mathematical Politics of Resource Allocation, Prepared for: Yafeng Shan (ed.) History and Philosophy of the Social Sciences, Cham: Springer.

Callon, M. (2007). What does it mean to say that economics is performative? In Do economists make markets? On the performativity of economics, ed. by D. MacKenzie, F. Muniesa, & L. Siu (pp. 311–357). Princeton: Princeton University Press.

Guala, F. (2001). Building economic machines: The FCC auctions, Studies in History and Philosophy of Science Part A, 32(3): 453-477.

Hitzig, Z. (2020). The normative gap: Mechanism design and ideal theories of justice. Economics and Philosophy, 36(3), 407-434.

Rilinger, G. (2023). Conceptual limits of performativity: assessing the feasibility of market design blueprints, Socio-Economic Review, 21(2): 885–908.

Roth, A.E. (2002), The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. Econometrica, 70: 1341-1378.

Spiegler, R. (2024). The Curious Culture of Economic Theory. Cambridge, MA: MIT Press.

Sönmez, T. (2024). Minimalist Market Design: A Framework for Economists with Pol- icy Aspirations https://arxiv.org/pdf/2401.00307

Keywords: market design, performativity, pragmatism, general equilibrium

Tax cuts for socially valuable work

Huub Brouwer*†[1] and Willem Van Der Deijl*‡[2]

[1]Tilburg University [Tilburg] – Netherlands
[2]Department of Philosophy (Tilburg University) – Netherlands

Abstract

Work often has benefits that extend beyond the economic benefits to the employer and the employee. Take an elementary school teacher. If the teacher does a good job at teaching, the students will benefit from this too. And well-educated children benefit society further, for instance by improving the functioning of democracy. This additional value is commonly referred to as the social value of work. Plausibly, the social value generated by some jobs, such as teacher or nurse, is higher than for other jobs, such as e-cigarette designer.
There are substantial labor shortages for socially valuable jobs throughout the industrialized world. In the United States alone, for instance, federal authorities project a shortage of 78.610 registered nurses and 55.000 teachers in 2025. At the same time, there are numerous workers who believe that their work does not contribute to the good of society, or even destroys social value (Graeber 2018). This is not surprising. In a market society that does not specifically steer labor towards socially useful work, we should expect some workers to be working in ways that are not socially useful.

Most liberal societies do not have systematic policies in place to stimulate workers to move into sectors of the labor market that are particularly useful to society (Schneider, Brun, Weber 2020). This is not to say that no such policies exist. A few countries have preferential migration policies for those workers with skills that are particularly useful and in short supply. Also, a limited number of countries promote students going into particular courses of study by giving them a discount on tuition fees or by allowing them to pursue their degree in English. However, one of the most central policy levers – taxation – is not used by any country. In this article, we argue that this is a mistake.

Our main claim is that governments should differentiate income taxation based on the use- fulness of the work for society. One example of this would be tax cuts for those who carry out socially valuable work such as nurses and teachers-but it could also include increased tax rates for those who carry out socially destructive work, such as e-cigarette design. Although economists have to some extent examined differential income tax rates based on the social usefulness of professions (Lockwood, Nathanson, Weyl 2017a), philosophers have so far re- mained silent on the issue. We hope our paper will contribute to changing this.

We examine tax cuts for socially valuable work from the perspective of two central approaches to the economics and philosophy of taxation (Fleurbaey 2018): efficiency approaches (in the tradition of Mirrlees 1971) and fair allocation approaches (inspired by Dworkin 2000; further

*Speaker
†Corresponding author: h.m.brouwer@tilburguniversity.edu
‡Corresponding author: W.J.A.vdDeijl@tilburguniversity.edu

developed by Fleurbaey 2008, chps. 4-5). We argue both approaches strongly favor using taxation to steer people to socially valuable work. We discuss various of setting up a system of differential tax rates based on the social value of jobs, reflecting on the balance they strike between efficiency and fairness.

We subsequently discuss three objections. First, that any concept of socially valuable work will be objectionably partial and, in particular, violate a commitment to neutrality in justification (Kymlicka 1989). Second, that a system of differential income tax rates based on the social value of people's jobs would be impractical and inefficient, because it would require gathering a great deal of information about the work that citizens do (Posner & Weyl 2013). And, third, that tax cuts for socially valuable work are inegalitarian, especially harming those with low pay and limited access to the goods of work (Gheaus & Herzog 2016). We argue that these concerns can be addressed to a large extent by carefully designing the tax, and that none of them are sufficiently good reasons to reject the proposal.

Working in a job that has little benefit to society not only affects the sense of satisfaction people get out of work, but also their sense of meaning, and even their dignity (Althorpe 2022; Smuts 2013; Tyssedal 2023; Wolf 2010). The fact that many workers have the sense that their work does not contribute much to society is preposterous in a world in which there is still so much that labor can contribute to meeting our human and societal needs.

References

Althorpe, Caleb. 2022. "Meaningful Work, Nonperfectionism, and Reciprocity." *Critical Review of International Social and Political Philosophy*, October, 1–22. https://doi.org/10.1080/13698230.2022.213775

Fleurbaey, Marc. 2008. *Fairness, Responsibility, and Welfare.* Oxford: Oxford University Press.

—-. 2018. "Welfarism, Libertarianism, and Fairness in the Economic Approach." In *Taxation: Philosophical Perspectives*, edited by Martin O'Neill and Shepley Orr, 37-59. Oxford: Oxford University Press.

Gheaus, Anca, and Lisa Herzog. 2016. "The Goods of Work (Other Than Money!)." *Journal of Social Philosophy* 47 (1): 70–89.

Graeber, David. 2018. *Bullshit Jobs*. Simon & Schuster New York, NY.

Kymlicka, Will. 1989. "Liberal Individualism and Liberal Neutrality." *Ethics* 99 (4): 883–905.

Lockwood, Benjamin B., Charles G. Nathanson, and E. Glen Weyl. 2017. "Taxation and the Allocation of Talent." *Journal of Political Economy* 125 (5): 1635–82. https://doi.org/10.1086/693393.

—-. 2017. "What If Socially Useful Jobs Were Taxed Less Than Other Jobs?" *The Harvard Business Review.* Retrieved on 11 November 2024, from .

Mirrlees, James. 1971. "An Exploration in the Theory of Optimum Income Taxation." *The Review of Economic Studies 38*(2): 175-208.

Posner, Eric, and E. Glen Weyl. 2013. "Benefit-Cost Analysis for Financial Regulation." *American Economic Review* 103 (3): 393–97.

Schneider, Florian, Fanny Brun, and Roberto A. Weber. 2020. "Sorting and Wage Premiums in Immoral Work." *University of Zurich, Department of Economics, Working Paper*, no. 353.

Smuts, Aaron. 2013. "The Good Cause Account of the Meaning of Life." *The Southern Journal of Philosophy*51 (4): 536–62. https://doi.org/10.1111/sjp.12044.

Tyssedal, Jens Jørund. 2023. "Good Work: The Importance of Caring about Making a Social Contribution." *Politics, Philosophy & Economics* 22 (2): 177–96. https://doi.org/10.1177/1470594X221148204.

Wolf, Susan. 2010. *Meaning in Life and Why It Matters*. Princeton University Press.

Statistics

- https://www.aacnnursing.org/news-data/fact-sheets/nursing-shortage.
- https://www.devlinpeck.com/content/teacher-shortage-statistics.

Keywords: taxation, justice, social value, liberal neutrality

Underdetermination in experimental game theory: the anti-holistic response

Maria Caamano*†[1]

[1]Universidad de Valladolid [Valladolid] – Spain

Abstract

Underdetermination is a central issue in the philosophy of science, whose discussion has developed from the early 20th century (Duhem 1906/1954, Quine 1953). The so called "Duhem-Quine problem" has been used as an umbrella term to refer to several problematic features emerging from the lack of a biunivocal correspondence between theory and evidence. Despite the centrality of the problem, underdetermination has typically not been directly addressed in debates about social science experimentation. A remarkable exception to this comes from some methodological analyses of experimental economics (see Hands 2001, Søberg 2005, Guala 2005, 2012, Bardsley et al. 2010, Jones 2008, 2012). A common denominator of these approaches was the proposal of anti-holistic strategies to get around the Duhem-Quine issue. According to the holistic version of this problem, whether an experimental result confirms or refutes the main hypothesis under test depends on which auxiliary hypotheses are associated with the main hypothesis. This implies that confirmation and refutation can only be established once the auxiliary hypotheses have been discarded as sources of error. The revision of auxiliary hypotheses hence emerges as a major task in social science experimentation. Anti-holistic strategies, consisting in examining the epistemic credentials of auxiliary assumptions, are not only relevant to validate the application of a given hypothesis, but also to set some constraints on the class of alternative hypotheses that could be applicable. The proposal focuses on the anti-holistic response to underdetermination, paying particular attention to this problem in experimental game theory.

The suggested analysis assumes that, when considering auxiliary hypotheses, two subsets are particularly relevant: a) those concerning what is assumed to be constant in an experimental design and b) those concerning the workings of the instrument. The relevance of a) stems from the fact that, in order to isolate the effect of the main variable of interest in an experiment, it is necessary to be able to determine which conditions remain stable (Guala 2012, Esterling, Brady and Schwitzgebel (EBS) 2021, 23). Knowledge of these elements provides evidence that there are the right conditions for the treatment to manifest its causal power, and that, keeping those conditions, but removing the main variable, the same results will not be obtained. The importance of understanding the workings of the instrument (i.e. b) above) should also be emphasized, since underdetermination can also be connected with the problem of instrumental opacity (Borsboom et al. 2009). As long as the measurement 'instruments' remain opaque, the same empirical data can be accommodated into many alternative interpretative frameworks. This has a straightforward methodological implication: scientists must have enough knowledge of the causal chain or network involved in the working of the measurement procedure.

*Speaker

† Corresponding author: mcaamano2@gmail.com

Where does experimental game theory stand in relation to a) and b)? Are the auxiliary hypotheses involved in testing game theoretic models being tested in turn? Do they relate to background conditions or to the working of the instrument? From the methodology of experimental economics, several authors have developed promising anti-holistic strategies in connection with the Duhem-Quine problem. As suggested by Søberg (2005), using a series of experiments, each testing one particular auxiliary assumption in conjunction with the main hypothesis, can help to rule out auxiliary assumptions like, for instance, those regarding the influence of factors like anonymity or the amount of the payoffs. Another important proposal is Guala's (2005) objective eliminative induction, where probabilities are attached to testing procedures, depending on how close the experiment is to an "ideal controlled experiment". Beyond general frameworks for anti-holistic strategies, a significant progress has been made in the cases of ultimatum game and public goods experiments (Jones 2012, 81-4, Andreoni & Blanchard. 2006, Andreoni 1990, Levine 1998, Santos 2009).

In view of the problems described above, experimental game theory has proposed two different strands of models to explain observed behavior: (1) models of social preferences, (2) models of bounded rationality (Hirokazu 2019, 28-33). In order to evaluate these options, it is important to have a clear idea of the progress made on the anti-holistic front. The present discussion explores how anti-holistic strategies have emerged in this field, highlighting both the existing successes and limitations.

References

Andreoni, J. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm- glow Giving." Economic Journal 100: 464–477.

Andreoni J. & E. Blanchard. 2006. "Testing subgame perfection in ultimatum games". Experimental Economics, 9: 307-321.

Bardsley N., R. Cubitt, G. Loomes, P. Moffatt, C. Starmer & R. Sugden. 2010. Experimental Economics: Rethinking the Rules. Princeton NJ: Princeton University Press.

Borsboom, D., A. O. J. Cramer, R. A. Kievit, A. Z. Scholten, and S. Franic. (2009). "The End of Construct Validity." In The Concept of Validity: Revisions, New Directions, and Applications, edited by R. W. Lissitz, 135–170. Charlotte, NC: Information Age Publishing.

Duhem, P. 1906/1954. The Aim and Structure of Physical Theory. Princeton: Prince- ton University Press.

Esterling, K. M., Brady, D., & Schwitzgebel, E. 2021. "The Necessity of Construct and External Validity for Generalized Causal Claims", https://www.datascienceassn.org/sites/default/files/The%20Necessity

Guala, F. 2005. The Methodology of Experimental Economics. Cambridge: Cambridge University Press.

Guala, F. 2012. "Experimentation in economics". In Mäki, U., editor, Philosophy of Eco- nomics, Handbook of the Philosophy of Science, North-Holland, Amsterdam: 597–640.

Hands, D. Wade 2001, Reflection Without Rules: Economic Methodology and Contemporary Science Theory. Cambridge: Cambridge University Press.

Hirokazu, T. 2019. "Laboratory Experiments in Game Theory". In: Kawagoe, T., Takizawa, H. (eds.) Diversity of Experimental Methods in Economics. Springer, Singapore. 23-37. https://doi.org/10.1007/978-981-13-6065-7 2

Jones, M. K. 2008. "On the Autonomy of Experiments in Economics". Journal of Eco- nomic Methodology, vol.15: 391-407.

Jones, M. K. 2012. " What to Do with a Problem Like Duhem-Quine? ", Revue de philosophie èconomique 2012/1 (Vol. 13): 79-102. DOI 10.3917/rpec.131.0079

Levine, D. K. 1998. "Modeling Altruism and Spitefulness in Experiments", Review of Eco- nomic Dynamics 1: 593-622.

Quine, W. O. 1953. ”Two Dogmas of Empiricism”. In From A Logical Point of View. Cambridge, Mass.: Harvard University Press: 20-46.

Søberg, M. 2005. ”The Duhem-Quine thesis and experimental economics: a reinterpretation”. Journal of Economic Methodology, vol.12: 581-597.

Keywords: Underdetermination, auxiliary assumptions, experimental game theory, ultimatum game, public goods experiments

What does the Trust Game say about Trust ?

Nicolas Camilotto*[1]
[1]Groupe de Recherche en Droit, Economie et Gestion *Centre National de la Recherche Scientifique, Université Côte d'Azur* – France

Abstract

Introduction and Motivation

This paper provides a historical and methodological analysis of the Trust Game, a widely used experimental design to study trust and social preferences. Although the influence of the Trust Game is considerable, it has not been subjected to the same level of critical scrutiny as other prominent experimental games. This paper aims to address this gap by tracing the game's historical trajectory, mapping its diverse applications across economic subfields, and critically evaluating the methodological challenges associated with its interpretation and the validity of its findings.

Historical Development of the Trust Game

The paper begins by situating the emergence of trust as a research topic within economics. Initially (late 1980s, early 1990s), economists primarily attempted to frame trust as a matter of calculated self-interest, a perspective known as "calculative trust" or "encapsulated trust." This view, rooted in incomplete contract theory, considered trust largely irrelevant when effective contracts or reputation mechanisms were present. David Kreps's early "Trust Game" (1990) reflected this, casting trust as a strategic decision within a one-sided prisoner's dilemma.

The critical turning point came with Berg, Dickhaut, and McCabe's (1995) Investment Game, now universally recognized as the standard Trust Game. This game fundamentally challenged the calculative trust framework. Experimental results from Berg et al. (1995) and subsequent replications consistently showed levels of trust and trustworthiness that contradicted the predictions of standard game theory based solely on self-interest. This finding spurred the widespread use of the Trust Game to study social preferences, marking a significant shift away from the earlier, purely calculative conception of trust.

Mapping the Uses of the Trust Game: A Bibliometric Analysis

To understand the Trust Game's varied applications, the paper uses a quantitative bibliometric approach. A corpus of 1098 papers (2008-2022) from the Web of Science Core Collection, mentioning "trust game"

*Speaker

or ”investment game” in the title, abstract, or key- words, is assembled. A bibliographic coupling network analysis is performed, linking papers based on shared references. This, combined with textual analysis of abstracts, reveals the structure of the research landscape and identifies key clusters.

Four main clusters emerge:

- **Psychology of Trust:** The largest, reflecting the Trust Game's significant impact in psychology, particularly neuropsychology. This cluster builds on Deutsch's (1958) earlier work, exploring the neural and cognitive foundations of trust.
- **Behavioral Economics:** This cluster focuses on building a theory of trust, incorporating psychological insights and challenging standard rational choice assumptions. It examines how experimental manipulations of the Trust Game affect behavior and investigates potential confounds (e.g., other-regarding preferences, risk aversion).
- **Experimental Economics:** Closely tied to development economics, this cluster uses the Trust Game to explore the link between trust, social capital, and economic outcomes. It often aims to inform policy, drawing on evidence-based policy principles.
- **Trust and Personality** This cluster, linked to psychology explore the use of the Trust game to analyse the relation between trust and personality.

The Problem of Validity: Contrasting Perspectives in Experimental and Behavioral Economics
The paper then examines the crucial issue of validity, focusing on external validity in the Trust Game. It highlights the *polysemy* of "external validity," often used to refer to both the *artificiality* of experimental results and their *generalizability*. Crucially, the paper contrasts how experimental and behavioral economics approach validity:

- **Experimental Economics:** Primarily concerned with *generalizability*. Researchers in this tradition often correlate Trust Game results with real-world behaviors (e.g., loan repayment) to assess the game's predictive power. However, this approach often lacks a strong theoretical framework to explain *why* these correlations exist. The focus is on the *social determinants* of trust.
- **Behavioral Economics:** Primarily concerned with *artificiality*. Researchers focus on how experimental manipulations within the Trust Game affect behavior drawing on psychological concepts. This approach seeks to develop a theory of trust, but often neglects the social context in which trust decisions are typically embedded. The focus is on the *psychological mechanisms* of trust.

Towards convergence? Challenges and opportunities
The paper concludes by advocating for a *rapprochement* – a closer integration – between experimental and behavioral economics in the study of trust. It suggests that combining the strengths of both approaches could lead to a more comprehensive understanding.

- Experimental economics could benefit from incorporating more robust theoretical frameworks (from behavioral economics) to guide its research and enhance the generalizability of its findings.
- Behavioral economics could benefit from greater attention to the social context and the external validity (in the sense of generalizability) of its findings.

However, the paper acknowledges the significant *challenges* inherent in this integration. Uniting the study of social determinants (experimental economics) and psychological mechanisms (behavioral economics) within a single experimental framework could introduce confounding effects, making it difficult to disentangle the various influences on trust behavior.

Keywords: Trust, Trust, Game, Experimental Economics, Behavioral Economics, Validity

Conventions and the performativity of economics

Jonas Campos*†[1]

[1]Universidade Estadual de Campinas = University of Campinas – Brazil

Abstract

Economics sometimes influences the economy-a phenomenon often referred to in recent decades as the performativity of economics. Since the early debates on the performativity of economics, scholars have acknowledged the importance of conventions (Callon, 2007; Felin & Foss, 2009; MacKenzie, 2006, 2010). However, they did not reference any concept of a convention. More recently, some authors have introduced David Lewis' (1969) concept of a convention to theorize the performativity of economics (Brisset, 2016, 2019; Guala, 2016). Using this concept, they have identified mechanisms through which economics can be performative and some limits it must face, thereby illuminating how performativity works.
In this paper, I argue that Lewisian conventions are too restrictive to theorize the performativity of economics in three main respects: (i) why individuals follow conventions; (ii) how conventions influence individuals; and (iii) how conventions relate to self-fulfilling prophecies. To address these limitations of Lewisian conventions, I propose an alternative conventionalist theory of the performativity of economics based on a different concept and theory of conventions. Moreover, I argue that even this alternative conventionalist theory cannot account for all cases of the performativity of economics, and suggest complementing it with the concept and theory of social norms.

First, in Lewis' concept, conventions are always followed due to a combination of common knowledge and individuals' preference for coordination. This combination has been argued to be absent in many conventions. Some have argued that common knowledge of a convention is often unrealistic, yet conventions persist (Binmore, 2008; Young, 1996). Others have argued that in some conventions, individuals do not have a preference for coordination in Lewis' sense (Sugden, 2005; Vanderschraaf, 1998). They prefer to conform when others conform, but they also prefer that others deviate.

Second, Lewisian conventions influence individuals in important ways. They provide incentives for individuals to conform and expectations about the likely behavior of others, allowing them to coordinate. However, it presupposes that individuals have a preference for coordination and share sufficiently similar background information and patterns of inference. It cannot change how individuals interpret their situation or what motivates them. Other- wise, it would cease to be a convention in Lewis' definition. In contrast, some authors argue that conventions sometimes influence how individuals interpret their situation and what motivates them, partly constituting their cognition and motivation.

Third, Lewisian conventions always involve self-fulfilling prophecies. When the convention is common knowledge, each individual expects others to conform, and these expectations fulfill themselves precisely because the convention is established. In addition, the self-fulfilling

*Speaker
†Corresponding author: jonas.campos103@gmail.com

character of these expectations (combined with a preference for coordination) motivates each individual to conform to the convention, in what Brisset (2019, p. 139) calls a ”mechanism of self-fulfilling prophecies.” Although it has received less attention from convention scholars, the relationship between conventions and self-fulfilling prophecies is important to a theory of the performativity of economics. In contrast to the theory based on Lewisian conventions, not all types of performativity involve processes of self-fulfillment. This is true only for what MacKenzie (2006) calls Barnesian performativity. Even Barnesian performativity may involve only self-fulfilling theories or models, rather than prophecies. Furthermore, even if Barnesian performativity involves self-fulfilling prophecies, the mechanism of self-fulfilling prophecies need not operate.

To address these limitations, I propose an alternative conventionalist theory of the performativity of economics. This theory is built upon the concept and the theory of conventions as (systems of) social rules with two defining properties: conformity with conformity; and arbitrariness (Dequech, 2009, 2013, 2017, 2024). This concept has an important mental dimension, as conventions can be or include (systems of) rules of thought. This dimension includes expectations as well as mental models, through which individuals select, organize, and interpret information. The theory of conventions built upon this concept explains how conformity with conformity happens and what causes arbitrariness. It is also closely related to a theory of how conventions influence economic agents.

This alternative concept and theory lead to a significantly broader theory of the performativity of economics. Economics has (systems of) social rules of thought that can become or influence conventions of thought or of behavior in the economy, and can thereby influence how economic agents think or behave. When they do, economics is performative through conventions. In the theory of the performativity of economics so conceived, different mechanisms can lead individuals to conform with conventions, most of which do not involve a preference for coordination or common knowledge-though some are consistent with one or both factors. This theory also admits various types of influence economics can have on individuals through conventions, including constitutive influences on cognition and motivation. Finally, it allows for a broader relationship between self-fulfilling prophecies and the performativity of economics. The latter need not involve expectations, even if it involves self-fulfilling theories or models. If it does involve expectations that self-fulfill, individuals need not be aware of their self-fulfilling character and this character need not lead any of them to conform-though this may sometimes happen.
Despite its broader scope, this alternative conventionalist theory does not fully account for how economics can be performative. Conventions cannot explain why some economic agents are influenced by economics independently of the (expected) conformity of others. To explain this kind of situation, I suggest we use the concept and theory of social norms as (systems of) social rules of thought or behavior with three defining characteristics: they prescribe what individuals *ought* to do, according to a set of values; they imply the possibility of social sanctions, and some individuals may conform because of this possibility; and at least some individuals internalize the norm and conform independently of sanctions (Dequech, 2009). The internalization of a norm can be influenced by the (expected) conformity of others, but it can also happen regardless of whether others (are expected to) conform. Social sanctions also motivate conformity independently of the (expected) conformity of others. Thus, the concept and theory of social norms, including their influence on agents, can complement a conventionalist theory of the performativity of economics.

Keywords: Performativity of Economics, Conventions, Self Fulfilling Prophecies, Social Norms

Why Bloomington is not in Virginia: Contrasting the Social Ontologies of the Ostroms and Buchanan

Massimo Cervesato*[1]

[1]Université Paris 1 Panthéon-Sorbonne - École d' èconomie de la Sorbonne – Universit´e Paris 1 Panthéon-Sorbonne, Ecole d'Economie de la Sorbonne – France

Abstract

Over the past decade, numerous commentators have revisited the contributions of the Ostroms to the development of public choice theory, particularly to explore the affinities be- tween the Bloomington school and the Virginia school. This renewed interest has generated a wealth of comparative studies, typically centred on the works of the three principal figures associated with these traditions: James M. Buchanan on one side, and Vincent and Elinor Ostrom on the other (Aligica and Boettke, 2009; Aligica, 2018; Herzberg, 2018; Duhnea and Martin, 2021; Johnson, 2023; Lewis and Meadowcroft, 2024). Lewis and Meadowcroft (2024, p. 365) are the first to challenge this interpretative rapprochement, demonstrating how "(Vincent) Ostrom became increasingly critical of Buchanan's continued reliance on rational choice theory to analyse constitutional decision-making, an issue about which the existing literature has said little." While their analysis of this divergence is insightful, their primary objective remains to articulate how the two bodies of work can nonetheless be interpreted as complementary, as they conclude: "In focusing on the pre-requisites for constitutional choice, Ostrom complements Buchanan's account of moral order which, as noted above, was almost certainly under-developed in its analysis of the capabilities required for people to develop an effective constitutional attitude (...). In this respect, Buchanan and Ostrom turned out to have complementary talents, interests, and emphases" (ibid., p. 383). In other words, despite acknowledging a fundamental point of divergence, Lewis and Meadowcroft ultimately align with the broader literature that seeks to underscore convergences rather than oppositions between the two schools.

Nevertheless, Ostrom and Buchanan use neither the same ontological framework nor the same research methods, and therefore do not arrive at exactly the same results on collective actions. Thus, while much of the existing literature emphasises the complementarities between their works, we argue that their divergences signal a deeper opposition in their understanding and explanation of collective action phenomena. To substantiate this claim, we build upon the comparative analysis of the Ostrom's and Buchanan's social ontologies initiated by Lewis and Meadowcroft (2024), examining how their divergent conceptions of the social world give rise to differing methodological practices, objectives, and ultimately opposed results in the study of collective action.

*Speaker

The growing interest in the ontological underpinnings of economic theories is evidenced by a burgeoning literature on the subject (T. Lawson, 2016, 2019; Martins, 2020; Lewis, 2021; Slade-Caffarel, 2024; Lewis and Runde, 2024). This approach is particularly relevant to the works of the Ostroms and Buchanan, as all three thinkers accord significant importance to the ontological foundations of their analyses (Wagner, 2018; Lewis, 2021; Lemke and Tarko, 2021). By confronting their respective conceptions of social order, we can draw out the implications of these ontological differences for their methods, objectives, and results.

The article is structured as follows. Section 1 outlines the common ground shared by Buchanan and the Ostroms. Section 2 demonstrates that they adopt distinct approaches to the problem of consensus in collective action. While Buchanan (Buchanan and Tullock, 1962; Buchanan, 1975, 2003) focuses on unanimity among individuals in exchange situations, the Ostroms emphasize reciprocal relationships among agents who share common knowledge and form "communities of shared understanding" (V. Ostrom, 1993, p. 170; V. Ostrom, 1980, 1997; Kiser and E. Ostrom, 1982; E. Ostrom, 1986, 1998). Section 3 explores how these ontological differences yield divergent epistemological frameworks, with Buchanan rooted in rationalist epistemology and the Ostroms drawing on pragmatist traditions (Sproule-Jones, 1982). Section 4 discusses the normative implications of these epistemological divergences: whereas Buchanan primarily investigates the relationship between *government* and the individuals they serve, the Ostroms focus on *governance* and the active role of individuals in collective decision-making (Johnson, 2022). Finally, Section 5 demonstrates how these onto- logical, epistemological, and normative differences lead to contrasting conclusions regarding individual rationality and the functioning of public economies.

Bibliography

Aligica, P.D. (2018) 'Artefactual and Artisanship: James M. Buchanan and Vincent Ostrom at the Core and Beyond the Boundaries of Public Choice', in R.E. Wagner (ed.) *James M. Buchanan: A Theorist of Political Economy and Social Philosophy*. Palgrave MacMillan, pp. 1105–1129.

Aligica, P.D. and Boettke, P.J. (2009) *Challenging Institutional Analysis and Development: The Bloomington School*. London: Routledge.

Buchanan, J.M. (1975) *The Limits of Liberty: Between Anarchy and Leviathan*. University of Chicago Press.

Buchanan, J.M. (2003) 'Public Choice: The Origins and Development of a Research Pro- gram'. Center for Study of Public Choice, George Mason University.

Buchanan, J.M. and Tullock, G. (1962) *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Duhnea, E. and Martin, A. (2021) 'Public choice theory: reuniting Virginia and Blooming- ton', in J. Lemke and V. Tarko (eds) *Elinor Ostrom and the Bloomington School: Building a New Approach to Policy and the Social Sciences*. Agenda Publishing.

Herzberg, R. (2018) 'Votes, Vetoes, Voice, and Exit: Constitutional Protections in the Work of James M. Buchanan and Vincent Ostrom', in R.E. Wagner (ed.) *James M. Buchanan: A Theorist of Political Economy and Social Philosophy*. Palgrave MacMillan.

Johnson, M. (2022) 'Two "Two Ostrom" Problems', *History of Political Economy*, 54(S1), pp. 69–96.

Johnson, M. (2023) 'On classical liberalism and democratic optimism: James M. Buchanan and Elinor Ostrom', *Journal of Public Finance and Public Choice*, 38(1), pp. 2–19.

Lawson, T. (2016) 'Comparing Conceptions of Social Ontology: Emergent Social Entities and/or Institutional Facts?', *Journal for the Theory of Social Behaviour*, 46(4), pp. 359–399.

Lawson, T. (2019) *The Nature of Social Reality: Issues in Social Ontology*. London and New York: Routledge.

Lemke, J. and Tarko, V. (2021) *Elinor Ostrom and the Bloomington School: Building a New Approach to Policy and the Social Sciences*. Agenda Publishing.

Lewis, P. (2021) 'Elinor's Ostrom's "Realist Orientation": An Investigation of the Ontological Commitments of her Analysis of the Possibility of Self-Governance', *Journal of Economic Behavior & Organization*, 189, pp. 623–636.

Lewis, P. and Meadowcroft, J. (2024) 'Constitutional Artisans: James Buchanan and Vin- cent Ostrom on Artifactual Man, the Constitutional Attitude, and the Political Economy of Constitutional Design', *Constitutional Political Economy*, 35(3), pp. 363–387.

Lewis, P. and Runde, J. (2024) 'Bloomington and Cambridge compared: varieties of ontological thinking, social positioning, and the self-governance of common-pool resources', *Cambridge Journal of Economics*.

Martins, N.O. (2020) 'Reconsidering the notions of process, order and stability in Veblen', *Cambridge Journal of Economics*, 44(5), pp. 1115–1135.

Ostrom, E. (1998) 'A Behavioral Approach to the Rational Choice Theory of Collective Action', *The American Political Science Review*, 92(1), pp. 1–22.

Ostrom, E. (1986) 'An Agenda for the Study of Institutions', *Public Choice*, 48(1), pp. 3–25.

Ostrom, V. (1980) 'Artisanship and Artifact', *Public Administration Review*, 40(4), pp. 309– 317.

Ostrom, V. (1993) 'Epistemic Choice and Public Choice', *Public Choice*, 77(1), pp. 163– 176.

Ostrom, V. (1997) *The Meaning of Democracy and the Vulnerability of Democracies: A Response to Tocqueville s Challenge*. University of Michigan Press.

Slade-Caffarel, Y. (2024) *Cambridge Social Ontology: An Introduction to Social Position ing Theory*. Routledge.

Sproule-Jones, M. (1982) 'Public Choice Theory and Natural Resources: Methodological Explication and Critique', *The American Political Science Review*, 76(4), pp. 790–804.

Wagner, R.E. (ed.) (2018) *James M. Buchanan: A Theorist of Political Economy and Social Philosophy*. Palgrave MacMillan.

Keywords: Ostrom (Vincent), Ostrom (Elinor), Buchanan (James M.), Social Ontology, Public Choice, Collective Actions

What is a structural model?

Christopher Clarke*[1]

[1]Erasmus University Rotterdam – Netherlands

Abstract

The traditional approach to causality in econometrics-championed by the tradition from Haavelmo and Koopmans through to Heckman and Wooldridge-is to understand causality in terms of structural parameters. This approach is known as the structural modelling approach to causality. Unfortunately, the structural modelling approach has never successfully articulated what it is for a parameter to be a genuinely structural parameter. This leaves the philosophical foundations of the structural modelling approach insecure.

These unsecure foundations have caused economists to turn to alternative ways of under- standing causality, in particular the Rubin Holland ”potential outcomes” framework (Ruben and Imbens 2015).

Following Herbert Simon, Kevin Hoover (2001, 2011, 2012, 2013) offers a theory of structural parameters that I call the ”quasi-interventionist theory”. As I interpret it, this theory is defined by two core claims: first, that structural parameters are counterfactually modular, meaning that each parameter remains unchanged under counterfactual variation of the others; and second, that structural parameters are directly controlled by decision makers.

This paper argues against Hoover’s quasi-interventionist theory of structural parameters. A key difficulty with Hoover’s second claim-that structural parameters are directly controlled- is the ambiguity surrounding the definition of direct control. One possible interpretation, inspired by Woodward’s interventionist framework, suggests that a parameter is directly controlled if it counterfactually depends on a policymaker’s actions while holding other parameters fixed. However, this definition leads to inconsistencies, as it fails to account for the necessary interdependencies among parameters in a system. A second possible interpretation is instead to redefine direct control in terms of causal order among parameters themselves. But this interpretation merely shift the explanatory burden: causal order among variables is defined in terms of causal order among parameters, which is a concept that remains itself ill-defined.

Given these challenges, this paper argues that Hoover’s theory should abandon the notion of direct control and instead focus on counterfactual modularity alone as the defining feature of structural parameters. However, even this revised approach faces a ”multiple parameterization problem”. As illustrated through coordinate system transformations, it is possible for different sets of parameters to satisfy counterfactual modularity simultaneously. Since Hoover’s theory does not provide a principled way to select among these competing parameterizations, it remains incomplete.

*Speaker

This analysis does not constitute a decisive refutation of the quasi-interventionist theory but highlights the need for alternative approaches that either refine Hoover's framework or replace it with a more satisfying account of structural parameters.

I conclude by sketching my own "extrapolationist" theory of structural parameters. The extrapolation theory offers an inferentialist rather than a reductionist theory of structural parameters (Williamson 2005 2006; Reiss 2012). This inferentialist theory says what it is for a reasoner to *believe* that a parameter is structural. Thus it provides a believe condition, but not a truth condition, for claims about structural parameters. Roughly put, it says: be- lieving that a set of parameters is structural means treating these parameters as inferentially independent when extrapolating the chances from one generic case to another.

References

Hoover, Kevin D. 2001. *Causality in Macroeconomics*. Cambridge University Press.

—. 2003. "Nonstationary Time Series, Cointegration, and the Principle of the Common Cause." *The British Journal for the Philosophy of Science* 54 (4): 527–51.

—. 2011. "Counterfactuals and Causal Structure." In *Causality in the Sciences*, edited by Phyllis Mckay Illari, Federica Russo, and Jon Williamson. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199574131.001.0001.

—. 2012. "Causal Structure and Hierarchies of Models." *Studies in History and Philosophy of Science Part C:Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (4): 778–86. https://doi.org/10.1016/j.shpsc.2012.05.007.

—. 2013. "Identity, Structure, and Causal Representation in Scientific Models." In *Mechanism and Causality in Biology and Economics*, edited by Hsiangke Chao, Szu-ting Chen, and Roberta L. Millstein, 35–57. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978- 94-007-2454-93

Imbens, Guido W., and Donald Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*.

Reiss, Julian.2012. "Causation in the Sciences: An Inferentialist Account." *Studies in His- tory and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (4): 769–77. https://doi.org/10.1016/j.shpsc.2012.05.005.

Williamson, Jon. 2005. *Bayesian Nets and Causality: Philosophical and Computational Foundations*. New York: Oxford University Press.

—. 2006. "Causal Pluralism Versus Epistemic Causality." *Philosophica* 77 (1). https://doi.org/10.21825/philosophi

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wooldridge, Jeffrey M. 2016. *Introductory Econometrics: A Modern Approach*. Six. Boston MA: Cengage Learning.

Keywords: causation, structural models

Qualitative Economics: Integrating Qualitative Research Methods into Economic Analysis

Leonardo Conte*[1,2]

[1]Université de Fribourg, D´epartement d'Economie Politique – Switzerland
[2]Franklin University Switzerland, Taylor Institute – Switzerland

Abstract

Economics is traditionally a *quantitative* science (Redman, 1997); a discipline that has increasingly adopted mathematical language (Weintraub, 2002) and statistical analysis (Porter, 1986) throughout its development (Morgan, 1990). As a result, today's economists exclusively rely upon quantitative research methods including econometrics and experimental work, neglecting any type of *qualitative* evidence, data, and research methods such as inter- views, ethnographies, and case studies (Basole & Ramnarain, 2016).
However, most social sciences and adjacent disciplines extensively adopt qualitative method- ologies (Berg & Lune, 2012) when tackling economic phenomena, issues, and topics; either in the scientific inquiry of consumption and production or for policy design and business practice purposes. It is the case of economic sociology (Slater, 2005; Warde, 2015), economic anthropology (Carrier, 2022; Hann, 2018), and applied fields deriving from economics such as consumer and marketing research (Belk et al., 2012) and organization studies (Mir & Jain, 2018), for instance.

Scholars already advocated for the use of qualitative methods in economics by listing their potential benefits (Bergmann, 2007; Bewley, 2002; Helper, 2000; Piore, 2006; Starr, 2014) or explaining the reasons of other social sciences being influenced by economic methods (Four- cade, 2018; Hirsch et al., 1987). Yet, these represent rare examples of economic methodology literature tackling the topic. Most scholars in the philosophy of social sciences did not account for the importance of underlining the methodological gap existing in modern economics.

Building upon this, I ask (a) *why* economists do not rely on qualitative inquiry, (b) *how* they could implement qualitative research, and (c) in *what* subject domains. In doing so, this work (1) unveils the potential contribution of qualitative methods to both economic theory and policy, (2) highlights the role of *sociocultural* factors over *behavioural* elements in economic analysis, and (3) suggests the need for an ontological, epistemological, and axiological shift towards what I called 'qualitative economics'.

As answers to the first question, preliminary results show fundamental ontological, epistemological, and methodological differences between economics and other social sciences. They typically reflect into imitating the natural sciences' practices (Mirowski, 1989) which, in turn, led to methodological monism (de Muijnck & Tieleman, 2022; Fischer et al., 2017) and a preference for *predictive* rather than *explanatory* power (see Akerlof, 2020). Further- more, institutional and cultural dynamics make economists adopt methods in line with policy

*Speaker

evaluation (Mitchell, 1998), their public intellectuals' role (Caballero, 2010; Earle et al., 2016; Maesse, 2015), and job market's power dynamics (Card & DellaVigna, 2013; Rossier et al., 2017).

As answers to the second question, it emerges the need to shift philosophical assumptions (Newman & Benz, 1998); that reflects into ontological (Dow, 2009), epistemological (Downward et al., 2002; Lee, 2016), and methodological changes (Chamlee-Wright, 2010b; Pickbourn & Ramnarain, 2016). In addition, the adoption of qualitative methods shall be reinforced by shifting economists' attention from measuring 'how much' to describing and understanding in-depth the 'how' (Miles et al., 2014) and the 'why' of economic events (Helper, 2000); from testing causal effects to generating theories (Finch, 2002); from hypothetico-deductive model building to inductive and descriptive-exploratory reasoning (Jemna, 2016).

As answers to the third question, few examples of qualitative economic research are found in policy evaluation (London et al., 2007), wellbeing (Hiswäls et al., 2017), inequality (Schor, 2017), finance (Ho, 2009), among others. Similarly, studies that embed the principles of qualitative inquiry include applications in health economics (Coast, 1999; Lincoln, 1992) and organization studies (Gioia et al., 2013; Shah & Corley, 2006), for instance. Moreover, promising efforts are traced in the latest trends of behavioural (Chater & Loewenstein, 2023; Schimmelpfennig & Muthukrishna, 2023), feminist (Berik, 1997; O'Hara, 1999; Schwartz, 1994), and Austrian economics (Boettke et al., 2004; Chamlee-Wright, 2010a).

Building upon these studies, and in line with the so-called 'empirical turn' (Angrist et al., 2017; see also Focardi, 2015), this paper proposes frameworks for embedding qualitative research methods in economics. In so doing, it overcomes the false belief – spread among economists – that qualitative data is merely 'anecdotal' (Piore, 2006) and therefore 'inferior' (Sandelowski, 2000) to other forms of 'hard' evidence (Akerlof, 2020). This work contributes to the field of economic methodology by revitalising the debate around qualitative research in economics, revaluating its underestimated role for the analysis of economic objects, and examining its possible uses for theoretical and empirical studies in economics.

Keywords: economic analysis, economic methodology, qualitative economics, qualitative research methods, social sciences

Collective Action Problems, Self-defeating Behaviors, and their Alleged Equivalence

Annalisa Costella*[1]

[1]Vrije University, Amsterdam – Netherlands

Abstract

Tragic self-defeating conative attitudes that are of philosophical interest are characterized by the following structure. A suboptimal outcome is brought about by many actions, each of which is 'optimal' for the 'self' performing it. Yet, there is an alternative better outcome that is i) available for all and ii) attainable if everybody performed an action that is suboptimal for the 'self' performing it. Self-defeat occurs both at the interpersonal level and at the intrapersonal one. At the interpersonal level, it happens when a suboptimal collective outcome is brought about by the actions of many individuals, each performing an action that is optimal for them. At the intrapersonal level, it arises when a suboptimal outcome for a person is brought about by many actions, each of which is optimal for the person's time-slice performing it. Call self-defeat at the interpersonal level a *collective action problem* and at the intrapersonal level a *self-defeating behavior*.

The apparent similarity between collective action problems and self-defeating behaviors has not gone unnoticed. In fact, it has been harnessed to draw normative implications. Equivalence in the structure of the two problems, it has been implied, warrants equivalence in the solutions to both or, more generally, normative insights on one problem have been used to illuminate the normative analysis of the other. I shall name this the *equivalence thesis*. While the thesis is intuitive and apparently compelling, there has been little to no discussion on whether the two problems are in fact relevantly alike and, if so, what features of the problems warrant the equivalence. This is no small oversight. Understanding as and when the thesis holds is of key importance. Indeed, if there is one problem of 'self-defeat' that can be instantiated at the interpersonal level or at the intrapersonal one, this would warrant that collective action problems and self-defeating behavior can always be treated as relevantly alike. In turn, the pressure on delivering a normative analysis on each of the two problems would translate into the pressure of analyzing the more general problem of self-defeat. Instead, if this is not the case – and there are relevant features that make the problems sometimes (or always) disanalogous – it would become pointless, or even harmful, to transfer normative analyses or solutions from one problem to the other, since there is no similarity between the two problems that warrants such transfer. More fundamentally, the falsity of the equivalence thesis would cast doubt on the meaningfulness of referring to the underlying self-defeating structure of collective action problems or self-defeating behaviors as an 'object' worthy of philosophical investigation.

This paper is a first attempt to formulate and analyze the equivalence thesis. I will argue that, in the general and loose version that has driven the intuition behind transferring normative implications from one problem to the other, it ultimately fails to hold. However, I will continue, this does not imply that there is no

*Speaker

possible formulation of the equivalence thesis that works. I set out to propose such a formulation, arguing that the link between structural similarity and normative equivalence holds if two conditions, which I call *control* and *nonwaivability*, are satisfied. As it turns out, the importance of the equivalence thesis – and its role in supplying normative insights on collective action problems and self-defeating behaviors – is both over- and understated. It is overstated, insofar as it is not true that the loose equivalence between the structures of the two problems warrants equivalence in the solution. It is understated, insofar as, were a version of the equivalence thesis true, the benefits for our philosophical toolkit would be sizeable. Indeed, it has been argued repeatedly that many real-life pressing problems such as climate change, addictive behaviors, and general failures to bring about desirable societal outcomes are characterized by self-defeat (Andreou, 2006, 2008; Sinnott-Armstrong, 2005). But even if one where to disagree that "there is no more important problem in the social sciences and no more difficult"(Elster, 1985, p. 141), the problem of selfdefeat remains a conceptual puzzle that poses relevant challenges to our moral and rational theorizing.

This paper proceeds as follows. In section 2, I present the literature that has implicitly or explicitly relied on the equivalence thesis. I offer a formulation of the thesis that I take to be faithful to that literature. In section 3, I showcase how solutions to one problem fail to transfer to the other. In section 4, I discuss the implications of section 3 and offer a revised version of the equivalence thesis. I argue that this revised version can explain the failure of solutions transfer and, at the same time, render justice to the importance of the equivalence thesis. I then consider the broader implications of this restatement. In section 5, I conclude.

Selected bibliography

Andreou, C. (2006). Environmental damage and the puzzle of the self-torturer. *Philosophy & Public Affairs*, *34*(1), 95–108.

Andreou, C. (2008). Making a clean break: Addiction and Ulysses contracts. *Bioethics*, *22*(1), 25–31.

Elster, J. (1985). Rationality, Morality, and Collective Action. *Ethics*, *96*(1), 136–155. https://doi.org/10.1086/292725

Sinnott-Armstrong, W. (2005). It's Not My. In *Perspectives on climate change: Science, economics, politics, ethics* (pp. 285–307). Emerald Group Publishing Limited.

Keywords: Collective Action Problems, Self, defeating Behaviour, Rational decision, making, Moral Decision, making

PREFERENCES: MENTALISM, BEHAVIORISM AND FUNCTIONALISM

Mikael Cozic*[1]

[1] Université Jean Moulin Lyon 3 - Faculté de Philosophie – Université Jean Moulin - Lyon 3 – France

Abstract

Neo-classical economic theory uses models to predict and explain the behavior of eco- nomic agents. For almost a century, the *preferences* of economic agents have been at the core of these choice models. An agent is typically assigned preferences defined over a set of mutually exclusive options, and it is assumed that she chooses the option she prefers among those from which she can choose. The notion of preference is equally central in the various branches of decision theory, some of which, such as individual decision theory and game theory, formulate abstract versions of the choice models used in economic theory. The importance of the concept of preference raises the issue of knowing how it is (or should) be interpreted: what exactly does (or should) it mean to attribute to agent *i* a preference for option *x* over option *y* ? Two main families of interpretation are typically constated: the behavioral and the mental interpretations. According the behavioral interpretations, y means that the agent chooses *x* over *y*. For mental interpretations, y means that the agent is in some mental state, typically viewed as both comparative and conative. For in- stance, the state of desiring *x* more than *y*. An important part of the literature devoted to this issue consists in taking side with one of these interpretations. To be more precise, it should be added that the semantic theses considered may be descriptive (" how preferences are interpreted? ") or normative (" how preferences should be interpreted? ") in nature. The two types of theses are often combined: for example, Hausman (2012) claims that the interpretation of preferences as comparative evaluations is both legitimate and dominant in economics. In contrast, Clarke (2016) argue that the behavioral interpretation belongs to the" official doctrine of neo-classical economics ", and that there are good reasons for this (see also Thoma, 2021). Still others believe that the behavioral interpretation dominates economics, while the mental one is favored by philosophical decision theory (Bradley 2017, p. 45). The first aim of this paper is to examine the very distinction between the two classes of interpretations. I am not the first to do so. Guala (2019), in particular, scrutinizes it and comes to a twofold conclusion: that there is an interpretation of preferences that is *neither mental nor behavioral* and that this interpretation is the most faithful to scientific practice. I will endorse a significant part of the claims made by Guala. However, I will favor a view that, at least on the surface, is different from his: a view according to which preferences fit well the functionalist conception of mental states, which is so widespread in the philosophy of mind. Preferences are better characterized as *both mental and behavioral* (if one tolerates a liberal conception of the behavioral): they are individuated simultaneously by their causal relations to behaviors and to other mental states (mostly, beliefs).

*Speaker

References

Angner E. ”What Preferences Really Are” *Philosophy of Science*, *85*(4), 2018, p. 660– 681.

Bradley R., *Decision Theory with a Human Face*, Cambridge, Cambridge University Press, 2017.

Clarke C., ”Preferences and Positivist Methodology”, *Philosophy of Science*, 83, 2016, p. 192-212.

Dietrich F. & List C.,” Mentalism versus Behaviourism in Economics: A Philosophy-of- Science Perspective”, *Economics and Philosophy*, 32(2), 2016, p. 249-281.

Guala F.,” Preferences: Neither Behavioural nor Mental”, *Economics and Philosophy*, 35(3), p. 1-19, DOI:10.1017/S0266267118000512

Hausman D.M., *Preference, Value, Choice and Welfare*, Cambridge, Cambridge University Press, 2012.
Thoma J. ”In Defence of Revealed Preference Theory”, *Economics and Philosophy*, 37(2), 2021, p. 163-187, DOI: https://doi.org/10.1017/S0266267120000073

Keywords: preferences, mentalism, behaviorism, functionalism

Coping with uncertainty

Ricardo Crespo*[1] and Fernando Tohmè*†[2]

[1]Universidad Austral and CONICET (U Austral and CONICET) – Argentina
[2]Universidad Nacional del Sur [Argentina] – Argentina

Abstract

Uncertainty is inherent to human existence. It stems from individual uniqueness, future unpredictability, and human agency. These factors make economic events intricate and resistant to generalization through inductive reasoning. While statistical analysis can reveal population trends, it is inapplicable to singular occurrences.

The singularity of economic events arises, in particular, from the reflexive character of human affairs and the reaction to predictions of these events (see Grunberg and Modigliani, 1954). Economic theory has often neglected the importance of uncertainty, reducing it to probabilistically tractable risk. However, some authors, such as Frank Knight, John Maynard Keynes, and George Shackle, have argued that uncertainty is immeasurable. Uncertain events must be handled differently if they cannot be assigned numerical weights. Even so, decisions made in such settings could still be deemed rational. However, in a wider sense: ”Decision-making procedures may be considered rational even when it is not possible to determine numerically definite probabilities of future states of the world” (Runde and Mitsuhara, 2003 p. 3).

In this paper, we will introduce ways of dealing with uncertainty, discussing, in particular, the original suggestions of Knight, Keynes, and Shackle (KKS), and emphasizing how the assumptions that lead to the decisions made by the agents can be seen as obtained through the logical process of abduction. We think it is relevant to look for methods of managing uncertainty by analyzing the similarities with the abduction process implicit in the KKS proposals. As we will explain, abduction is not a way of decision-making, but a way of arriving at likely hypotheses that may provoke decisions and actions.

We start by presenting the ideas of Knight, Keynes, and Shackle in section 2. For Knight, ”the actual procedure of making decisions in practical life is a rather inscrutable or ‘intuitive’ formation of ‘estimates’, subject to a wide margin of error or uncertainty” (1921, p. 314). Knight quotes Alfred Marshall who considered that business managers make decisions based on ”trained instinct” rather than knowledge (1921, p. 211). For example, a manufacturer considering whether to increase the capacity of the firm will take into account measurable factors, but the final decision will be based on estimates of the outcomes (1921, p. 226).

For Keynes, in face of an uncertain situation, the only possibility is to combine different inputs in the most reasonable way: make a precise observation of the situation; compare it to similar situations faced previously; and imagine people’s reactions taking into account conventions, experience, and intuitions. This interpretation is consistent with the theory of knowledge developed in the Treatise on Probability and other works, in which intuition or

*Speaker
†Corresponding author: ftohme@gmail.com

direct knowledge plays a salient role. Furthermore, in his biographical essay on Ramsey, Keynes claims that: ”the basis of our degrees of belief – or the a priori probabilities as they used to be called – is part of our human outfit, perhaps given us merely by natural selection, analogous to our perceptions and our memories rather than to formal logic” (1973, pp. 300-301).

For Shackle, intuition, feelings, imagination, judgements, instinct, are the tools with which reality is imperfectly managed in this uncertain world.

In section 3 we discuss how the ideas of Knight, Keynes and Shackle can be reframed as involving instances of abduction and the related concept of inference to the best explanation (IBE).

In Section 4, we review other economic analyses of uncertainty. The post-World War II revolution in economics treated uncertain events as if they were simply risky. It was not until the emergence of behavioral economics that these alternative views of uncertainty came back to the forefront. While Behavioral Economics adopted criticisms of the representation of uncertainty in Mainstream Economics, it was rather late in advancing alternative formalism.

These formalisms can be better understood through the lens of abductive reasoning. This view posits that agents generate hypotheses about future outcomes and base their decisions on these hypotheses. This has two implications: it provides a comprehensive model of decision-making and it sheds light on why agents may deviate from the predictions of main- stream models, leading to perceived ”mistakes.” We discuss some recent advances along this line.

References

Grunberg, Emile and Franco Modigliani (1954). ”The Predictability of Social Events”, Journal of Political Economy LXII/6: 465-78.

Keynes, John Maynard (1973). The General Theory and After: Part II. Defence and Development, The Collected Writings of John Maynard Keynes (Vol. 14). London: Macmillan.

Knight, Frank H. (1921). Risk, Uncertainty and Profit. Mineola: Dover Edition, 2006.

Runde, Jochen and Sohei Mizuhara (2003). ”Introduction”. In: Jochen Runde and So- hei Mizuhara (eds.) The Philosophy of Keynes’s Economics. London: Routledge, pp. 1-16.

Keywords: Uncertainty, F. Knight, J.M. Keynes, G. Shackle, Abduction, New theories of uncertainty

Revising the Moral Arguments Black Reparations in the US in light of the Economic Arguments

Johne Davis*[1]
[1]University of Amsterdam – Netherlands

Abstract

There are currently two kinds of arguments for paying reparations to the descendants of enslaved individuals in the US, one economic regarding how it should be done and one moral regarding why it should be done. However, these arguments were developed independently of one another, and rely on different conceptual foundations. Successful public policy for paying reparations needs both kinds of arguments and needs them to be mutually supporting. This paper argues that in this case the burden of adjustment falls on the current moral arguments. Their reliance on Locke's natural rights thinking and neoclassical economics individualism makes them philosophically problematic, unrealistic, and ultimately unpersuasive. The eco- nomic arguments for reparations rely on Stratification economics, and represent individuals as members of social groups. They provide support for concepts of group rights and collective moral responsibility, and provide a basis for more realistic and persuasive moral arguments for reparations.

There are a number of methodological-philosophical issues in the paper. One is how does economic and social science analysis provide or fail to provide foundations for normative policy recommendation, and what adjustments in either might this entail? Much current methodological thinking about values in economics disparages the value neutrality idea. But this takes us little distance toward investigating how values in policy recommendation meant to produce action get their purchase and basis in economic and social science analysis. This paper examines one case of this on a difficult and socially sensitive matter: reparations for slavery.

A second issue concerns how individual-level and group-level ontological concepts operate in the leading plan for black reparations developed in Stratification economics. Both types of concepts are involved, but their boundaries, points of application, and weights in this plan need to be made clear to explain what is morally compelling about that plan. In general, the methodology of economics has been more concerned with epistemological issues. But these discussions presuppose ontological assumptions about who economics is about and who policies apply to.

A third issue is how Locke's state of nature foundation for social equality in his influential reparations arguments can be dismissed while still preserving social equality as a founda-tional ideal. If 'nature' does not/cannot provide grounds for moral arguments, what kinds of grounds are there for those arguments? In general, reparations arguments of all kinds, for example as associated with war and ethnic cleansing, advance justice claims framed as rectifying past moral wrongs and crimes. Yet is the concept of rectification meaningful without being accompanied by a concept of what rectification restores? Are the ideals of rectification and restoration both required in social justice thinking?

*Speaker

Economics methodology is especially concerned with what the ideal of objective science requires of economics. Normative ideas have entered into this in connection with the idea that concepts are often value-laden. Yet this conclusion is quite modest when it comes to dealing with the role of normative thinking in economics especially because economics has al- ways been an active policy science involved in laying economic foundations for how economies ought to be governed. This paper interrogates this state of affairs and in so doing comments on the general state of economic methodology. But more practically it also aims to make a contribution to a fundamental contemporary moral issue, one important not just because of the magnitude of the crime of African American slavery in the US but also important because comparable injustices are regularly occurring in the present time.

Keywords: reparations moral arguments, reparations economic arguments, Locke, neoclassical economics, Stratification economics, group rights, collective moral responsibility

JEL codes: A13, D63, J15, J88, Z13

Keywords: reparations moral arguments, reparations economic arguments, Locke, neoclassical economics, Stratification economics, group rights, collective moral responsibility

Co-Designing Nudges: Measuring the Impact of Expert and Citizen Collaboration

Samuel De La Cruz Solal*[1] and Guilhem Lecouteux†[1]

[1]Groupe de Recherche en Droit, Economie et Gestion – Centre National de la Recherche Scientifique, Universitè Côte d'Azur – France

Abstract

This study aims to examine the differences in recommendations generated through deliberation among 'experts' and 'citizens' when asked to create a nudge. According to Blomkamp (2018), co-design is a novel method for creatively engaging citizens and stakeholders to find solutions to complex problems. Co-design holds great promise for policy makers. The aim of this approach is to shift the process from a top-down approach to an evolutive bottom-up process, within which citizens' contributions can be supported by lab experiments organised by experts – such an approach is labelled by Richardson & John (2021) a *nudge-plus*. The inclusion of citizens in the process of construction of nudges may thus increase their ethical acceptability by reducing the paternalistic risk of letting experts decide what is best for others. For example, one proposed way is to incorporate initiatives for citizen dialogue and deliberation, known as 'thinks' (John et al. 2019). 'Thinks' are participatory initiatives whereby citizens and others discuss policy issues and policy proposals in an open, transparent and reflexive way, such as citizen assemblies discussing how to address climate change. This study will follow an experimental protocol with three distinct groups. The first group E (Experts) composed of participants with prior knowledge of nudges will engage in a design- thinking workshop to create nudges for students on another campus. The second group C (Citizens) composed of participants without prior knowledge of nudges will undergo the same workshop but design nudges for their own campus. The last Group EC (Expert + Citizens), in which citizens will collaborate with one expert (who previously participated in Group E) in a similar workshop. To the best of our knowledge, the existing literature has not explored this specific approach of integrating experts and citizens in a co-design process for developing nudges, underscoring its originality.

Our analysis will focus primarily on three aspects: (i) the possible differences between the recommendations of each group; (ii) the influence of the presence of one expert in the EC group (and whether such groups will exhibit mostly expert-driven or citizen-driven outcomes); (iii) the acceptability of each proposal, depending on the nature of the nudge, and on the nature of the group that composed the nudge. A complementary goal will be to measure the efficiency (or at least, the perceived efficiency *ex ante* from different groups) of the nudges proposed by the different groups. As pointed out by Daniel Priolo et al. (2023), most of the studies conducted so far to assess the acceptability of nudges nonetheless have suggested a relatively high acceptability by people, with some fluctuations depending on the details of the intervention. The typical design of these studies consists in presenting nudge scenarios to experimental participants (e.g. through web surveys) and asking them if they

*Speaker

†Corresponding author: guilhem.lecouteux@univ-cotedazur.fr

find the intervention acceptable or not. Participants thus make a hypothetical acceptability judgment in at least two senses. First, they judge the acceptability of the descriptions of nudges without experiencing the nudge. Second, even when they do experience the nudge insofar as subjects must make decisions (Bang et al., 2018), these decisions remain hypothetical in the sense that subjects do not experience the consequences of their choice. Furthermore, Weber (2015) suggests that the acceptability of a nudge may increase after its implementation, meaning that measures of acceptability should take this effect into account. Additionally, behavioural outcomes will be measured, such as compliance rates with the proposed behavioural interventions (e.g., increased recycling, reduced smoking, or enhanced energy conservation). These findings will help determine whether expert-designed nudges outperform citizen-designed nudges or whether a collaborative approach yields the best results.

Our main hypotheses to be tested are: H1: recommendations between the groups E and C will differ; H2: the opinion of the expert will strongly influence the group EC outcome; H3: if disclosed, groups C and EC outcomes will be more acceptable, whereas those of group E will be less acceptable. For the efficiency, we will ask other subjects to rate the *ex-ante* perceived efficiency. We hypothesis that group EC will be more efficient and more accept- able.

The overall objective of the paper is thus to critically assess the added value of citizen participation in behavioural policy design, both in terms of efficiency and acceptability. Data collection is expected for May-June 2025.

References

Bang, H. M., Shu, S. B., & Weber, E. U. (2018). The role of perceived effectiveness on the acceptability of choice architecture. Behavioural Public Policy, 4(1), 1–21.
https://doi.org/10.1017/bpp.2018.1

Blomkamp, E. (2018). The promise of co-design for public policy 1. In *Routledge hand- book of policy design* (pp. 59-73). Routledge.

Daniel Priolo, Isabelle Milhabet, Marilena Bertolino, Tom Juille, Dorian Jullien, Guilhem Lecouteux, Isma¨el Rafa¨ı & Pierre Thèrouanne (2023) Would you like some coffee with your sugar? A natural field experiment on the efficiency and acceptability of setting zero sugars as a default in coffee-vending machines, Comprehensive Results in Social Psychology, 7:1-2, 25-41, DOI: 10.1080/23743603.2023.2214964

John, P., Cotterill, S., Moseley, A., Richardson, L., Smith, G., Stoker, G. and Wales, C. (2019) Nudge, Nudge, Think, Think: Experimenting with Ways to Change Citizen Be- haviour. Manchester, UK: Manchester University Press.

Richardson, L., & John, P. (2021). Co-designing behavioural public policy: lessons from the field about how to 'nudge plus'. *Evidence & Policy*, *17*(3), 405-422.
Weber, E. U. (2015). Climate change demands behavioral change: What are the challenges?
Social Research: An International Quarterly, 82(3), 561–580.
https://doi.org/10.1353/sor.2015.0050

Keywords: nudge, expert, citizen, acceptability, efficiency

Is value a substance? A deconstruction of the institutional approach

Giustino De Michele*†[1,2]

[1]Université Paris 8 Vincennes-Saint-Denis – Université Paris 8 Vincennes-Saint-Denis – France

[2]Centre Interdisciplinaire d'Étude des Littératures d'Aix-Marseille – Aix Marseille Université, AMIDEX-Civis3i (MSCA Actions) – France

Abstract

What is value? Is it a substance? And for that matter, what is a substance? And what is a science of value, whether this is a substance or not? These questions seem eminently philosophical. They are in fact properly metaphysical, according to Aristotle's definition of a science of principles, and of the principle of any scientific endeavour. Nevertheless, they are also properly economic: it is precisely through these questions that, following and developing an institutional approach, French economist Andrè Orlèan, in his 2011 volume *L'empire de la Valeur: Refonder l''economie*, proposes a new conception of an economic theory, or, as he puts it, a new foundation of economics.
Orlèan's basic assumption is a critique of both a classical and a neoclassical approach ac- cording to which value would be a substance (whether it is defined in terms of labour or of utility) of which prices (and in particular their semiotic vehicle, which is to say: money) would be a secondary manifestation.

In order to oppose the notion of a substance of value, Orlèan draws on various disciplines: alternative economic theories (Thorsten Veblen), sociology (in the wake of Georg Simmel and in particular of Emile Durkheim, but Marcel Mauss's and even Georges Bataille's influence seems present in his arguments), and even literary studies. In fact, Orlean's most effective reference (see also Andrè Orlèan and Michel Aglietta, *La violence de la monnaie*, 1982) is to French literary theorist Renè Girard.

It is through the Girardian mimetic theory that Orlèan comes to develop a theory of the institution that liquidity is (currency being a case of liquidity, in his account) and thus to reinscribe a (marginalist) theory of equilibrium into a multiple-equilibrium institutional framework. In this context, value is not a substance, nor is it measurable according to any natural or stable standard. It is rather the result of a social act, of a collective, conventional (although implicit) engagement through which a social structure takes its very shape.

Moreover, Orlèan's effort is not meant to be theoretical per se. One of his aims is indeed to (once again) reinscribe economics in the wider context of social sciences. But the framework that he draws is meant to be more coherent with quantitative analysis, and in particular with such phenomena as financial bubbles, than standard economics is. Written in 2011 (see also Andrè Orlean, *De l'euphorie à la panique: Penser la crise financière*, 2009), *L'empire*

*Speaker
†Corresponding author: giustinodemichele@gmail.com

de la valeur considers that the 2008 financial crisis has revealed the limits of mainstream economic theories, incapable as they are to deal with instability and more generally with actual financial dynamics.

Can we go one step further? This paper, after outlining Orlèan's position and its merits, attempts to show how Orlean's own theory remains based on a metaphysical conception of value. More than this: as a matter of fact, this conception of value is substantialist, just like those that Orlean criticizes. Symptomatically, the process through which Orlèan himself identifies this (institutionalist) substance-value, that is, the "*puissance of the multitude*" (the power of a multitude, an expression borrowed from Baruch Spinoza), is perfectly described by his reformulation of the mimetic theory. However, the author does not seem to be aware of this analogy, and thus falls prey to the same criticism that he directs at his theoretical rivals (classical or neoclassical). Notably, he does not seem to be aware that the power of a multitude is a conventional, or at least a contingent (semiotic) value.

The analysis of Orlean's theoretical gesture, which strives to "refound economics", along with the identification of its strengths and limitations, and the description of the multi- faceted context from which it emerges, inevitably leads to the necessity of questioning the boundaries that separate economics from other disciplines. Furthermore, it calls into question the very nature of the position from which economics can assert itself as a legitimate and distinct scientific field.
In this light, a deeper reflection on the relationship between economics and its neighbouring fields is not only crucial for understanding the discipline's evolution, but also for envisioning its potential future trajectory. Emphasising and problematising the interdisciplinary institutionalist approach calls for the development of new methodologies, better equipping economics to address complex theoretical and societal issues.

Keywords: Value, Institutionalism, Philosophy, Mimetic Theory, Money

Reconceptualizing Individual Agency for the Ecological Transition

Maxime Desmarais-Tremblay*[1]

[1]Bureau d'Économie Théorique et Appliqué – CNRS, University of Lorraine, France – France

Abstract

For decades economists have been quasi-unanimous in their support for a carbon tax as a fair and efficient tool to reduce greenhouse gas emissions. Yet, such proposals have been dismissed at the federal level in the United States. They sparked a tax revolt in France in 2018. Canada did adopt a modest tax in 2019, but it is likely to be repealed in the near future. This gap between the policies recommended by experts and the political support they receive has led economists to study the 'beliefs and attitudes' of individuals about climate change policies (Douenne and Fabre 2022; Dechezleprˆetre et al. 2024; Meilland, Kervinio, and Mèjean 2025). While economists feel confident about measuring individual beliefs and confronting them to facts, they rarely embrace a rich normative conception of individuals (see Davis 2003). The standard policy approach in economics is consequentialist (see, Adler 2019). Policies are usually derived either from a social choice aggregation of individual preferences, or from a representative agent facing a tradeoff. Confronted with repeated failures to implement carbon pricing policies, many economists have broadened their conceptualization of political processes (Klenert et al. 2018; Blanchard, Gollier, and Tirole 2023; Combet and Pottier 2024). But references to moral philosophy are largely absent from this literature.
Could a different conception of the individual change our perspective on the right policies to reduce GHGs emissions? Beyond beliefs and preferences, persons also hold and express values. How should we account for the diverse and contradictory attitudes and values expressed by individuals? What is the nature and extent of individual responsibility in reducing GHG emissions?

According to Elizabeth Anderson's (1993) pluralist-expressive theory of action, individuals value different things in different ways. The theory is rational to the extent that individuals must give themselves, and others, reasons for their evaluations and their actions. Individuals have to make sense of their actions for themselves, in light of the past choices they have made. Individuals can take responsibility for their authentic choices, but it leaves open the question of the articulation of individual and collective responsibility. Actions are ap- propriate when they conform to public (expressive) norms. These norms in part determine expectations for economic behaviour. Norms can change and will have to change. But a too rapid transformation creates a distance between the values embedded in the legal and economic institutions and the expectations against which individuals made their life choice. This, in turns, can lead to resentment, and a sense of alienation (Taylor 1979).

How can a more comprehensive moral theory help us to make sense of individual attitudes to- ward climate change policies? A more realistic moral foundation could go beyond postulating

*Speaker

that individuals have 'distributive concerns'. For instance, a carbon tax can be considered unfair if it violates horizontal equity by exempting many industries and types of consumers. A concern for vertical equity might lead people to support progressive green transfers to compensate for other injustices in the fiscal system. But regressive policies that increase the cost of carbon might also be supported by higher order desires to change one's behaviour in the same way that anti-smoking policies were supported by some smokers. Moreover, individuals' value for equality might lead them to prefer banning combustion-engine cars to distributing fines to rebellious buyers. Besides, a sense of corrective justice might lead individuals to reject Coasian equivalence and defend the polluters pays principle. Finally, individuals often care for people who are closest to them, family members, then fellow citizens. Thus, the cheapest way to reduce emissions by a ton might not always be the best course of action because they might prefer to see the indirect benefits of political intervention closer to home.

References

Adler, Matthew D. 2019. Measuring Social Welfare: An Introduction. New York: Ox- ford University Press.

Anderson, Elizabeth. 1993. Value in Ethics and Economics. Cambridge, MA: Harvard University Press.

Blanchard, Olivier, Christian Gollier, and Jean Tirole. 2023. 'The Portfolio of Economic Policies Needed to Fight Climate Change'. Annual Review of Economics 15: 689–722.

Combet, Emmanuel, and Antonin Pottier. 2024. Un nouveau contrat ´ecologique. Paris: Presses Universitaires de France.

Davis, John B. 2003. The Theory of the Individual in Economics. London: Routledge.

Dechezleprˆetre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. 2024. 'Fighting Climate Change: International Attitudes Toward Climate Policies'. American Economic Review. Forthcoming. https://socialeconomicslab.org/research/wor papers/fighting-climate-change-international-attitudes-toward-climate-policies/.

Douenne, Thomas, and Adrien Fabre. 2022. 'Yellow Vests, Pessimistic Beliefs, and Car- bon Tax Aversion'. American Economic Journal: Economic Policy 14 (1): 81–110.

Klenert, David, Linus Mattauch, Emmanuel Combet, Ottmar Edenhofer, Cameron Hep- burn, Ryan Rafaty, and Nicholas Stern. 2018. 'Making Carbon Pricing Work for Citizens'. Nature Climate Change 8 (8): 669–77.

Meilland, Auriane, Yann Kervinio, and Aurèlie Mèjean. 2025. 'International Climate Jus- tice: What the People Think'. Environmental and Resource Economics 88 (1): 225–47.

Taylor, Charles. 1979. Hegel and Modern Society. Cambridge: Cambridge University Press.

Keywords: carbon tax, individual agency, responsibility, morality, climate justice, economic policy

Experimental elicitation of the ’beliefs’ of large language models with uncertain lotteries

Eléonore Dodivers[1] and Guilhem Lecouteux*[2]

[1]Groupe de Recherche en Droit, Economie et Gestion – Centre National de la Recherche Scientifique, Universit´e C^ote d’Azur – France
[2]Universit´e C^ote d’Azur, CNRS, Gredeg (France) – CNRS : UMR7321 – France

Abstract

Large Language Models (LLMs) are attracting considerable interest from the general pub- lic, and their democratisation is leading to a growing number of users using them as advisory tools in areas for which these models were not specifically designed (e.g. using ChatGPT for advice on health, financial investments, education, etc.). There is also a growing but rather limited literature that seeks to elicit the preferences of LLMs such as ChatGPT using the tools of experimental economics, such as Chen et al 2023, who aim to survey GPTs on different budget allocation choices, or Kim et al 2024, who reproduce the experimental protocol of Choi et al 2014 to measure the risk preferences of GPTs.

The aim of this paper is threefold. First, we aim to clarify ’what’ is measured when eliciting the preferences and beliefs of agents such as GPT. We propose an externalist interpretation `a la Dennett, i.e. that the preferences and beliefs are not inner mental states of LLMs, but rather the language of external observers to rationalise LLMs’ responses. If LLMs have been trained on a human-generated corpus, then the preferences and beliefs could be seen as the ’statistically more likely’ preferences and beliefs of humans (more precisely, the WEIRD population, see Atari et al 2023). This means that studying the preferences and beliefs of LLMs can indirectly inform researchers about the underlying preferences and beliefs of humans, and more precisely of the humans who feed the corpus on which the model is based – the choice of corpus is thus key to shaping the preferences of the LLM.

Second, we aim to propose an experimental protocol to elicit the preferences and beliefs of different LLMs, with three dimensions in mind: (i) determining the number of questions required to obtain exploitable answers (since the answers of LLMs are partly random de- pending on the temperature setting of the model; the requests will be submitted via the application programming interface (API) so that prompts can be submitted en masse with Python), (ii) determining the form of questions and answers (i.e. either a few questions with long answers or many short questions with many short answers), and (iii) investigating possible framing effects that could help to increase the ’quality’ of the answers (Dodivers et al 2025 document for instance that GPT3.5’s levels of donation in a Dictator Game should be interpreted as focal points rather than the expression of fairness, suggesting a high sensitivity to the context for some models).

* Speaker

Third, we aim to apply our protocol to the case of choices between uncertain lotteries. The states of the world of the lotteries refer to real financial scenarios (i.e. the likely evolution of the stock price of different companies, including the companies that own the LLMs we are testing), and we ask different LLMs (GPT 3.5, GPT-4, GPT o1-mini for OpenAI, and different versions of Gemini, Claude, Le Chat) to rank in terms of preferences a series of 10 portfolios whose outcomes depend on the realisation of the lotteries. These portfolios are designed to test the validity of the behavioural foundations of the expected value model (Wakker 2010, de Finetti 1931), which allows us to elicit the agent's beliefs about the different scenarios. Preliminary results from a pilot study suggest that the performance of GPT3.5 and GPT4 differ in terms of completeness and transitivity, and that their performance may be influenced by the company mentioned in the different scenarios. For the complete and transitive responses (about 25% of the queries), additivity and monotonicity appear to be systematically checked, meaning that we can infer beliefs about the states of the world for these responses. Data collection is expected in April 2025, we will test different forms of prompts, varying the description of the profile of the individual that the LLM is supposed to advise (e.g. 'risk neutral', 'risk preferences such that the person is indifferent between X and Y', etc.). This could also highlight possible paternalist tendencies for the LLM, and whether it respects the preferences of the person it is supposed to advise.

References

Atari, M., Xue, M. J., Park, P. S., Blasi, D., & Henrich, J. (2023). Which humans?. https://osf.io/preprints/psyarxiv/5b26t/download

Chen, Y., Liu, T. X., Shan, Y., & Zhong, S. (2023). The emergence of economic ratio- nality of GPT. Proceedings of the National Academy of Sciences, 120(51), e2316205120.

Choi, S., Kariv, S., Mu¨ller, W., & Silverman, D. (2014). Who is (more) rational?. American Economic Review, 104(6), 1518-1550.

De Finetti, B. (1931). Sul Significato Soggettivo della Probabilittext`a. Fundamenta mathe- maticae, 17.

Dennett, D. C. (1989). The intentional stance. MIT press.

Dodivers, E., Jelovac, I. Rafa¨ı, I, & Ventelou, B. (2025). Uncovering the Fairness of AI: Focal Point, Inequality Aversion, and Altruism in ChatGPT's Dictator Game Decisions. Working paper.

Kim, J., Kovach, M., Lee, K. M., Shin, E., & Tzavellas, H. (2024). Learning to be Homo Economicus: Can an LLM Learn Preferences from Choice. arXiv preprint arXiv:2401.07345.

Ross, D. (2012). The economic agent: Not human, but important. In U. M¨aki (Ed.), Philosophy of economics (pp. 691–735). Amsterdam: North-Holland.

Keywords: large language models, AI human advice, belief elicitation

Algorithms and Autonomy: Liberal Responses to the Preference-Shaping Power of Recommender Systems

Malte Dold*

[1]Pomona College – United States

Abstract

Algorithms in the form of recommender systems analyze user data to suggest items (such as products, movies, music, articles, or social connections) that users are likely to find ap- pealing. These systems can enhance user experiences by curating content and personalizing interactions based on past behavior and the behavior of others with similar tastes. However, algorithms also play a critical role in shaping preferences across various domains. They cu- rate decision-making environments through default options, framing effects, and gamification techniques, subtly nudging users toward certain behaviors. On social media, algorithms influence attitudes by prioritizing emotionally engaging or sensational posts. Personalized news feeds and search results further limit exposure to diverse perspectives, reducing opportunities for trial-and-error learning and fostering greater conformity within the system. Importantly, users often remain unaware of how algorithms shape their preferences, contribute to filter bubbles, or reinforce simplistic narratives. These dynamics raise complex ethical questions for liberals, particularly regarding the emphasis they place on individual autonomy.

In this project, I explore what the history of liberalism can teach us about addressing these challenges. Specifically, I examine three notions of autonomy in the liberal canon: autonomy as negative freedom or opportunities (Nozick/Hayek/Sugden), autonomy as stemming from a well-ordered hierarchy of motivations (Frankfurt/Dworkin), and autonomy as self- authorship or individuality (Mill/Raz). According to the opportunity conception, a person acts autonomously if they are free from coercion or external control, particularly by the state. The hierarchical conception understands autonomy as arising when first-order desires align with reflective higher-order preferences. The individualist conception, by contrast, views autonomy as achievable when the social conditions exist for individuals to engage in experiments of living allowing them to develop their own vision of the good life.

These different notions of autonomy lead to distinct policy responses to the preference- shaping power of algorithms. The opportunity conception cautions against state interventions in algorithms, arguing that such interventions face intricate knowledge problems and might lead to unintended consequences. The hierarchical conception endorses algorithms and recommender systems that help individuals align their desires with their higher-order preferences, while cautioning against those that undermine the reflective process. Finally, the individualist conception highlights the importance of fostering experimentation and ensuring exposure to diverse viewpoints, enabling individuals to craft their own paths. I will argue for the individualistic conception of autonomy, arguing that it provides the most relevant and effective framework for addressing the challenges posed by algorithmic recommender systems.

Keywords: autonomy, algorithms, behavioral normative economics, welfare

*Speaker

Reaping the Benefits of Interactive Scientific Pluralism: The Case of Ecological Economics

Jeremias Düring*†[1]

Bergische Universität Wuppertal – Germany

Abstract

Since its institutionalization in the late 1980s, ecological economics has positioned itself as a distinctly pluralistic school of economic thought (Norgaard, 1989). From its inception, there has been vigorous debate regarding the precise meaning of ”pluralism” within this school. Adhering to a ”methodological pluralism”, prominent figures within ecological eco- nomics advocate for a ”big tent” view (Howarth, 2008) that tolerates heterodox currents and more mainstream environmental economists within the same school of thought. Conversely, other ecological economists, most prominently Clive Spash (2015, 2024), argue for the exclusion of environmental economics from the tent and advocate for exclusive and intensified collaboration with other heterodox schools of thought.

The resolution of this pluralism question not only determines the composition of the epistemic community of ecological economics but also how it positions itself within the broader discipline of economics, specifically in relation to the mainstream and other heterodox schools of thought. Moreover, it provides novel insights for the broader pluralism debate within the history and methodology of economics, which addresses whether economics as an academic discipline has become excessively narrow, potentially impeding its capacity to address significant issues within its domain (such as inequality, climate concerns, or gender imbalances). This paper argues that a modified version of Hasok Chang’s (2012, 2022) interactive scientific pluralism is the best fit for addressing the pluralism question within ecological economics and the larger pluralism debate in economics.

This modified version of interactive pluralism builds upon a previously unexplored opportunity to integrate Chang’s original account of scientific pluralism with Helen Longino’s (Kellert et al., 2006; Longino, 2002, 2013) work on the pluralist stance and on critical contextual empiricism (CCE). As will be elucidated in the paper, there exists sufficient common ground to extend Chang’s (2022) three original ”modes” of interaction (integration, co-optation, and competition) between different ”systems of practice” by a fourth mode of ”critical interaction” (Longino 2013). The extended account of interactive pluralism offers a middle ground between the two factions within the internal pluralism debate in ecological economics. It mediates between the excessively permissive big tent view and the exclusionary and overly puristic position of Clive Spash. According to interactive pluralism, no school should be excluded from the ecological tent; however, this does not imply that all members coexist in harmony.

The first section introduces the internal pluralism debate in ecological economics and con- textualizes it within the broader discourse on the lack of plurality in economics. It will be demonstrated that Spash justifies the

*Speaker

† Corresponding author: jeremias.duering@uni-wuppertal.de

exclusion of mainstream economists with reference to a significant position within the wider pluralism discourse in economics. This position, known as structured pluralism, was proposed by historian of economics Sheila Dow

(2004) and posits that economics is organized into distinct schools of thought with clearly delineated boundaries between them. The subsequent section provides a more detailed analysis of structured pluralism, which yields two outcomes: First, it becomes evident that Spash's argument for purging ecological economics of mainstream economists is not substantiated by structured pluralism. While clear boundaries exist between the various schools of thought, they are considerably less well-defined than Spash posits, and, more importantly, these boundaries are permeable.

The fundamental premise of structured pluralism is that economics advances through critical discourse between opposing schools at these very boundaries. While this assertion may be valid, the second outcome of this section is that structured pluralism remains too narrow to adequately address the pluralism question for ecological economics or to serve as a fully convincing account for the broader debate in economics. Critical interaction across school of thought boundaries certainly constitutes one important mode by which schools of thought can interact with one another. However, solely emphasizing this agonistic mode of interaction overlooks more cooperative forms of engagement between different schools of thought. A more comprehensive and flexible conceptualization of pluralism is needed and will be provided in the third section.

Here, the above-mentioned modified version of interactive scientific pluralism will be introduced. It integrates Chang's original three modes of interaction (integration, cooptation, competition) with a fourth mode of critical interaction derived from Longino's work on scientific pluralism. Interactive pluralism precludes the exclusion of mainstream economists from ecological economics. However, this does not imply that everybody just coexists in harmony. The school of ecological economics, after all, is an epistemic community. As in any com- munity, there will be opportunities for cooperation, aligned with the modes of integration and co-optation, but there will also be criticism and competition among its members. This need not be detrimental to the community if these modes of interaction are harnessed and carefully channeled through agreed-upon rules.

This insight can be extended to the broader pluralism debate in economics: The discipline does not lack plurality; indeed, considerable plurality is manifested in numerous schools of economic thought. Nevertheless, the primary issue lies in the insufficient interaction among these schools along the lines of the four proposed modes of interaction. The reasons for this are multifaceted. Within the limited scope of this paper, I can only briefly highlight the significance of an improved social organization of economics that facilitates and incentivizes these types of interactions.

References

Chang, H. (2012). Is Water H2O?: Evidence, Realism and Pluralism. Springer Nether- lands. https://doi.org/10.1007/978-94-007-3932-1

Chang, H. (2022). Realism for Realistic People: A New Pragmatist Philosophy of Science. Cambridge University Press. https://doi.org/10.1017/9781108635738

Dow, S. C. (2004). Structured Pluralism. Journal of Economic Methodology, 11 (3), 275–290. https://doi.org/10.1080/1350178042000252965

Howarth, R. B. (2008). Editorial. Ecological Economics, 64 (3), 469. https://doi.org/10.1016/S0921-8009(07)00577-0

Kellert, S. H., Longino, H. E., & Waters, C. K. (2006). Introduction. The Pluralist Stance. In S. H. Kellert, H. E. Longino, & C. K. Waters, Scientific Pluralism (pp. vii–xxix). Uni- versity of Minnesota Press.

Longino, H. E. (2002). The Fate of Knowledge. Princeton University Press. https://doi.org/10.1515/9780691187013

Longino, H. E. (2013). Studying Human Behavior: How Scientists Investigate Aggression and Sexuality. University of Chicago Press.

Norgaard, R. B. (1989). The case for methodological pluralism. Ecological Economics, 1(1), 37–57. https://doi.org/10.1016/0921-8009(89)90023-2

Spash, C. L. (2015). The content, direction and philosophy of ecological economics. In J. Martinez-Alier (Ed.), Handbook of Ecological Economics (pp. 26–47). Edward Elgar Publishing. https://doi.org/10.4337/9781783471416.00006

Spash, C. L. (2024). Foundations of social ecological economics: The fight for revolutionary change in economic thought. Manchester University Press. https://doi.org/10.7765/9781526171498

Keywords: Ecological Economics, Pluralism Debate, Scientific Pluralism, Modes of Interaction

The Epistemic Boundaries of Economic Informality: Rethinking Methodological Approaches in Economic Policy

Ceyhun Elgin*[1]

[1]Bogazici University – Turkey

Abstract

The informal economy has long been a subject of economic inquiry, yet its methodological treatment remains contested. While economic models often categorize informality as a deviation from formal institutional structures, this paper argues that informality is not merely a residual category but a distinct epistemic domain requiring a reconceptualization of economic methodology. Drawing on empirical findings from fiscal policy responses, gendered labor informality, and cross-national comparisons, this study critically examines how different methodological approaches-ranging from structuralist to agent-based and institutionalist perspectives-shape our understanding of informal economies.

Traditional economic models have often treated informality as a byproduct of inadequate regulation, weak institutions, or economic underdevelopment. The dualist framework, for instance, contrasts formal and informal sectors, assuming that informality is an anomaly to be eradicated through economic modernization. Meanwhile, the structuralist perspective sees informality as a necessary function of capitalist economies, serving as a shock absorber during economic downturns. Yet, both approaches tend to neglect the methodological implications of treating informality as a dynamic and evolving economic space rather than a static category.

This paper challenges these conventional paradigms by arguing that informality must be analyzed not only as an economic phenomenon but as an epistemic construct that demands methodological pluralism. The persistence of informal economies across diverse contexts suggests that informality cannot merely be interpreted as a temporary stage of economic development. Instead, informality should be understood as an integral and evolving feature of economic life, shaped by specific institutional, political, and social conditions.

One of the key challenges in studying informality is the measurement problem. Informal economies, by definition, operate outside officially recorded data, making them difficult to quantify. Traditional econometric techniques rely on proxy variables such as cash transactions, self-employment rates, or discrepancies between national accounts and survey data. However, these measures often fail to capture the complexity and heterogeneity of informal economic activities. This paper explores alternative methodological approaches, including:

- **Agent-Based Models (ABMs):** ABMs simulate the interactions of heterogeneous agents within an economy, offering insights into how informal practices emerge and evolve.

*Speaker

By modeling decentralized decision-making and adaptive behavior, ABMs allow for a more nuanced understanding of informality beyond the constraints of aggregate macroeconomic models.

- **Institutionalist Perspectives:** New institutional economics provides an alternative framework for understanding informality by examining the role of formal and informal rules in economic transactions. This approach emphasizes the interplay between state regulation, social norms, and economic incentives, suggesting that informality is not merely the absence of formal institutions but the presence of alternative institutional arrangements.
- **Comparative-Historical Analysis:** By situating informality within broader historical and political-economic contexts, this approach helps reveal how different policy regimes, labor market structures, and governance models shape the informal economy. Case studies from different regions, including Latin America, Southeast Asia, and post-socialist economies, illustrate how informal practices are embedded in specific socio-economic and political histories.
- **Behavioral Economics and Informality:** Standard economic models often assume that economic actors behave rationally within the constraints of formal institutions. How- ever, behavioral economics challenges these assumptions by highlighting cognitive biases, social norms, and risk perceptions that shape decision-making in informal markets. For instance, informal entrepreneurs may choose to remain outside the formal sector not only due to cost-benefit calculations but also due to trust deficits in government institutions or cultural preferences for informal networks.

The way informality is conceptualized has direct implications for economic policy. Many policy approaches seek to formalize informal economies through taxation policies, labor regulations, and financial inclusion programs. However, such efforts often fail because they assume informality is a problem to be solved rather than a complex system with its own logic and resilience.

This paper argues that traditional policy interventions based on standard economic method- ologies frequently lead to unintended consequences. For example, policies aimed at increasing tax compliance among informal businesses often ignore the social and economic structures that sustain informality, such as family networks, informal credit systems, and community- based trust mechanisms. Similarly, attempts to integrate informal workers into formal labor markets may overlook the agency of workers who deliberately opt for informal arrangements due to greater flexibility or cultural factors.

Moreover, methodological biases in mainstream economic research have contributed to a misalignment between policy prescriptions and the realities of informal economies. Standard macroeconomic models often exclude informality from their frameworks, treating it as an externality rather than an intrinsic part of economic systems. This omission leads to a lack of policy tools tailored to informal economies, resulting in ineffective governance and economic policies that exacerbate inequality rather than addressing its root causes.

To address these gaps, this paper proposes a methodological shift towards pluralism in eco- nomic research on informality. A more integrative approach-one that combines quantitative modeling with qualitative insights, institutional analysis, and interdisciplinary perspectives- can lead to better policy design and more accurate assessments of informal economies.

This study calls for a broader epistemological shift in economic methodology. By critically examining the conceptual boundaries of informality, it challenges the notion that informal economies exist as residual spaces outside formal economic systems. Instead, informality should be viewed as an evolving economic order with its own structures, institutions, and rationalities.

In light of this, future research should:

- **Develop more robust mixed-method approaches** that integrate qualitative field- work, ethnographic insights, and survey-based empirical models to better capture informal economic dynamics.

- **Explore the intersections of informality with other economic paradigms**, such as feminist economics, postcolonial critiques, and ecological economics, to enrich theoretical understandings of informality.
- **Examine the role of digitalization and platform economies** in reshaping informal work, particularly in the gig economy and decentralized finance (DeFi) spaces.

By synthesizing insights from economic methodology, political economy, and the sociology of economics, this paper contributes to broader discussions in economic epistemology. It advocates for an interdisciplinary approach to understanding informal economies that goes beyond binary distinctions between formal and informal, instead recognizing the fluidity and complexity of economic life.

In conclusion, the study of informality requires a rethinking of economic methodologies that move beyond conventional models and assumptions. By acknowledging the epistemic boundaries of informality and adopting a pluralist approach, economists can develop more accurate, context-sensitive, and effective policy frameworks. This paper ultimately aims to contribute to the broader discourse on economic methodology and informality by providing a critical lens through which to reassess the relationship between economic theory, policy, and lived economic realities.

Keywords: informality, economic policy, institutional economics

Policy Post-Growth and the Lack of Diversity in the Scenario Framework

Rawad El Skaf*[1]

[1]Leibniz University Hannover – Germany

Abstract

In December 2023, the president of the 28th Conference of the Parties Sultan Al Jaber defended during an online debate that ”there is no science out there, or no scenario out there, that says that the phase-out of fossil fuel is what’s going to achieve 1.5C” (The Guardian, 3 December 2023). While much could be said regarding the inaccuracies and falsehoods in this statement, my focus in this paper is on the link between science, scenarios, and policy from a philosophy of science perspective.

There are many definitions of the concept ”scenario” (Brewer 2007; Schwartz 1991; van der Heijden 1996, Selin 2006), but in all cases the emphasis is on the idea that scenarios are not predictions, and are not supposed to be true, but rather they are different projections of alternative futures which aim to be useful for formulating strategies and making decisions. Introduced by Herman Kahn in the 1960s, scenarios were first used to plan for aspects of nuclear warfare in the early Cold War period (Galison 2014). Since then, scenario-based science has been democratised, especially in research on topics with important societal and environmental consequences.

Why use scenarios in the context of climate change research? Research into climate change depends on a myriad of factors, but two categories of information are essential. The first is to know what changes are going to take place in a future climate. For instance, how much the temperature will rise over the course of the 21st century, how much the Arctic Sea ice will decline, what is the extent of the instability of polar ice sheets and the resulting additional sea-level rise. The second category of information essential to climate change research is the socio-economic structure of the future world. This latter will release a quantity of greenhouse gases (GHGs) and therefore further mitigate or exacerbate climate change. It will also suffer the various effects of climate change and try to adapt to its new environment. The complexity and the deep uncertainties characterizing these two systems taken separately – climatic and socio-economic – as well as the complexity of their interaction, make scientific prediction of these two futures impossible, even probabilistically. This has led climate change research to move towards a scenario-based science.

Scenarios could be individual; a researcher could imagine, design, and analyse scenarios for her own purposes. However, in a globally coordinated climate change research, shared common scenarios are needed to analyse and compare climate change mitigation, adaptation and impact studies by a highly interdisciplinary and international community. For that, the Integrated Assessment Modeling Consortium has coordinated the development of the ”SSP-RCP scenario framework”, which already served as the basis to thousands of research

* Speaker

papers and projects (O'Neill et al. 2020) and the latest IPCC (Intergovernmental Panel on Climate Change) 6th assessment report (AR6). This will continue to be the case for the next AR7 report. In this framework, SSPs (Shared Socioeconomic Pathways) consist of a limited set of alternative socioeconomic futures that are both represented in short qualitative narratives and with quantitative projections of key drivers. However, all five SSPs share a common Global Economic Growth Assumption (GEGA). That is, all current SSPs assume, with varying degrees, that the economies of the Global North and those of the Global South will keep on growing until the end of the century.

This results in the following well-known problem: GEGA is both historically and theoretically tightly correlated with GHG emissions. SSP-based mitigation scenarios, both consistent with GEGA and aim at achieving 1.5◦C or below 2◦C targets complying with the Paris Agreement, need to find ways to solve this problem. These solutions tend to be either (a) negative emissions-oriented mitigation scenarios, which means that they rely on some technology to remove GHG from the atmosphere in the future, or (b) decoupling-oriented mitigation scenarios, which means that they mainly rely on the contested hypothesis that one could absolutely decouple economic growth and GHG emissions, or more generally the assumption that economic growth can continue while reducing environmental harm.

This is not just a scientific problem; mitigation scenarios are a central science for policy tool. They aim at presenting decision-makers with a range of a "manageable" set of options to achieve a certain mitigation target. This is done to a large extent via the IPCC's reports where "most of the scenarios in the AR6 database are SSP-based" (Riahi et al. 2022, p.305) and Edenhofer and Kowarsch's (2015) Pragmatic-Enlightened Model (PEM) model. However, having only SSP mitigation scenarios consistent with GEGA restricts the range of solutions in a value-laden manner, contrary to the IPCC's general mandate of being "policy-relevant and yet policy-neutral, never policy-prescriptive". This restriction is value- laden because GEGA is a value-laden assumption and could be challenged by appealing to different values, as evidenced by the "post-growth" (1) literature (e.g. Kallis et al., 2018; Kuhnhenn, 2018; Hickel, 2019; Otero et al., 2020; Keyßer and Lenzen, 2021; Hickel et al., 2021; Li et al., 2023; Kikstra et al., 2024; Slamerˇsak et al., 2024). By relaxing GEGA for the Global North, post-growth scholars aim at proposing alternative mitigation scenarios. However, for these scenarios to be constructed, explored, and assessed more systematically, they need to be fully integrated into the scenario framework, something not yet done.

In this paper I argue that this integration could and even should be done. To spell out my argument, the paper is structured as follows. In section 2, I introduce the scenario frame- work and explain how SSP-scenarios are derived from SSP-pathways and are used to do mitigation (as well as adaptation) comparative analysis. Most importantly, in this frame- work an SSP-scenario needs to be consistent with an SSP-pathway, thus underlying the importance of GEGA in all five SSPs. In section 3, I critically present how scenarios are used to guide decision-making following Edenhofer and Kowarsch's (2015) PEM model. This approach promises to uphold IPCC's general mandate by deferring controversial value-laden judgements to decision-makers. Here I argue that by upholding GEGA in the theoretical framework and deriving only GEGA compatible mitigation scenarios, scientists are considering this assumption as consensual. However, GGEA is controversial and thus should be reflected in the mitigation options. In section 4, I reconstruct and synthesise in a single deductive argument several criticisms coming from the post-growth literature and calling for the integration of their research into the scenario framework. I then propose two ways for doing this. In section 5, I argue that this integration enhances the diversity of scenarios that could be constructed under the scenario framework. This is important, since from a philosophy of a value-laden science perspective, diversity is a well-suited guide to "objectivity". Integrating post-growth into the scenario framework thus enhances this latter's objectivity and hence IPCC's policy-neutrality. I conclude in section 6 with some thoughts on plausibility, but defend that, in some cases (to which post-growth belongs), diversity comes first; for plausibility or feasibility assessment across dimensions (Brutschin et al., 2021) to be done correctly, they should be done a posteriori. For that, an integration of post-growth into the scenario framework is needed.

Keywords: Climate change economics, scenario framework, diversity, objectivity, IPCC

The ”Mortality Paradox”: History and Methodology of Poverty Measurement

Judith Favereau*[1]

[1]Leibniz University Hannover – Germany 1Universitè Lumière Lyon-2 – Universitè Lumiére-Lyon 2, Triangle – France

Abstract

Poverty has an abundance and diversity of statistical measures. Yet, despite this diversity, when poverty is measured several paradoxes emerge, leading to refinements of its measurement. Sketching the history of poverty measurement can be seen as drawing the history and resolution of these paradoxes. Nonetheless, one paradox appears persistent for most of the poverty measures: the so-called mortality paradox. The more poverty kills, the fewer poor people there are, and the lower the poverty measure is. Lefebvre, Pestieau, and Ponthière (2013, 2018, 2019a, 2019b) and Mukherjee (2007) speak of the ”missing poor” or the ”mortality paradox.”

This paradox finds its origins in reflections on the question of income distribution within poverty measurement. Sen (1976) showed that withdrawing a portion of income from the poor has no impact on poverty measurement, as the number of poor remains unchanged. This observation has given rise to income distribution-sensitive measures of poverty, the most widely used in applied economics being that developed by Foster et al. (1984). But above all, this has shown the need to think of poverty in terms of deprivation, to distinguish between different levels of poverty and their distribution. Sen (1976) and Mukherjee (2001) axiomatize this deprivation. Mukherjee (2001) goes a step further, thinking of it in multidimensional terms. The measures proposed by Sen (1976), Foster et al (1994) and Mukherjee (2001) provide a mathematical solution to the distribution paradox. However, these measures are still insensitive to population variations, i.e. they do not allow for the missing poor. Kanbur and Mukherjee (2007) then extend Mukherjee’s (2001) previous axiomatization by adding an axiom that allows the missing poor to be integrated using a time dimension. Such axiomatizations have led to an analysis in terms of social welfare. Indeed, Fleurbaey and Ponthi`ere (2022) and Fleurbaey, Leroux and Ponthi`ere (2014) have proposed some possible compensations for these premature deaths.

For all that, more general ethical and epistemological reflections on the definition (shaping) of poverty as deprivation have not yet been conducted in connection with the mortality paradox. Nevertheless, Sen (1993, 1998) opens the way by questioning the value of a mortality indicator to measure economic successes and failures. According to Sen (1993, 1998), mortality data reveal the deprivations faced by individuals. These deprivations are invisible when one focuses solely on economic data, such as GDP per capita: ”For example, one country can have much lower life expectancy than its less wealthy counterpart because its citizens have poor access to health care and basic education. Mortality data can be used to evaluate policy and to identify vital aspects of economic deprivation in particular nations and in specific groups within nations” (Sen, 1993, p. 40).

* Speaker

This paper aims to place mortality as a crucial dimension of poverty measurement. To this end, I primarily draw on Amartya Sen's works on both mortality and development ethics. First, I show how mortality data reveal crucial privations that are invisible with common poverty measures. This raises a central methodological question on measurement in general about what numbers reveal or hide. For instance, on the one hand for international comparison, one might prefer a thin description. However, such a description might conceal crucial poverty characteristics (e.g. privations). On the other hand, one might favor revealing those crucial characteristics, which might make the measure heavy, thus hardly transposable. The question becomes what to decompose and how much. Mortality data encapsulates several privations, which also emphasize how people live. Consequently, it might highlight what to focus on and how to do so, without being unduly burdensome. Secondly, mortality data combined with participatory approaches can be mobilized as an ethical tool. This relates to one crucial ethical issue concerning poverty measurement, namely whether it should be expressed, for instance, through income, basic needs, primary goods, or capabilities. Once more, retracing the history of these criteria emphasizes several paradoxes. Capabilities as being open, plural, and context-dependent might solve some of these paradoxes. However, they have been subjected to significant criticism for their difficult operationalization. In- deed, they should be identified through public discussion, but how to implement such a public discussion is unclear. This would require precisely defining both the topic and the frame of the discussion. Participatory approaches as developed by Sabina Alkire could be employed to define the frame. Mortality data, which reveals privations, could act as a proxy for capabilities and could be the topic of this public discussion.

References

Fleurbaey, M., & Ponthiere, G. (2022). Measuring well-being and lives worth living. Eco- nomic Theory, 1-20.

Fleurbaey, M., Leroux, M. L., & Ponthiere, G. (2014). Compensating the dead. Journal of Mathematical Economics, 51, 28-41.

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. Econometrica: journal of the econometric society, 761-766.

Hacking, I. (2000). Fac ̧onner les gens: le seuil de pauvret ́e. L'E ̀re du chiffre: syst`emes statistiques et traditions nationales. Quebec: Presses de l'Universit ́e du Qu ́ebec, pp. 17-36.

Kanbur, R., & Mukherjee, D. (2007). Premature mortality and poverty measurement. Bul- letin of economic Research, 59(4), 339-359.

Lefebvre, M., Pestieau, P., & Ponthiere, G. (2019a). Missing poor and income mobility. Journal of Comparative Economics, 47(2), 330-366.

Lef ̀ebvre, M., Pestieau, P., & Ponthiere, G. (2019b). Premature mortality and poverty measurement in an OLG economy. Journal of Population Economics, 32, 621-664.

Lefebvre, M., Pestieau, P., & Ponthi`ere, G. (2018). FGT Old-Age Poverty Measures and the Mortality Paradox: Theory and Evidence. Review of Income and Wealth, 64 (2), 428-458.

Lefebvre, M., Pestieau, P., & Ponthi`ere, G. (2014). Mortalit ́e diff ́erentielle et pauvret ́e par ˆage. Revue fran ̧caise d' ́economie, 29(4), 173-196.

Lefebvre, M., Pestieau, P., & Ponthi`ere, G. (2013). Measuring poverty without the mor- tality paradox. Social Choice and Welfare, 40, 285-316.

Mukherjee, D. (2001) Measuring Multidimensional Deprivation. Mathematical Social Sci- ences, 2001 : 233-251.

Sen, A. (1998). Mortality as an indicator of economic success and failure. The economic journal, 108(446), 1-25.

Sen, A. (1993). The economics of life and death. Scientific American, 268(5), 40-47. Sen, A. (1976). Poverty: an ordinal approach to measurement. Econometrica, 219-231.

Keywords: poverty, measurement, methodology

A Defence of Preference-Based Approaches to Welfare Evaluations in Public Policy

Roberto Fumagalli*[1]

[1]King's College London, London School of Economics, University of Pennsylvania – United Kingdom

Abstract

Economists, philosophers and policymakers (henceforth, EPPs) commonly attempt to ground *welfare evaluations* of public policies on information concerning individuals' preferences (e.g. Angner, 2016, Manzini and Mariotti, 2014, Rubinstein and Salant, 2012). However, several empirical findings indicate that individuals' *actual* preferences - i.e. the preferences individuals happen to have in the examined choice settings - frequently fail to meet the information conditions and the consistency conditions putatively required to reliably track individuals' welfare (e.g. Gilboa et al., 2009, on cases where individuals' actual preferences rest on inaccurate information about the available choice options; Sugden, 1991, on cases where individuals' actual preferences violate transitivity). In light of these findings, many EPPs take individuals' *informed* preferences - i.e. the preferences individuals are able to form on the basis of accurate information and considerate judgments concerning their options - to reliably track individuals' welfare (e.g. Bernheim, 2009, Hausman, 2012, ch.7-8). According to these *preference-based approaches* to welfare evaluations (henceforth, PREF), individuals are well-off to the extent that their informed (rather than actual) preferences are satisfied, and the satisfaction of individuals' informed preferences reliably tracks individuals' welfare (e.g. Fumagalli, 2024, Hausman, 2022).
In recent years, several prominent authors have criticized EPPs' reliance on PREF for hampering central democratic and liberal values such as recognition respect, individual agency, anti-paternalism, liberal neutrality between different conceptions of the good, and socio- economic justice. This paper combines insights from behavioral economics, political philosophy and cross-disciplinary well-being research to defend PREF against these objections. My main claim is that the proponents of PREF can successfully address those objections and that EPPs are justified in relying on PREF. If correct, this result has important implications for welfare evaluations in public policy. For PREF and other approaches to welfare evaluations (e.g. Nussbaum, 2000, on capability approaches; Sugden, 2018, ch.4-5, on opportunity-based approaches) ground different evaluations of policies' welfare implications and ground such evaluations on dissimilar justificatory principles (e.g. Adler and Fleurbaey, 2016, Fumagalli, 2021, Thoma, 2021).

The paper is organized as follows. Section 1 outlines the main tenets of PREF and distinguishes influential versions of PREF (e.g. Griffin, 1986, ch.1-2, on preference satisfaction theories of welfare; Hausman, 2024, on the evidential account of welfare). Sections 2-6 explicate and address five influential objections put forward against PREF, namely: the objection from *recognition respect*, which holds that EPPs should not adopt PREF because adopting PREF leads EPPs to disrespect the individuals they target (e.g. Terlazzo, 2016); the objection from *individual agency*, which holds that EPPs should not adopt PREF because PREF

*Speaker

equate agency with consistent preferences and thereby fail to recognize the continuing identity of individuals with inconsistent preferences (e.g. Sugden, 2010); the objection from *anti-paternalism*, which holds that EPPs should not adopt PREF because adopting PREF leads EPPs to endorse unjustified (or otherwise objectionable) paternalistic interventions (e.g. Rizzo and Whitman, 2020, ch.7-8); the objection from *liberal neutrality*, which holds that EPPs should not adopt PREF because adopting PREF leads EPPs to violate liberal neutrality between different conceptions of the good (e.g. Arneson, 2006); and the objection from *socio-economic justice*, which holds that EPPs should not adopt PREF because adopting PREF hampers EPPs' ability to address the socio-economic factors that make it difficult for individuals to form and act on welfare-enhancing preferences (e.g. Nussbaum, 2001).

References

Adler, M. and Fleurbaey, M. Ed. 2016. *Oxford Handbook of Well-Being and Public Policy.* Oxford University Press.

Angner, E. 2016. Well-being and economics. In G. Fletcher (Ed.), *The Routledge Hand- book of Philosophy of Well-being*, 492-503. Routledge.

Arneson, R. 2006. Desire formation and human good. *Royal Institute of Philosophy Supplement*, 59, 9-32.

Bernheim, D. 2009. Behavioral welfare economics. *Journal of the European Economic As- sociation*, 7, 267-319.

Fumagalli, R. 2021. Theories of well-being and well-being policy. *Journal of Economic Methodology*, 28, 124-133.

Fumagalli, R. 2024. Preferences versus opportunities: on the conceptual foundations of normative welfare economics. *Economics & Philosophy*, 40, 77-101.

Gilboa, I., Postlewaite, A. and Schmeidler, D. 2009. Is it always rational to satisfy Sav- age's axioms? *Economics & Philosophy*, 25, 285-296.

Griffin, J. 1986. *Well-Being: Its Measure and Importance.* Clarendon Press.

Hausman, D. 2012. *Preference, Value, Choice, and Welfare.* Cambridge University Press.

Hausman, D. 2022. Enhancing welfare without a theory of welfare. *Behavioral Public Policy*, 1-16.

Hausman, D. 2024. Subjective total comparative evaluations. *Economics & Philosophy*, 40, 212-225.

Manzini, P. and Mariotti, M. 2014. Welfare economics and bounded rationality: the case for model-based approaches. *Journal of Economic Methodology*, 21, 343-360.

Nussbaum, M. 2000. *Women and Human Development: The Capabilities Approach.* Cam- bridge University Press.

Nussbaum, M. 2001. Adaptive preferences and women options. *Economics & Philosophy*, 17, 67-88.

Rizzo, M. and Whitman, D. 2020. *Escaping Paternalism: Rationality, Behavioral Eco- nomics, and Public Policy*. Cambridge University Press.

Rubinstein, A. and Salant, Y. 2012. Eliciting welfare preferences from behavioral datasets. *Review of Economic Studies*, 79, 375-387.

Sugden, R. 1991. Rational choice: a survey of contributions from economics and philosophy. *The Economic Journal*, 101, 751-785.

Sugden, R. 2010. Opportunity as mutual advantage. *Economics & Philosophy*, 26, 47– 68.

Sugden, R. 2018. *The Community of Advantage*. Oxford University Press.

Terlazzo, R. 2016. Conceptualizing adaptive preferences respectfully. *Journal of Political Philosophy*, 24, 206–226.

Thoma, J. 2021. On the possibility of an anti-paternalist behavioural welfare economics. *Journal of Economic Methodology*, 28, 350-363.

Keywords: Welfare, Preferences, Choices, Methodology, Policy Evaluation.

Channels of Model Performativity

Benjamin Genta*[1]

[1]University of California [Irvine] – United States

Abstract

In the early days of the Covid-19 pandemic, the sociologist Zeynep Tufekci (2020) wrote in *The Atlantic* about the predictions of pandemic models: ”Right answers are not what epidemiological models are for (...) These models are not snapshots of the future.” Rather, she sees their main use as providing knowledge of possible scenarios on which to act on, not as a tool to predict the world accurately. Some philosophers of science have taken an approach along these lines (Schroeder 2021; Douven 2023).
One reason for why a model might fail to be predictive is because the dissemination of the model itself changes the world. Let us take the example of a pandemic model which makes certain correct assumptions about how people interact. If this model projects an alarmingly high number of deaths due to illness, then upon learning of the model, people's interactional behavior might change-say, by voluntary social distancing. If people's behavior changes because of their exposure to the model, the assumptions that led to the model in the first place (assuming they were correct to begin with) will not hold anymore.

Tufekci writes that such cases ”can make a model look false. But that's not always what happened. It just means we won. And that's why we model.” In other words, overshooting can sometimes be a model virtue since it will lead to socially desirable outcomes. This thought of relegating the predictive success of certain models-echoed particularly throughout the coronavirus pandemic-is gaining traction. But this seems to turn scientific inquiry upside down: instead of predicting or explaining the world, the model is meant to change behavior, presumably for the better.

This kind of phenomenon-the world changing because of theory-has been discussed at length in sociology under the banner of *performativity* (e.g. Callon 1998). The pandemic has reinvigorated interest in scientific performativity among philosophers of science. Van Basshuysen et al. (2021), for example, argue that models can perform a pandemic, while Van Basshuysen (2023) suggests that some models should be evaluated based on their performative impact.

The performativity models raises important epistemic questions. Many of these concerns have been raised by philosophers working on *looping, reflexive,* and/or *interactive* effects (Hacking 2007; Godman & Marchionni 2022). While performativity is often conflated with these effects (Marchionni et al. 2024), there are fundamental differences. For instance, discussions of looping and reactive effects mostly focus on the change a theory might have on the behavior of the individual described by that theory.

However, in the case of social science models-as opposed to cases of the performativity of psychological concepts-individual behavioral change is only one dimension of performativity.

*Speaker

Often, it is not even the most significant one. Furthermore, while looping effects and reactivity are typically seen as ”unintended effects” of models (Northcott 2022), performative effects can be deliberate-in fact, in paradigmatic cases, models are explicitly designed to reshape the world.

Despite the considerable literature on performativity, little work has been done on *how* models perform the world. Sociologists have largely focused on empirical case studies (Boldyrev & Svetlova 2016; Healy 2015), often critiquing the performative effects of economics without attempting to extract general features.

My goal is to fill this gap by developing a theoretical framework for under- standing the mechanisms through which models perform. Via case studies from economics and public health, I argue that models perform the world when they change at least one of the following: (1) the legal infrastructure; (2) the material or technological infrastructure; or, (3) the beliefs or preferences of individuals the model purports to represent.

Francesco Guala (2007) makes a similar distinction, suggesting that performative models may inform institutional design or influence individual behavior. However, he considers the former-what I identify as changes to legal and material/technological infrastructure-to be a less ”distinguished” form of performativity, focusing instead on belief and behavior change. I will show that much of the most interesting work on performativity involves examining how
(1) and (2) are transformed.

The talk is structured as follows: first, I lay out the basics of performativity (Austin 1962). I then discuss some recent accounts (such as Van Basshuysen et al. (2021) and Hacking (2007)), lay out their shortcomings, and propose an alternative way of understanding how models may perform. To support my framing, I present three case studies: kidney matching markets (Roth 2015), the Black-Scholes-Merton model of option trading (Mackenzie 2006), and a pandemic model (Ferguson et al. 2020). I then consider some philosophical upshots of this framing: namely, it gives us a more nuanced understanding of what it means for a model to be performative (and, in particular, what it means for the model to be *successfully* performative).

References

Austin, J. L. (1962/1975). *How to Do Things with Words*. Harvard University Press.

Boldyrev, I., & Svetlova, E. (2016). After the turn: How the performativity of economics matters. *Enacting dismal science: New perspectives on the performativity of economics*, 1-27.

Callon, M. (1998). Introduction: the embeddedness of economic markets in economics. In: Callon, M. (ed.), *The Laws of the Markets*, New York, Blackwell Publishers, 1–57.

Douven, I. (2023). Pandemics and flexible lockdowns: In praise of agent-based modeling. *European Journal for Philosophy of Science*, *13*(3), 35.

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., ... & Ghani, A. C. (2020). *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand* (Vol. 16). London: Imperial College London.

Godman, M., & Marchionni, C. (2022). What should scientists do about (harmful) interactive effects?. *European Journal for Philosophy of Science*, *12*(4), 63.

Hacking, I. (2007). Kinds of people: Moving targets. *Proceedings of the British Academy*, 151, 285–318.

Healy, K. (2015). The performativity of networks. *European Journal of Sociology/Archives. Europ´eennes de Sociologie*, *56*(2), 175-205.

MacKenzie, D. (2006). Is economics performative? Option theory and the construction of derivatives markets. *Journal of the History of Economic Thought*, *28*(*1*), 29–55.

Marchionni, C., Zahle, J., & Godman, M. (2024). Reactivity in the human sciences. Euro- pean Journal for Philosophy of Science, 14(1), 1-24.

Roth, A. E. (2015). Who Gets What-and Why: the new economics of matchmaking and market design. Houghton Mifflin Harcourt.

Schroeder, S. A. (2021). How to interpret Covid-19 predictions: Reassessing the IHME's model. *Philosophy of Medicine*, *2*(*1*).

Tufekci, Z. (2020). Don't believe the COVID-19 models: That's not what they're for. *The Atlantic*.

Van Basshuysen, P., White, L., Khosrowi, D., & Frisch, M. (2021). Three ways in which pandemic models may perform a pandemic. Erasmus Journal for Philosophy and Economics, 14(1), 110-127.
Van Basshuysen, P. (2023). Austinian model evaluation. Philosophy of Science, 90(5), 1459- 1468.

Keywords: Performativity, Models, Mechanism Design

Assessing Development Projects through the Lens of the Human Development and Capability Approach: How a More Comprehensive Methodological Framework Can Reshape Development Policies.

Pietro Ghirlanda*†[1] and Enrica Chiappero-Martinetti[1]

[1]Università degli Studi di Pavia [Italia] = University of Pavia [Italy] = Universitè de Pavie [Italie] – Italy

Abstract

Research Context and Aims

Disadvantaged countries are historically threatened by a series of multifaceted crises that endanger individuals' and communities' capabilities. Among these crises, one that is often forgotten is a measurement crisis concerning the assessment of development projects. Project assessment is in fact crucial for understanding the effectiveness of actions and commitments taken for addressing emergencies, deciding how and where to allocate the available resources, and designing future interventions aimed at improving the multiple dimensions of human wellbeing. However, it is deeply influenced by the institutional context that establishes the "rules of the game" for development interventions.

In this sense, the methodology which currently represents the 'gold standard' in development economics, i.e. randomized controlled trials (RCTs), is difficult to apply in some important circumstances: a) when development programs are broad in scope and aim to integrate economic value creation with social and environmental sustainability, b) when they have been already implemented and thus it is required a retrospective assessment, c) when their effects unfold in a medium/long-term horizon and are not limited to the short-term, d) when their implementation in a certain domain is likely to produce cross-sectorial and cross-thematic spill-overs in other domains, e) when an operational approach to evaluation is equally relevant.

Accordingly, the qualities of RCTs, which are a consolidated, scientifically rigorous, and easily replicable method, go together with some limitations that risk compromising the aim of development policies and creating a sort of "technocratic illusion." Namely, being characterized by the necessity to create a manageable experimental setting to isolate few measurable variables and assess the impact of a "treatment" on a restricted sample of participants vis- `a-vis a control group that does not receive the treatment, it can be tricky for RCTs to achieve external validity (Achen, 2022). Moreover, other elements that bring their reliability into question are their feigning ignorance about the content of experience, common knowledge, and best practices (Pritchett, 2020), their lack of a holistic perspective, and their being guided by a rule-utilitarian inspiration indifferent to distributive concerns (Henderson, 2022). In particular, the so-called "equipoise" requirement, i.e. conducting experiments only

*Speaker

† Corresponding author: pietro.ghirlanda@unipv.it

when there is reasonable uncertainty over the benefits of treatments-without creating clear disadvantages in the control group-is difficult to satisfy in development contexts.

In light of this scenario, this paper aims to propose the human development and capability approach (HD&CA) as a richer informational basis for assessing development projects and going beyond reductionist approaches. Moreover, by capturing important aspects often ignored by traditional evaluation instruments, such an approach could also help to reform the institutional context of development policies and consequently orient funding allocation and project design more effectively.

Methodology and Results

The HD&CA is a social justice theory that focuses on the evaluation of human development in terms of the substantive opportunities people have to achieve doings and beings they have reason to value in several dimensions of their lives. Its constitutive multidimensional and complex character, which allows researchers to weigh the heterogeneity of both personal and social conversion factors as well as of the role of individual agency and empowerment in reaching achievements, and its vagueness, which transforms the evaluation of achievements from a binary issue into a matter of degree, make it particularly suitable for the assessment of contested subjects like development interventions (Clark, Biggeri, & Frediani, 2019; Garc´es-Velàstegui, 2022; Pham, 2018).

Accordingly, this approach offers several advantages: a) it provides a unifying lens through which to assess different case studies while highlighting their specificities; b) it enables giving priority to the quality of life of the worst-off by placing individuals and communities into their respective contexts, instead of focusing just on the maximization of average utility; c) it encompasses not only direct, short-term, tangible benefits, but also more long-term, indirect, and intangible ones; d) it considers the process through which outcomes are achieved rather than only outcomes themselves.

In light of these advantages, our paper takes inspiration from the HD&CA and introduces a multidimensional framework aimed at rethinking how development interventions are assessed. The framework in question consists of six dimensions:

1) the ownership of development projects and the partnerships that can guarantee their success;

2) the participation and level of engagement of local actors, either formal or informal, as well as the governance mechanisms implemented;

3) to what extent individual and community agency and empowerment have been activated by the projects;

4) what have been the tangible and intangible impacts generated in the short, medium, and long run;

5) what have been, if any, the cross-sectorial and cross-thematic spillover effects produced;

6) the environmental, social, and economic sustainability of the projects.

Moreover, the paper tests this framework by applying it to a specific case study: the assessment of three interventions that have been run in Kenya by AICS (the Italian Agency for Development Cooperation) concerning the important issues of maternal and child health and tuberculosis infection. The projects assessed are the SonoMobile, the HIPS-TB, and the ATI-TB projects. Each project is analyzed through a mixed methodology informed by the research framework, combining retrospective quantitative surveys and qualitative in-depth interviews conducted with a purposive-selected sample of the main projects' beneficiaries. In this sense, the analysis of our case study confirms how applying the HD&CA offers the

opportunity to develop not a deterministic method but a holistic way of thinking that can be adapted to different contexts and operationalized through various methods in compliance with the initial objectives of different development interventions. In this sense, it can help to better understand the strengths and weaknesses of the projects under analysis and con- tribute to orienting future funding allocations and project design.

Bibliography

Achen, C. H. 2020. RCTs versus observational research. In J. Widner, M. Woolcock & D. Ortega Nieto (eds.), The case for case studies: Methods and applications in international development: 52–60. Cambridge, UK: Cambridge University Press.

Clark, D. A., Biggeri, M., & Frediani, A. A. 2019. Participation, empowerment, and capabilities: Key lessons and future challenges. In Clark, D. A., Biggeri, M., & Frediani, A. A. (eds.), The capability approach, empowerment and participation. Rethinking international development series: 385–402. Cambridge, UK: Cambridge University Press.

Garc´es-Velàstegui, P. 2022. Using the capability approach and fuzzy set qualitative comparative analysis in development policy evaluation. Journal of Comparative Policy Analysis: Research and Practice, 24(2): 179–197.

Henderson, H. 2022. The moral foundations of impact evaluation. Journal of Human Development and Capabilities, 23(3): 425–454.

Pham, T. 2018. The capability approach and evaluation of community-driven development programs. Journal of Human Development and Capabilities, 19(2): 166–180.

Pritchett, L. 2020. Why ”feigned ignorance” is not good economics (or science generally). Lant Pritchett Personal Website, November 6. https://lantpritchett.org/why-feigned- ignorance-is-not-good-economics-or-science-generally/.

Keywords: development economics, project evaluation, randomized controlled trials, human development & capability approach, development policies, health

Why be a frame-sensitive reasoner? On the epistemic foundations of ultraintensional preference formation

James Grayot*[1]

[1]University of Porto, Institute of Philosophy – Portugal

Abstract

In *Frame it Again* (FIA), J.L. Bermúdez (2020) delivers a long-overdue analysis of the concept of framing, expertly covering the philosophical and scientific implications that the phenomenon has posed for the study of human reasoning, judgment, and decision-making (2020). Bermúdez's goal is clear from the outset: it's to clarify what is really involved in framing and to investigate whether there are benefits to being a 'frame-sensitive reasoner' such that even contentious cases of framing might be reconciled with certain mainstream conceptions of rationality, both those that give formal structure to the neat and contained worlds of microeconomics and decision theory as well as those that provide rules of conduct in uncertain social and political realities. His defense of frame-sensitive reasoning runs roughly as follows:
In contrast to classical perspectives about rationality derived from the axioms of rational choice theory in which decision-makers are viewed as utility maximizers with consistently ordered preferences (see, e.g., Savage 1961, Resnik 1987), we know that individuals are highly sensitive to how decisions are presented to them and that it is not uncommon for them to knowingly hold inconsistent preferences over their options and/or outcomes. Unlike cyclical preferences, which are paradigmatic of irrationality in the classical perspective, 'quasi- cyclical' preferences are those in which the decision-maker would not change or alter their preference orderings upon learning that two options/outcomes share the same extension, i.e., contain the same properties or attributes. More specifically, quasi-cyclical preferences are deemed *not irrational* if they are informed by an agent's taking an 'ultraintensional' perspective toward their options. Ultraintensionality is the representational context in which agents frame their options, placing particular emphasis or focus on some properties/attributes over others. By adopting an ultraintensional perspective towards their options, an individual may behave as though they have cyclical preferences when, in fact, the true values they associate with each option under framing would reveal their preferences to be non-cyclical.

Further, building upon the non-irrationality of quasi-cyclical preferences, Bermúdez also argues that there are many contexts in which one *ought to be* a frame-sensitive reasoner. In scenarios of great social or political complexity, where there is no unitary or 'objective' point of reference to establish a framing of a dilemma or debate, one should strive to con- sider as many (relevant) frames as possible before settling on a choice. The more flexible and openminded one is in considering different frames, the more likely one will be to reveal underlying emotions or opinions that may bias (positively or negatively) their judgments toward a decision, and thereby, to conscientiously compare and contrast the relevant features of the competing perspectives or viewpoints at issue. This is especially pertinent in cases of interpersonal decision-making, where a collective

*Speaker

decision may become ”deadlocked” due to irreconcilable values over some controversy. In short, the benefit of being a frame- sensitive reasoner is that it equips individuals with the ability to distinguish ”fact-based” from ”value-based” judgments, and thus, provides them with tools for achieving a discursive ”equilibrium”, i.e., an outcome that is practically rational from all (relevant) perspectives.

While this paper is sympathetic toward the view that frame-sensitive reasoning can be a reliable path toward rational decision-making, it remains skeptical about some of its theoretical underpinnings and justifications. In short, this paper asks: Why be a frame-sensitive reasoner? –specifically, what justifies a given framing toward a decision as rational? Starting from the assumption that Bermúdez is primarily interested in the *normative* virtues of framing, FIA can be seen as having two overlapping goals: the first is to make the case that quasi-cyclical preferences are at least not irrational; the second is to defend the prudence of frame-sensitive decision-making. While the two goals are clearly connected, they each require some unpacking.

Establishing the first goal depends on demonstrating that quasi-cyclical preferences are in- deed not irrational. Following Bermúdez, I take it to be a relatively straightforward matter whether quasi-cyclical preferences, as ultraintensionally-formed preferences, are not irrational (they are not! –However, what this means in terms of their *formal permissibility* needs further clarification). By contrast, and more importantly for this paper, establishing the second goal depends on demonstrating that ultraintensional preference formation is more than merely not irrational, but that its prudence is encapsulated by a form of *procedural rationality* that is amenable to both microeconomic and decision-theoretic modeling as well as to the demands of practical reasoning in a complex world. For this reason, a robust justification of ultraintensional preference formation must begin with an investigation into
(i) the epistemic foundations of ultraintensionality as a representational context, and (ii) the cognitive requirements of decision-makers as epistemic agents. The focus of this paper is about how that investigation should proceed, and in how far we can get toward a theory of frame-sensitive reasoning as procedurally rational *without* committing question-begging appeals to extant theories of rationality. I do this by couching ultraintensional preference formation in two broader extra-normative schools of thought. The first approach is that of naturalized rationality, specifically ecological and embodied approaches to rationality. The second approach is that of virtue epistemology, specifically the line of virtue epistemology that treats understanding as necessary for, if not more intellectually rigorous than, mere knowledge.

I will argue that the virtue epistemic approach provides a promising avenue toward an extra- normative justification for ultraintensional preference formation (and so, for frame-sensitive reasoning in general) by promoting the role of understanding in frame-based decision-making. Yet, I also show how this move opens the theory up to concerns that go beyond what is considered in FIA; in particular, it puts pressure on the theory to identify and commit to cognitive requirements for agents to have understanding. This turns out to be a difficult task given that what counts as understanding will likely differ across decision contexts, and this will obfuscate our search for what is cognitively required for agents to have understanding in each context. This leads to what I will call the ‘meta-framing’ problem: Whether or not a theory of frame-sensitive reasoning can be seen as procedurally rational is itself subject to framing what constitutes understanding across decision contexts. So, while both approaches provide potential extra-normative support for ultraintensional preference formation, neither approach on its own is singularly equipped to provide absolute justification. Rather, each approach seems to hint at gaps left unaddressed by the other. This dynamic, though it brings us closer to appreciating what a general theory of frame-sensitive reasoning needs to provide to satisfy Bermúdez’s second goal, the ‘meta-framing’ problem hints at more fundamental challenges to the study of human rationality.

Keywords: Framing, Frame sensitivity, Reasoning, Decision making, Preferences, Ultraintensional preference formation, Ecological rationality

Health, Death, and Well-Being: Toward an Intermediate Account of Well-Being in Terminally Ill Patients

Lior Nissim Grinman *[1]

[1]The Hebrew University of Jerusalem – Israel

Abstract

In behavioral welfare economics, present-biased behavior is often considered a cognitive error that leads individuals to make suboptimal decisions regarding their long-term well- being. To address this, behavioral economic interventions – such as nudges, precommitment strategies, and other behavioral tools – are employed to correct present-biased choices.

iIn this paper, I argue that present bias should not automatically be regarded as a suboptimal behavior that undermines well-being. I illustrate this through the case of terminally ill patients, showing that present-biased choices – such as opting for comfort care over aggressive treatment that could potentially prolong life – may genuinely reflect their true preferences.

The underlying intuition behind this claim is that, in cases of terminal illness, individuals may have a genuine preference for the present. Unlike broader health contexts – such as exercising regularly or maintaining a healthy diet – where consistently postponing long-term benefits may indicate a failure of rational decision-making, terminal illness presents a unique challenge: the future itself is uncertain and may not exist at all. In this context, prioritizing the present over an uncertain future may reflect a true preference rather than a cognitive error.

To ground this intuition, I examine whether a theory of well-being can justify the view that not correcting present bias is often the appropriate response in end-of-life care. To this end, I develop a theory of well-being for terminally ill patients. Following Anna Alexandrova's (2017) mid-theory approach to well-being, I integrate empirical constraints from well-being measurement tools in end-of-life care with philosophical theories of well-being. Specifically, I identify three key constraints that any plausible theory of well-being for terminally ill patients must satisfy:

- Multidimensionality: Well-being in terminal ill patients is assessed across multiple dimensions, including physical, emotional, social, and spiritual aspects.
- Patient-centered focus: Policies and interventions prioritize the patient's subjective experience, either directly through self-reports or indirectly through clinician observation tailored to patient needs. All metrics ultimately focus on improving patient-centered care, and assessing their subjective feeling.
- Adaptability to Context: Policies and interventions for terminal illness emphasize short- term adaptability. Given the dynamic and acute changes in a terminal patient's condition, well-being is assessed frequently and adjusted to align with evolving needs.

*Speaker

I begin by reviewing three major theories of well-being – 'the big three'– and argue that they are insufficient to account for these constraints. I then propose that a hybrid theory of well-being, akin to Susan Wolf's fitting-fulfillment view, offers a more suitable framework for understanding why present-biased decisions by terminally ill patients can be seen as im- proving their well-being.

To further develop this perspective, I propose extending Wolf's framework from a focus on 'Doings' to 'Beings.' In the context of terminal illness, traditional goal-directed activities ('Doings') may no longer be feasible, necessitating a shift toward the intrinsic value of activities such as 'being present,' cultivating connections and relationships, and finding existential meaning ('Beings'). This extended framework captures how terminally ill patients can do well, even when their choices are present-biased.

Beyond contributing to the development of an account of well-being for terminally ill patients – an important task in itself – this paper also makes significant contributions to behavioral welfare economics. First, it demonstrates that present-biased behavior can enhance well- being in certain contexts and, therefore, should not necessarily be considered detrimental. Second, if the life of terminally ill patient is going well for them, at least in part, due to present-biased choices, then behavioral interventions aimed at correcting present bias may be misguided in this context. Rather than nudging terminal patients toward specific treatments, such interventions should exercise greater caution in respecting their present-oriented preferences and avoid imposing behavioral corrections that prioritize long-term well-being at the expense of present well-being.

Keywords: rationality, behavioral welfare economics, well-being, mid-level theory, terminal illness, end of life care, present bias, meaning in life

The Origin of Money – a Methodologically Informed Appraisal of the Debate

Jakob Gschwandtner*[1]

[1]Johannes Kepler Universität – Austria

Abstract

Questions concerning money have taken centre stage in personal and social life as well as in political discourse for hundreds of years. In the theory of money, however, no agreement is in sight. One central point of contention is the origin of money. Here, two camps stand in opposition to each other: the Commodity Theorists and the Chartalists. Even though both accounts have been explicitly stated and debated for well over a hundred years now, no clear path towards a common understanding of money has emerged. Given its role in informing effective (monetary) policy, scientific progress in the field of money theory is clearly desirable. After briefly describing the two theories, we sketch how our methodological reflection serves that end. According to Carl Menger (1871, 1909), a major representative of the Commodity Theory, money emerged because of the requirements of barter and trade. In a pure barter economy, exchange only takes place if two individuals get together who desire precisely the goods which the respective other one has to offer. In a reaction to this problem of the double coincidence of wants, individuals start to accept goods which they might not need themselves but which they expect to be able to exchange for their desired goods. In a gradual process, driven by individual actions, a single good takes up the role of a universally accepted medium of exchange – it becomes money.

Georg Friedrich Knapp (1905), often referred to as the founder of the Chartalist approach, offers a different account. Money, he argues, is the result of actions by the state or other central authorities. By levying taxes and determining what kind of payments it accepts in settling them, the state establishes what counts as money. Crucially, this could be an already widely used commodity like cattle or gold as well as previously worthless tokens without other use.

In recent years, there have been some efforts to facilitate a dialogue between the two approaches: For instance, Aydinonat (2008) provides a rigorous analysis of Mengers methodology and proposes to understand his account as an abstract model as opposed to a historical description – this sheds new light on its compatibility with alternative accounts. Ikeda (2008) and Semanova (2014) discuss, whether Menger is accurately represented as taking a radically anti-statist position in his explication of the origin of money; Rallo (2020) argues that modern proponents of Chartalism have exaggerated Knapp's opposition to Mengerian explanations; and Rapka (2023) proposes to integrate supposedly Chartalist insights into a Mengerian framework. Taking those valuable contributions into account, we attempt to offer a systematic appraisal of the debate. We do so by paying special attention to the following, interconnected methodological issues:

*Speaker

Essentialism and diverging definitions: The first question is whether Commodity Theorists and Chartalists use the term ”money” in the same sense. Our approach is to look at the quite explicit definitions of ”money” which both Menger and Knapp offer and to analyse how they relate to each other. Taking this route requires us to reject strongly essentialist interpretations: Our task is not to gain insights about the true essence of money but to propose conceptualisations of money which allow for fruitful research.

Methodological principle – methodological individualism or historical analysis: Knapp, being closely connected to the German historical school, puts forwards a historical explanation of the development of money. Menger, on the other hand, is committed to methodological individualism and aims at explaining money as the unintended result of the interplay of intended actions. How strongly features the Methodenstreit in the debate on the origin of money?

Explanatory aim – how-actually, how-possibly: Adjacent to the question of methodological principle is the underlying explanatory aim. We explore the possibility of under- standing the Mengerian account as a how-possibly explanation. Treating it not as a description of the actual historical development but as a model, aimed at elucidating the underlying mechanism, might allow to align Mengerian and Chartalist insights. Here, we connect Aydinonat’s approach to recent publications on how-possibly explanations. (Castan˜on, 2021; Gru¨ne-Yanoff and Verreault-Julien, 2021)

The role of evidence: This might be especially important because current authors – most prominently Graeber (2011) – argue that anthropological and historical evidence of early credit money disproves the Mengerian origin story. However, as the previous considerations show, it is not trivial to determine the role which empirical evidence plays in deciding the debate. At the very least, some methodological and epistemological groundwork is required, in order to arrive at empirically testable hypotheses.

References:

Aydinonat, E. (2008). The Invisible Hand in Economics. How Economists explain unintended social consequences. New York: Routledge.

Castan˜on, G.A. (2021). Possibilist Explanation: Explaining How-Possibly Through Laws. Erkenntnis, 86, 835–852.

Graeber, D. (2011): Debt: The First 5.000 Years. Brooklyn: Melville House.

Gru¨ne-Yanoff, T./ Verreault-Julien, P. (2021): How-possibly explanations in economics: any- thing goes? Journal of Economic Methodology, 28(1), 114-123.

Ikeda, Y. (2008). Carl Menger’s monetary theory: A revisionist view. The European Journal of the History of Economic Thought, 15(3), 455–473. https://doi.org/10.1080/09672560802252347

Knapp, G.F. (1905): Staatliche Theorie des Geldes. Leipzig: Duncker & Humblot.

Menger, C. (1871). Grunds¨atze der Volkswirthschaftslehre, Vol. 1. Vienna: Wilhelm Braumüller.

Menger, C. (1909): Geld. Reprinted in: Hayek, A. (ed.): The Collected works of Carl Menger. Vol. IV. 1936. London: PERCY LUND, HUMPHRIES & CO. LT, 1-116.

Rallo, Juan Ram´on. (2020). Georg Friedrich Knapp was not a ”Chartialist”. History of Political Economy 52:4

Rapka, P. (2023). Debt and the origin of money. In: Sielska, A./Turowski, K./ Jasin´ski, L. (eds.): Mengerian Economics. Cheltenham: Edward Elgar, 87-105.

Semenova, A. (2012). Carl Menger’s theory of money’s origins: Responding to revisionism. The European Journal of the History of Economic Thought, 21(1), 107–141. https://doi.org/10.1080/09672567.2012

Keywords: Menger, Knapp, The Origin of Money, Money theory

Taxing Carbon, Framing Responsibility: Who Values the Environment?

Mathieu Guigourez*[1]

[1]Centre d économie de la Sorbonne – Université Paris 1 Pantéon-Sorbonne, Université Paris 1 - Panthéon-Sorbonne, Université Paris 1: Panthéon Sorbonne – France

Abstract

Putting a price on carbon is not something intuitive and can be justified in several ways. The Hotelling rule (Hotelling, 1931) is the first attempt to calculate the optimal depletion rate of a scarce resource by economists. The rule states that the optimal depletion rate or carbon emission rate in our case - shall reflect the real interest rate on the market. This rule has been used by prominent environmental economists (Nordhaus, 1993, 2017; Ulph and Ulph, 1994) to advocate for the implementation of an optimal carbon price that reflects the observed behaviours of economic agents on the market without making any normative claims on the agents' behaviours. A second possible way is to conceive carbon emissions as a market failure, i.e. a negative externality that the economist ought to internalize. In a Pigouvian framework (Pigou, 1920), it does not matter how much economic agent pollutes as long as that pollution is internalised through a tax or compensation system that enables a transition to a greener production system. Carbon taxes are, thus, usually presented as one the most 'rational' ways to deal with the climate crisis (Blanchard et al. (2023), for an overview of the literature on carbon taxes since the 1970's one can refer to Timilsina (2022).

The main assumption running through this literature is that, without the economist, the economic agent would barely lift a finger to do anything about climate change, as though his/her responsibility was not at stake. While the introduction of a carbon tax may be heavily rejected by the general opinion (Douenne and Fabre, 2022), other critics have argued that such a carbon tax may imply even the main assumption running through this literature is that, without the economist, the economic agent would barely lift a finger to do anything about climate change, as though his/her responsibility was not at stake. While the introduction of a carbon tax may be heavily rejected by the general opinion (Douenne and Fabre, 2022), other critics have argued that such a carbon tax may imply even more consumption of polluting goods (Ballet et al., 2005; Heal and Schlenker, 2019; Hartmann et al.,2023; Kaufmann et al., 2024). Carbon taxation has, then, been conceptualized in total opposition to individual responsibility (Bazin et al., 2004; Ballet et al., 2005, 2007). More specifically, we argue that the formulation of a carbon tax, even if it's not explicitly stated, always implies some acceptance of our individual responsibility for environmental issues, or the lack of it. Traditionally, the responsibility for correctly – and optimally – valuing the environment fell to the economist (Pigou, 1920; Hotelling, 1931), leaving aside the responsibility of the agent. As such, this paper proposes to delve into this literature on carbon pricing by examining carbon tax models that attempt to incorporate a degree of individual responsibility from the economic agent regarding climate issues.

*Speaker

The main ambition of this proposal is to examine and broaden our understanding of moral responsibility at stake in carbon tax models. The first model of environmental responsibility examined (Bazin et al., 2004; Ballet et al., 2005, 2007) conceives a threshold of taxation upon which the carbon tax has a crowding-out effect on our individual responsibility and, thus, becomes a ”licence to pollute.” The second model (Mattauch et al., 2022) examines a level of taxation that accounts for evolving preferences in regards to ecological issues. This model argues that, if not accounted for, those changes in behaviours lead to sub-optimal tax rates; as such, public policies should always account for the potential change in behaviours they may occur. The third model (Hartmann et al., 2023) argues that hidden carbon taxes are more efficient than salient ones at reducing our carbon footprint. They argue that salient carbon taxes may enhance our moral self-awareness – what they call ”moral licensing” – which lead to higher consumption levels of taxed goods. Bazin et al. (2004, p.131) frame such arguments with the following maxim: ”Since I pay, I am permitted to consume.” These three models point out the importance of considering how taxes can have a crowding-out effect and implicitly alleviate individual responsibility regarding environmental issues. Yet, there still lacks a deeper questioning of the responsibility at stake in their argument, which is precisely the gap this paper aims to fill. We can distinguish two competing ways of conceptualising individual moral responsibility involved here. On the one hand, the first way of conceiving a moral responsibility to curb our carbon footprint (which we frame as ‘ecological responsibility’) is to adapt the way we conceptualize preferences. One possible way to do so is via the endogenisation of preferences in regard to environmental issues and policies. Instead of taking the agent’s preferences to be exogenous (i.e. as a given), the economist should account for the adaptation of the agent’s consumption when faced with these ecological challenges. In other words, the endogenisation of preferences refers to cases in which 1) the economic agent adapts their consumption choices to reduce their individual impact or 2) that such consumptions unconsciously evolves after the introduction of a policy / tax. The literature dealing with the endogenisation of preferences is still budding (O’Hara and Stagl, 2002; Daube and Ulph, 2016; Dasgupta et al., 2016; Mattauch et al., 2022) but offers strong objections to standard carbon tax models. This first conception of endogenous change in the preference may be framed as ”accordant responsibility”. Whether our curb in carbon emissions is conscious or unconscious is not determinant for such a conceptualisation, as it only refers to acting according to the economist’s policy. Such a framework is compliant with green nudges policies (Schubert, 2017; Wensing et al., 2020; DesRoches et al., 2023). The second way of conceiving an ecological responsibility is to embed an ethical reflection in the decision-making process – hence a ”procedural responsibility” – leading an economic agent to autonomously reduce their individual ecological impact. In other words, such a procedural responsibility aims to look at critics of Rational Choice Theories in order to question and broaden the utilitarian framework upon which we conceive individual rationality. Basing this approach upon Sen’s work on commitment (Sen, 1973, 1977, 1985, 2005), I conceive this notion of responsibility as an imperative that would be rationally chose by an economic agent despite any incentive or political enforcement. If public policies only rely on what we call accordant responsibility, they may end up enforcing suboptimal tax rates (Mattauch et al., 2022) or enhance the consumption of polluting goods (Bazin et al., 2004; Hartmann et al., 2023; Kaufmann et al., 2024). A deeper understanding of individual responsibility is thus needed in order to avoid such traps. There is, therefore, a general discrepancy in the literature on carbon pricing which main aim is to optimally internalize an externality and that of individual responsibility which aims at autonomously internalizing an externality. The main objective of this proposal is twofold: 1) to analyse and compare alternative models of carbon taxes that account for the formation of preferences to reduce our own ecologic impact and 2) to question if this accounting of individual responsibility can stick with a carbon pricing Approach or if there needs a deeper understanding of what ecological responsibility stands for.

Keywords: Carbon Tax, Individual Responsibility, Endogenous Preferences, Moral Licensing

”Feminist Political Economy and the Migration Choices of Undocumented Latina Women”

Aubrial Harrington*[1]

[1]Arizona State University [Tempe] – United States

Abstract

Mainstream economic analyses often frame migration decisions within rational choice models, emphasizing wage differentials and labor market access as primary motivators. How- ever, such models inadequately address structural inequalities, gendered labor segmentation, and intersecting oppressions that profoundly influence the migration choices of undocumented Latina women. This paper employs a feminist political economy perspective to examine how global care economies, economic precarity, and neoliberal policies disproportionately impact these women's migration decisions. Rather than treating migration as an individual market choice, I argue that it must be understood as a gendered response to eco- nomic coercion, social reproduction crises, and the structural failures of neoliberal policies to provide economic security for women across borders. The dominant economic framing of migration as a rational labor market decision has become increasingly contested. Traditional economic models frequently overlook unpaid labor, social reproduction, and care work-domains predominantly occupied by women. This oversight leads to an incomplete understanding of labor dynamics and economic contributions. Traditional economic models often ignore the role of social provisioning, gendered labor segmentation, and the informal economy, all of which disproportionately shape the migration choices of women (Benerìa, Berik, and Floro 2016). Latina migrants enter feminized labor sectors- domestic work, caregiving, and service economies-where they face systematic exploitation, legal vulnerability, and an undervaluation of their labor (Ehrenreich and Hochschild 2002). Insights from feminist political economy is essential for understanding power relations and hierarchies that sustain contemporary capitalism. This perspective challenges the narrow focus of traditional economic models and underscores the necessity of incorporating a feminist lens to fully comprehend the complexities of labor and migration.

The global care economy significantly influences migration patterns, particularly among women from the Global South. Migrant women often fill domestic and care roles in wealthier nations, not simply due to wage disparities but as a direct response to structural inequalities that devalue women's labor and limit their economic autonomy in their home countries. Feminist political economy provides crucial insights into the reproduction of inequality through globalized labor relations, demonstrating how the division of care labor and social reproduction is central to understanding migration trends (Cantillon et al. 2023).

Care work is not only underpaid and undervalued but also deeply shaped by racial and gendered hierarchies that define who performs this labor. Latina migrants enter economies where their labor is essential yet invisible, often working in undocumented or precarious

*Speaker

conditions with limited protections (WILPF, 2023). The feminist critique of economic rationality challenges the notion that these labor patterns are freely chosen, instead arguing that they are dictated by historical economic structures and migration policies that reinforce dependency and exclusion.

Migration functions as both a strategy of survival and a form of agency within deeply unequal transnational economic structures. Feminist political economy highlights that migration is not simply about individual economic gain but about navigating global economic systems that are stacked against women, often leaving them with limited choices (Osirim, 2018). Latina migrants engage in informal labor markets that provide economic opportunities but also expose them to labor exploitation, workplace abuses, and precarious legal statuses.

Migration choices must be understood in relation to gendered economic precarity, insttutionalized exclusion, and the violence embedded in migration regimes themselves (Elias and Rai, 2019). This structural violence manifests not only in economic exclusion but also in state-enforced restrictions that push Latina migrants into highly surveilled and legally precarious forms of labor. In this context, agency is constrained, and Latina migrants must navigate a system designed to exploit their labor while denying them security.

State policies reinforce economic marginalization and labor segmentation, creating an environment where Latina migrants are both economically necessary and legally disposable. Certain feminist political economy works critique neoliberal economic policies for generating pervasive insecurity for migrant women, stripping them of labor protections while benefiting from their work. Latina migrants are positioned at the intersection of gendered labor extraction and restrictive immigration enforcement, where state policies simultaneously criminalize and exploit their labor.

These policies do not merely respond to economic conditions; they actively produce and sustain labor precarity by restricting pathways to citizenship, enforcing work authorization barriers, and expanding surveillance and criminalization of undocumented workers. Feminist political economy highlights the structural role of the state in shaping labor markets, particularly in how neoliberal states deregulate labor protections while increasing border enforcement (Cantillon et al, 2023). This contradiction creates a vulnerable workforce that is both economically essential and politically excluded.

Drawing on feminist critiques of economic rationality and commercial modernity, this pa- per situates the experiences of undocumented Latina migrants within broader discussions of economic marginalization. Feminist political economy today must address the ways in which Latina migrants are systematically positioned within global economic structures that extract their labor while denying them full social and political inclusion. A feminist political economy perspective does not simply critique these conditions-it offers a transformative framework for rethinking economic justice. By valuing social reproduction, care economies, and gendered labor contributions, feminist political economy provides a pathway for reimagining migration policies that prioritize human dignity over market efficiency. This requires not only policy interventions to protect migrant labor rights but also a broader reconsideration of economic systems that devalue women's work while depending on it.
By centering the lived experiences of undocumented Latina migrants, this paper challenges dominant economic frameworks that overlook the complexities of migration decisions. Incorporating feminist political economy perspectives reveals how structural inequalities, the undervaluation of care work, and neoliberal policies shape migration patterns. Understanding these dynamics is crucial for developing policies that address the root causes of migration and support the rights and well-being of migrant women. Rather than treating migration as an individual economic choice, we must recognize it as a deeply gendered response to structural coercion, economic exclusion, and a crisis of global social reproduction.

Keywords: Migration, Feminist Political Economy, Undocumented Latina migration, Economics of Immigration

The insuffciencies of game theory: complementary perspectives by Simone Weil

Aviad Heifetz*[1]

[1]The Open University of Israel – Israel

Abstract

December 14, 2024

In the world picture that game theory depicts,

1. Each individual has a set of available actions;
2. Each action profile of the individuals determines a (possibly stochastic) outcome;
3. Each individual is characterized by her stable preferences over outcomes;
4. Experiencing an outcome which is more preferred by the individual implies wellbeing for her.

This framework, though very general, has two important blind spots:

- Human beings' wellbeing does not depend solely on the experience of outcomes (even if very generally construed); and
- The link between (profiles of) actions and outcomes might be extremely unstable.

We will present potential remedies for these blind spots, as discussed by the French philosopher Simone Weil (1909-1943).

The first blind spot: consequentialism

In game theory experiencing an outcome determines one's wellbeing. This is a far-reaching generalization of the idea of food consumption that brings satiety. But an important aspect of human flourishing is precisely the lack of satiety, the ever-remaining sense of wonder and yearning that accompanies deeds aimed at satiation and gratification. 'Outcome' and 'yearning' are antonyms - the one finalized, achieved, the other never finalized and never achieved.

Had we cared only about consequence experience we would have delegated our choices to Nozick's 'experience machine':

*Speaker

Suppose there were an experience machine that would give you any experience you desired. Super-duper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. (. . .) Would you plug in? What else can matter to us, other than how our lives feel from the inside? (. . .)

First, we want to do certain things, and not just have the experience of doing them. (...)

A second reason for not plugging in is that we want to be a certain way, to be a certain sort of person. Someone floating in a tank is an indeterminate blob. There is no answer to the question of what a person is like who has long been in the tank. Is he courageous, kind, intelligent, witty, loving? It's not merely that it's difficult to tell; there's no way he is. Plugging into the machine is a kind of suicide. (. . .)

Thirdly, plugging into an experience machine limits us to a man-made reality, to a world no deeper or more important than that which people can construct. There is no actual contact with any deeper reality, though the experience of it can be simulated. Many persons desire to leave themselves open to such contact and to a plumbing of deeper significance. (. . .)

Perhaps what we desire is to live (an active verb) ourselves, in contact with reality. (And this, machines cannot do for us.) Without elaborating on the implications of this, which I believe connect surprisingly with issues about free will (. . .), we need merely note the intricacy of the question of what matters for people other than their experiences.[1]

When we are free, we continuously renounce the experience machine. This involves a continuous refusal concurrently with the pleasures themselves. For example, we do not devour our food, we savor it - unless we are starved to the extent that we lost our ability to refuse food; erotic experiences - according to thinkers from Plato to present-day psychologists - involve not only a Freudian drive discharge of gratification, but concurrently also sublime, undischarged and unfulfilled yearning. It is when we lose our capacity to refuse while consuming, we lose our liberty in one of the following two contrasting ways.

One way involves excessive consumption of pleasures of certain kinds, allowing for addiction, e.g. of various substances. Once addicted, we are not free any more to refrain from consuming. Becker and Murphy (1988)'s notion of rational addiction is thus a contradiction in terms. This is because rationality, in the sense of human reason, is about choosing freely again and again and again, whereas addiction is about permanently renouncing one's freedom to choose. The second way of losing our capacity to refuse while consuming involves refraining altogether from certain pleasures, by delegating our liberties to some sect guru or leader. That second way, in which we stop thinking for ourselves by ourselves, is the essence of fundamentalism. None of these two ways of losing our liberty, neither addiction nor fundamentalism, is compatible with genuine wellbeing.

To avoid both extremes, Simone Weil proposed that genuine progress in the realm of work hinges on a balance of challenges rather than on lush outcomes:

The only mode of production absolutely free would be that in which (...) the difficulties to be overcome would have to be so varied that it would never be possible to apply ready-made rules; not of course that the part played by acquired knowledge should be nil; but it is necessary that the worker should be obliged always to bear in mind the guiding principle behind the work in hand, so as to be able to apply it intelligently to ever-new sets of circumstances. Furthermore, it goes without saying that the degree of complexity of the difficulties to be solved must never be too great, on pain of bringing about a split between thought and action. (. . .)

To achieve this end it would be enough if man were no longer to aim at extending his knowledge and power indefinitely, but rather at establishing, both in his research and in his work, a certain balance between the mind and the object to which it is being applied.[2]

Weil's idea of progress as a balance of challenges is thus an alternative to the nonsustainable ideal of unbounded economic growth.

2 The second blind spot: the instability of the map from actions to outcomes

Whereas in game theory we consider (possibly stochastic) outcomes as completely determined by everybody's actions, in real life the way we read the dependence of outcomes on actions might, in fact, change abruptly and instantaneously:

If at night, on a deserted road, instead of a tree I think I see a man lying in wait, a threatening human presence forces itself on me, and (...) makes me tremble even before I know what it's all about; I draw near, and suddenly everything is different, I no longer tremble, I read a tree and not a man.[3]

This is still a happy example, where my reading of the outcome of advancing along the deserted road has stabilized: Once I identified the benign dead tree trunk, it will not transmute back in my mind to a threatening human presence. However, this is the exception rather than the rule. In many important circumstances, our reading of action outcomes continues to flicker back and forth:

If in times of war, or civil unrest, unarmed men are sometimes killed, it is because the sight of what is vile in these people and begs to be destroyed penetrates the souls of the armed men at the same time as the sight of the [unarmed men's] clothes, their hair, their faces (. . .)

If [in contrast] few crimes are committed during the ordinary course of life, it is because when a human being stands in front of us we read in the colours which enter through our eyes something which must to some degree be respected.

There is the same difference between these two states as between those of the man walking on the deserted road when he reads in an appearance first a man lying in wait, then a tree. (...)

The change takes place like a flash; each of the two readings appears, when it is there, to be the only one real or possible.[4]

If the outcome of the act of killing a neighbor keeps flickering in one's mind between peace time, civil war and peace again, stable preferences over outcomes do not necessarily induce stable preferences over actions or over action profiles. Thus, even if an individual has stable preferences over outcomes, as assumed in game theory, these preferences might still be poor predictors for her choices: to predict her choices, we would have to understand how she reads the outcomes of actions at the time of choosing.

There may therefore be an additional, possibly more fundamental stable characteristic of individuals, beyond their preferences over outcomes: character.

Character is an invariant that appears throughout the reactions of a human being (acts, gestures, attitudes, words), and, if it were itself defined, would define the possible reactions. (...) The reaction of a human being with respect to determinate circumstances, that force him to get involved in some manner or other, depends upon the circumstances and upon his character; the possibility of a choice is illusory. When we have to choose, the choice is in fact already made.[5]

Thus, like preferences in game theory, character determines choice, but at a stage much prior to the time of actual action. This is because character, as it has already evolved, determines how we would read outcomes. Simone Weil brings the following example:

For the sailor, the experienced captain, whose ship has in a sense become like an extension of his body, the ship is a tool for reading the storm, and he reads it quite differently than the passenger. Where the passenger reads chaos, unlimited danger, fear, the captain reads necessities, limited dangers, the means of escape from the storm, a duty to act courageously and honourably.[6]

The passenger's reading of the storm keeps flickering. In contrast, the long sailing apprenticeship the captain has gone through, that turned him into a master sailor, not only stabilizes his reading of the storm; it has also forged his character so that he cannot but do his best for the safety of his passengers.

Simone Weil then poses two open questions concerning character:

(...) our problem with respect to our character is double. First of all, how can we modify it ourselves, (...) and go beyond it in the direction of the highest value?[7]

That is, how to undo some of our passions and emotions, like fear, and incorporate them within higher-level sentiments, like courage? What kind of apprenticeship would be conductive to acquiring moral habits, in analogy with the apprenticeship of sailing a ship? Is the physical apprenticeship automatically conductive also towards the moral realm, as in the captain's example, or perhaps additional dimensions of apprenticeship are required?

Subsequently, Simone Weil poses her second open question concerning character:

how can we keep the empire of circumstances in certain cases from breaking it, and transforming it in such a way that we are exiled from ourselves?[8]

That is, how to bolster our character to resist the 'empire of circumstances' that threaten to break our character - torture, oppression, physical and emotional misery?

Simone Weil therefore poses here a constrained optimization problem, but where the 'budget constraint' pertains to our spirit rather than to matter.

References

Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. Journal of political Economy, 96(4), 675-700.

Nozick, R. (1974). Anarchy, State, and Utopia. Basic Books. Weil, S. (1958). Oppression and liberty. Routledge.

Weil, S. (1990). Essay on the notion of reading. Philosophical Investigations 13(4), 297-303.

Weil, S. (2015). Notes on the concept of character. In Late Philosophical Writings. University of Notre Dame Press.

[1]Nozick (1974).
[2]Weil (1958) (our emphasis).
[3]Weil (1990).
[4] Ibid.
[5]Weil (2015).
[6]Weil (1990).
[7]Weil (2015).
[8] Ibid.

Scientific objectivity and normative economics: the case of fairness theories

Conrad Heilmann*[1] and Stefan Wintein[1]

[1]University of Erasmus University Rotterdam – Netherlands

Abstract

Can normative economics be objective in any substantial sense of the word? On the face of it, that may seem an absurd question. Surely, normative economics investigates value-laden concepts and makes value-laden claims. At the same time, theories in normative economics can and should be analysed as part of the scientific endeavour that we call economics. In this article, we use accounts of scientific objectivity from philosophy of science to analyse in how far content and claims of normative economics can be argued to be objective. That is to say, we will investigate to what degree normative economic theory may be argued to cohere with different accounts of scientific objectivity. We analyse the case of "fairness theories" in philosophy and economics. These are theories that have conceptualised what fairness is and what it means to be fair. We analyse fairness theories and show that they can indeed be seen as somewhat objective – if we adopt the conception of objectivity as "anti-subjective". Fairness theories in philosophy and economics are objective in the sense that they are anti- subjective theories that make at least "positionally objective" claims.

Section preview:

1. Scientific objectivity as "anti-subjective"

We introduce three important accounts that conceptualize scientific objectivity as "anti- subjective". On these accounts, scientific objectivity is analysed with regards to the degree of (anti-)subjectivity inherent in the theory or science in question. For instance, the famous demand of the "view from nowhere" eschews any sort of subjectivity (Nagel, 1986). This view is aptly summarised by Fine (1998): 'In short the ideal of method requires us to be no-one in particular.' Rather less strict, Amartya Sen's (1993) conception of "positional objectivity" allows for position-relative but observer-invariant scientific claims. And in feminist standpoint theory, different perspectives of individuals are said to be epistemically beneficial, not because of their individuality, but because of the insights they afford (Wylie 2003).

2. The case of fairness theories

We analyze the case of fairness theories with the above accounts of scientific objectivity. Fairness is an important normative concept, central to theories in both philosophy and eco- nomics, as well as to moral theory and practice (Wintein and Heilmann, 2021). Interestingly, there are several parallels between different fairness theories in both disciplines. Both disciplines have developed, independently of each other, theories that see fairness as something established by using so-called "claims" (Broome 1990, Thomson 2019) or by using "preferences" to posit "envy-free" distributions. We show that the claims-based theories of fairness

*Speaker

strive to a ”view from nowhere” and minimally achieve ”positional objectivity”. We also show that the preference-based fairness accounts are less subjective than one may think, and can also be argued to achieve a weaker form of ”positional objectivity”.

3. Results

The central specific claim of our article is: Fairness theories in philosophy and economics are objective in the sense that they are anti-subjective theories that make positionally objective claims. We have established this claim by analysing different fairness theories and showing in detail how their content holds up to the different accounts of scientific objectivity as anti- subjective. We also show that our analysis affords a feminist critique of both preference-based and claims-based fairness theories: standpoint theory helps to point out that there is room and need for contingent (rather than reductive) fairness theories. More broadly, we show that it is fruitful to analyse normative theories in economics with conceptions of scientific objectivity: such analyses help appreciate the degree of objectivity inherent in them, and is thus a contribution to the critical and constructive assessment of such theories.

References

Broome, J (1990) Fairness. Proceedings of the Aristotelian Society 91: 87-101.

Fine, A (1998), The Viewpoint of No-One in Particular, Proceedings and Addresses of the American Philosophical Association 72(2): 7–20.

Nagel, T (1986) The View from Nowhere, OUP.

Sen, A (1993), Positional Objectivity, Philosophy and Public Affairs 22, No.2: 126–145.

Thomson, W (2019) How to Divide When There Isn’t Enough: From Aristotle, the Tal- mud, and Maimonides to the Axiomatics of Resource Allocation. Cambridge University Press.

Wintein, S & Heilmann, C (2021) Fairness and Fair Division. In: Heilmann, C. & J. Reiss (eds.) The Routledge Handbook of Philosophy of Economics. Routledge, 255-267.

Wylie, A (2003), Why Standpoint Matters, in Science and Other Cultures: Issues in Philosophies of Science and Technology, Figueroa, R and Harding, S (eds.), pp. 26–48.

Keywords: objectivity, fairness

Majority voting on nudges

Tobias Henschen*[1]

[1]University of Cologne – Germany

Abstract

In *The Community of Advantage*, Sugden directs two well-known criticisms against behavioral welfare economics. His first criticism states that (like neoclassical welfare economists) behavioral welfare economists assume a "view from nowhere": that the analyst of (neo- classical and) behavioral welfare economics is an "impartially benevolent spectator" who addresses a "benevolent autocrat" (Sugden 2018: 18-22). Sugden (2018: 14) criticizes this assumption because it runs counter to the type of contractarianism that he thinks should organize society, and that envisages that policies or institutions "are assessed from the several viewpoints of individual members of society", and not "from the standpoint of a benevolent social planner". Sugden's second criticism says that behavioral welfare economics assumes the existence of an "inner rational agent" with rational (i.e. complete and transitive) and context-independent preferences, and that this assumption lacks any psychological foundation.

In the debate initiated by these two criticisms, several authors try to defend their work or field. Bernheim (2021: 13), for instance, argues against the first criticism that both neoclassical and behavioral welfare economics "admit contractarian interpretations", and that there is "nothing peculiar or unnatural about the 'social planner' perspective". He also argues that the second criticism applies to an approach that he refers to as "BRP" or "behavioral revealed preference", and that it leaves the generalized framework of Bernheim and Rangel unscathed (Bernheim 2021: section 4). In his response to Bernheim, Sugden (2021: 423) concedes that neoclassical and behavioral welfare economics admit contractarian interpretations, and that the generalized framework doesn't depend on the assumption of the existence of an inner rational agent. Sugden (2021: 422-3) also insists, however, that behavioral welfare economics becomes problematic when "the recommended methods of policy appraisal are designed to correct citizens' errors", and that the methods of the generalized framework continue to invoke "processes of latent reasoning" and "latent preferences".

The debate of Sugden's two criticisms is already three or four years old. In this paper, I will nonetheless revisit this debate because there is at least one problem that remains implicit in the debate and merits being made more explicit. This problem becomes visible once we realize that in behavioral welfare economics, token-level relations between individual choices and welfare levels need to be understood as causal, and no longer as constitutive, as in neoclassical welfare economics. If the token-level relations between individual choices and welfare levels are causal, evidence of these relations is evidence of the truth of counterfactuals

*Speaker

like 'If individual *i* chose *x* and her choice is made in the absence of mistakes, her level of welfare would increase'. These counterfactuals make reference to preferences or reasoning processes that are latent in the sense of 'counterfactual'. Methodologically speaking, how- ever, there is nothing wrong with counterfactual entities. Natural and social sciences refer to such entities on a regular basis. Sugden's insistence on the applicability of his second criticism to the generalized framework seems accordingly a bit exaggerated.

The real problem affecting the generalized framework is a problem of weak evidence. In the generalized framework, evidence of relations of token-level causation between individual choices and welfare levels is interpreted in terms of evidence of the presence or absence of mistakes. Mistakes are unobservable because they are defined in relation to individual welfare levels. Bernheim therefore proposes to approximate evidence of the absence or presence of mistakes by registering ex post statements that testify to the presence of mistakes, and by providing evidence of the absence or presence of what he calls "incorrectly informed indirect judgments" (Bernheim 2016: 48). But choices that are made in the presence of incorrectly informed indirect judgments are not necessarily mistaken, and ex post statements are difficult to register and neither verifiable nor especially trustworthy. As a result, there are likely to be many borderline cases, and the false positive rate of identifying mistakes is likely to be high among paternalists and low among anti-paternalists.

How is the problem of weak evidence to be solved? In this paper, I investigate one possible solution that involves a continued search for evidence of context-independent preferences. Bernheim (2016: 20-21, 2021: 3) and Sugden (2018: 11) share the belief that context-independent preferences don't exist. But they also seem to think that context- independent preferences are necessarily those of an inner rational agent, and the question is whether context-independent preferences necessarily lack any psychological foundations. I hypothesize that there are psychological foundations to context-independent preferences when they are first-order preferences for which individuals form (context-dependent) second- order preferences. I also argue that these first-order preferences should be prioritized over context-dependent first-order preferences if the second-order preferences are well-informed. And I elaborate on a possible case, in which individuals are likely to form well-informed second-order preferences for context-independent first-order preferences: on the case where they vote for or against the implementation of a nudge. This case is of particular interest to theorists who believe that there is no view from nowhere that (behavioral) welfare economists or contractarian theorists could adopt. Unlike this view from nowhere, simple majority decisions (for or against the implementation of the nudge) seem to be true to the letter of the type of contractarianism that Sugden endorses.

References

Bernheim, B. D. 2016. The good, the bad, and the ugly: A unified approach to behavioral welfare economics. *Journal of Benefit-Cost Analysis* 7(1): 12-68.

Bernheim, B. D. 2021. In defense of behavioral welfare economics. *Journal of Economic Methodology* 28(4): 385-400.

Sugden, R. 2018. *The community of advantage: A behavioral economist's defense of the market*. Oxford: OUP.

Sugden, R. 2021. A response to six comments on The Community of Advantage. *Journal of Economic Methodology*28(4): 419-430.

Keywords: benevolent autocrat, inner rational agent, behavioral revealed preference, causal evidence, majority voting, nudges

Construct validity and mechanistic extrapolation in the psychology of poverty

Lari Hokkanen*†[1]

[1]Helsingin yliopisto = Helsingfors universitet = University of Helsinki – Finland

Abstract

Within behavioural public policy, the psychology of poverty seeks to produce generalizable, mechanistic, and policy-relevant knowledge, and has already made seminal contributions to the understanding of poverty from a behavioural and experimental perspective. This paper explores what consequences there are for the mechanistic extrapolation and generalization of heterogeneous causal effects, if one evaluates the paradigmatic experimental designs with the concept of construct validity. With the concept I emphasize its theoretical side of connecting high-order constructs with empirical operationalizations and inspecting experimental interventions as imperfect operationalizations. I argue that the field's goal of producing robust, policy-relevant causal generalizations is at odds with the current practices of construct validation, particularly regarding the interventions and mechanisms under study. The practices and their downstream consequences to experimental design and resulting causal inferences do not easily allow generalizations and extrapolations made. This argument has wide-reaching implications for debates surrounding the heterogeneity of findings in behavioural developmental economics and public policy. The psychology of poverty is an illustrative case to discuss problems related to heterogeneity, as it has always been "beyond nudging" and is interested in causal mechanisms across contexts.

Building on a behavioural economics perspective, the psychology of poverty began with two seminal experimental studies a decade ago that produced bold causal generalizations about the negative psychological effects of poverty and resource scarcity. Since then field has expanded significantly, and the paradigmatic experimental interventions include priming, payday, harvest, and laboratory strategies (with some field experiments). Now, the main construct, scarcity, is operationalized flexibly and is often linked to "universal" cognitive mechanisms that apply across contexts. However, when flexible operationalizations are left implicit, challenges arise. Cognitive mechanisms may vary or interact with other kinds of mechanisms across contexts. The heterogeneity of scarcity can produce varying effects depending on duration and intensity-as could be in different welfare-state settings, as some evidence points out-which broad generalizations may fail to capture. Moreover, the nature of the experimental interventions (what they changed and what they did not) along with their relations to the higher-order construct of scarcity, is not always clear. Some field experiments operationalize scarcity in a way that result does not give evidence for or against the scarcity theory: for example, a quasi-experimental payday study might study the main effects of short time-periods between payments, which is not testing the theory, if the time-period is too short to have theorized cognitive effects. Tendency to overlook construct validity leads to an incomplete understanding of how interventions, assumed mechanisms, and outcomes relate to the various behaviours and (choice) environments of the poor, thus hindering building generalizations. Lack of construct validity can also partly explain the mixed evidence-base of the field.

*Speaker

† Corresponding author: lari.hokkanen@helsinki.fi

The psychology of poverty holds great value, with experiments remaining central in the future too. The field has shown that poverty has cognitive effects through causal mechanisms, even though work remains to be done on the specifics. To get further, instead of applying the paradigmatic cognitive mechanism schema by default to new studies, researchers could con- sider more how empirical operationalizations reflect the phenomenon with its social-cultural aspects, treating interventions as local and context-dependent, and embracing-sometimes messy-mixed-methods triangulation to untangle how interventions are actually implemented and how mechanisms seem to work. The incoherent results also showcase the importance of explicating the phenomenon, its model and how that model is operationalized, as single studies can study only part of the phenomenon. Many of the outlined issues lack a straightforward statistical solution and instead demand a more rigorous theoretical approach, as emphasized by many psychometrics classics. Moreover, insights from the interventionist and mechanistic approaches to explanation can help balance the trade-off between granularity and generality in causal claims and thus recalibrate the aim of the field more realistic: generalizations are sensitive to background factors and causal relations are variable-relative, which implicates careful approach to constructing operationalized variables in experimental research, as for example, different indicators of scarcity (subjective, objective, relative) and social economic status can yield different results. Collaboration between behavioural researchers and poverty scholars could further help by developing policy-relevant constructs of poverty and scarcity and impose constrains to possible space of mechanisms, even though there is no need nor possibilities for a ”grand theory” of poverty. To end, I argue that issues related the field are of interested to all behavioural approaches to development and public policy. To complement recent arguments for the ”heterogeneity revolution” in applied behavioural science, that is, focusing on heterogenous treatment effects and context, issues and improved processes of construct validity are of great importance.

Keywords: construct validity, mechanisms, behavioural public policy

Thick Concept, Thin Consensus: Normative Judgments in Estimating Inequality of Opportunity

Moritz Hörl*[1]

[1]Wirtschaftsuniversität Wien – Austria

Abstract

Social science indicators play a crucial role in contributing to societal progress by serving two primary purposes: informing the public and effectively guiding policy decisions (Thoma, 2024). This paper argues that current efforts by economists to empirically estimate inequality of opportunity (IOp) fall short of fulfilling either of these objectives.

Rooted in discussions about fairness and responsibility in philosophy, the IOp approach emerged as an effort to combine normative principles with empirical analysis. Over the past decades, it has gained increasing prominence in the field of economics (Roemer and Trannoy 2015). The approach adheres to a central principle: inequalities attributed to circumstances are considered as unfair since they are beyond the control of an individual whereas inequalities arising from varying effort levels are deemed acceptable. According to this principle, individuals should be held accountable only for their effort and choices, while compensated for inequalities due to circumstances. However, distinguishing between effort and circum- stances requires normative assumptions prior to empirically analyzing IOp. In addition to classifying variables as either circumstances or effort, a key consideration involves addressing the effect of circumstances on effort. Specifically, researchers must decide if people should be held accountable for the influence of circumstances on the level of effort they exert in life. While this issue was brought up in the initial debates surrounding IOp (Jusot et al. 2013), the recent literature has shifted focus to statistical developments, such as the incorporation of machine learning methods into the IOp framework.

This paper revisits these foundational debates, arguing that current empirical practices in IOp research overlook the significance of critical normative assumptions. This, I contend, results in a situation of distributive epistemic injustice and undermines the ability to effectively use IOp estimates for policy guidance. Specifically, I show that researchers solely adhere to Roemer's (1998) operationalization -albeit for different reasons- of not holding individuals accountable for the effects of circumstances on effort. As a result, IOp estimates are only estimated within this specific normative framework, limiting their relevance to those who agree with Roemer's perspective. While Roemer (1998) argues for compensating individuals for the effects of circumstances on effort, Barry (2005), drawing on Dworkin's (1981) perspective, contends that individuals should be held accountable for their observable effort, regardless of how circumstances may have influenced it. Thus, differences in effort levels due to varying circumstance variables should not be accounted for. Instead, Barry (2005) contends that only the actual observable effort of people should be considered. The decision is not only important from a moral standpoint but also has substantial implications for empirical estimates of IOp. Ramos and Van de Gaer (2016) report that most findings suggest between 20% and 40% of the overall effect of circumstances can be attributed to the effect of circumstances on effort.

*Speaker

Focusing exclusively on Roemer's (1998) approach undermines the capacity of IOp estimates to inform all members of the public, as alternative normative viewpoints are not considered when estimating IOp. While this primarily poses an information problem that could be addressed by estimating multiple IOp indicators based on different normative assumptions, using social science indicators like IOp to guide policy requires additional legitimization procedures. It is not only impractical to base evidence-based policy on multiple divergent IOp indicators, but it is also essential for people to have reasons to trust and accept that guiding policy with a specific IOp operationalization is warranted. I then introduce two alternatives proposed in the literature to address this issue.

The democratization approach seeks to democratize value judgments in empirical research (Schroeder 2021), while the stakeholder approach aims to leverage insights from various stakeholders with different expertise (Alexandrova and Fabian 2022). I consider these proposals in more detail within the context of the IOp framework. By drawing on recent insights from behavioral economics (Andre 2024, Dong et. al 2022, Fehr et. al 2024), I argue that democratizing the value judgments in IOp analysis rejects the current empirical operationalization, resulting in overestimated levels of IOp. This is the case since most people indeed hold others responsible for the effect of circumstances on effort.

In order to compare the democratization and stakeholder approach, an expert survey was conducted to elicit the normative values of experts regarding this issue. Experts are often considered to have special normative knowledge given their expertise (Alexandrova 2017) and thus, represent an important stakeholder group alongside the general public. The sur- vey results indicate that experts have different fairness preferences concerning the effect of circumstances on effort. Therefore, following the stakeholder approach would necessitate additional procedures, such as deliberative discussions that include both experts and members of the general public, to more appropriately handle normative assumptions in the IOp framework. Given that none of these alternatives is currently employed to provide political legitimization for the existing IOp operationalisation, current estimates are insufficiently justified to guide policy decisions.

To make inequality of opportunity (IOp) estimates more inclusive, researchers should present multiple estimates based on distinct normative frameworks, particularly those of Roemer (1998) and Barry (2005), as both offer theoretically defensible perspectives. Providing the public with reasonable alternative frameworks fosters greater distributive epistemic equality. Beyond offering multiple estimates, legitimate policy guidance requires clear justification for selecting one operationalization over another. This can be achieved through democratization or stakeholder-based approaches, which determine which normative framework is most appropriate for informing policy decisions. Recent attempts to sidestep normative debates in the IOp literature by focusing on "inherited inequality" (Ferreira and Brunori, 2024) do not resolve these issues, as they retain the same methodological foundations and underlying value judgments as traditional IOp measures.

Instead of trying to eliminate normative assumptions, IOp researchers should acknowledge and incorporate them into their analyses. Embracing these inherent value judgments ensures that IOp estimates are credible, inform public discourse, and effectively guide policy decisions.

Works cited

Andre, P. (2024). Shallow meritocracy. Review of Economic Studies, 1–36.

Alexandrova, A. (2017). A Philosophy for the Science of Well-Being. Oxford University Press.

Alexandrova, A. and Fabian, M. (2022). Democratising measurement: or why thick concepts call for coproduction. European Journal for Philosophy of Science, 12(1):7.

Barry, B. (2005). Why Social Justice matters. Polity.

Dong, L., Huang, L., and Lien, J. W. (2022). They never had a chance: Unequal opportunities and fair redistributions. Technical report, CeDEx Discussion Paper Series.

Fehr, D., Müller, D., and Preuss, M. (2024). Social mobility perceptions and inequality acceptance. Journal of Economic Behavior & Organization, 221:366–384.

Ferreira, F. H. and Brunori, P. (2024). Inherited Inequality, Meritocracy, and the Purpose of Economic Growth. III Working Paper, 147.

Jusot, F., Tubeuf, S., and Trannoy, A. (2013). Circumstances and efforts: How important is their correlation for the measurement of inequality of opportunity in health? Health economics, 22(12):1470–1495.

Ramos, X. and Van de Gaer, D. (2016). Approaches to inequality of opportunity: Principles, measures and evidence. Journal of Economic Surveys, 30(5):855–883.

Roemer, J. (1998). Equality of Opportunity. Harvard University Press.

Roemer, J. and Trannoy, A. (2015). Equality of Opportunity. In Handbook of Income Distribution. Elsevier.

Schroeder, A. (2021). Democratic values: A better foundation for public trust in science. The British Journal for the Philosophy of Science, 72(2):545–562.

Thoma, J. (2024). Social science, policy and democracy. Philosophy & Public Affairs, 52(1):5–41.

Keywords: Inequality of Opportunity, Thick Concepts, Normative Judgments, Econometric Modeling, Distributive Epistemic Injustice

Anti-Reductionist Strategies for Macroeconomists

Paul Hoyningen-Huene*[1,2]

[1]Leibniz University Hannover – Germany
[2]University of Zurich – Switzerland

Abstract

Discussions about reductionism in the sciences tend to be highly controversial, and eco- nomics is no exception. The purpose of this paper is not to engage directly in this debate but rather to clarify it from a meta-level perspective by analyzing the possible and impossible argumentative moves within this discourse. I focus on the more challenging position-defending the autonomy of macroeconomics and arguing against its reduction to microeconomics (see, e.g., Hoover 2008, 2009, 2010, 2015). This position is inherently more difficult to support than its opposite because it involves an impossibility claim: namely, that any attempt to reduce macroeconomics to microeconomics will ultimately fail. Arguments for impossibility claims are particularly demanding, as they require considering and refuting an infinite number of possible reductions.
A crucial first step in this argument is to clarify the precise nature of the reduction in question. There are at least five principal forms: ontological, epistemic, explanatory, prognostic, and methodological reduction. The next step is to examine the presuppositions underlying the distinction between micro- and macroeconomics. It turns out that one must assume ontological reduction; otherwise, the question of reducibility in other senses becomes trivially answerable in the negative. Most economists accept ontological reduction since macroeconomic phenomena introduce no fundamentally new entities or interactions beyond those at the micro level. Therefore, I take ontological reduction as a given.

Now, let us assume we are concerned with epistemological reduction. The anti-reductionist macroeconomist seeks to argue that knowledge at the macro level enjoys a degree of autonomy from knowledge at the micro level. This means that macroeconomic knowledge cannot be fully derived from microeconomic knowledge, even if the latter were complete. This claim challenges what I call the fundamental reductionist intuition: if a system is composed of smaller parts, and if those parts, their composition, and their interactions are known, then the behavior of the system as a whole must be derivable from knowledge of its components. This intuition is driven by the assumption that the macro level has no additional degrees of freedom beyond those present at the micro level. In other words, all macro-level changes must originate from micro-level changes, and there can be no independent macro-level variation. Technically speaking, the macro level is said to supervene on the micro level. The key question, then, is how the anti-reductionist can argue against this fundamental intuition.

Three primary anti-reductionist strategies emerge:

Conceptual Incommensurability: This argument asserts that the macro and micro levels are qualitatively distinct in ways that prevent conceptual bridging, making derivation of macro-level knowledge from micro-

*Speaker

level knowledge impossible. This position appears in various sciences, including economics. For instance, Hoover (2015) argues that macroeconomic laws operate through ”billiard ball” causality, whereas the micro level is characterized by individual intentionality. Since causality and intentionality belong to fundamentally different conceptual domains, he contends that macroeconomic causal laws cannot be reduced to micro-level intentional states. However, this argument is risky. In other fields, seemingly unbridgeable conceptual gaps have been overcome-consider the relationship between thermodynamics and mechanics. While heat and mechanical motion initially appeared irreconcilable, statistical mechanics established conceptual links between them. This suggests that even stark phenomenological differences may not preclude conceptual integration.

Interdependence Between Levels: A second strategy challenges the assumption that the macro and micro levels are clearly separable-a necessary condition for epistemological reduction. The idea is that for macroeconomic knowledge to be reducible to microeconomic knowledge, the micro-level state of a system must be determined independently of any macro-level states. However, as Hoover (2008, 2015) argues, this condition does not hold in economics. Individual choices at the micro level are influenced by agents' expectations and knowledge of macroeconomic variables, implying that the micro level depends on the macro level. This argument, however, has a potential weakness: it is not macro-level phenomena per se that influence the micro level, but rather micro-level agents' beliefs about macroeconomic variables. Since beliefs reside at the micro level, the supposed macro-to-micro dependence might be illusory.

Pragmatic Uselessness of Reduction: A third anti-reductionist strategy holds that macro- to-micro reduction is not merely impossible but also unnecessary. A key inspiration for this argument comes from Hilary Putnam's example of a solid square peg passing through a solid round hole (Hoover 2015). This problem can be fully analyzed using geometry, without reference to the molecular composition of the peg or the hole. Although many molecular configurations can result in solidity, these details are irrelevant to determining whether the peg fits. According to Hoover, macroeconomics exhibits a similar structure: many different micro-level realizations may underlie a given macroeconomic law, yet understanding these micro-level details does not enhance comprehension of the macro-level regularity. Consequently, reductionist attempts become meaningless. However, one might question whether macroeconomic laws are as robust as the peg-hole example suggests. Macroeconomic laws are far less secure, and the conditions of their validity are often only loosely understood. In this light, micro-level insights might actually contribute to improving macroeconomic understanding, casting doubt on the applicability of this third anti-reductionist strategy to economics.

This admittedly essayistic and anecdotal discussion highlights that, as in many other sciences, anti-reductionist arguments in economics are intricate and elusive.

References:

Hoover, Kevin D. ”Does Macroeconomics Need Microfoundations?” In The Philosophy of Economics: An Anthology. 3rd Edition, edited by Daniel M. Hausman, 315-33. Cambridge: Cambridge University Press, 2008 (2001).

Hoover, Kevin D. ”Microfoundations and the Ontology of Macroeconomics.” In The Oxford Handbook of Philosophy of Economics, edited by Harold Kincaid and Don Ross, 386-409. Oxford: Oxford University Press, 2009.

Hoover, Kevin D. ”Idealizing Reduction: The Microfoundations of Macroeconomics.” Erkenntnis 73 (2010): 329-47.

Hoover, Kevin D. ”Reductionism in Economics: Intentionality and Eschatological Justification in the Microfoundations of Macroeconomics.” Philosophy of Science 82, no. 4 (2015): 689-711.

Keywords: Autonomy of macroeconomics, ontological reduction, epistemic reduction, antireductionist strategies

Between economics and philosophy: a reappraisal of the Rawls-Harsanyi debate

Herrade Igersheim*[1] and Juan Carvajalino*

[1]Bureau d'économie théorique et appliquée (BETA) – université de Strasbourg, CNRS : UMR7522, Université Nancy II – P E G E 61 avenue de la For^et Noire 67085 STRASBOURG CEDEX, France

Abstract

On the one hand, John Harsanyi, both economist and philosopher, was born in Budapest in 1920, obtained a PhD in philosophy in the same city in 1947 and moved to Australia in 1950. In 1956, Harsanyi switched to economics and obtained a PhD at Stanford, continuing his academic career at Berkeley from 1964. In 1994, John Harsanyi was awarded the Nobel Prize in Economics, together with Reinhard Selten and John Nash, for their pioneering work on equilibrium in non-cooperative game theory.
On the other hand, John Rawls, the most famous political philosopher of the 20th century, was born in Baltimore in 1921 and studied philosophy at Princeton, where he obtained his doctorate in 1950. Rawls went on to teach at prestigious institutions such as Oxford, Cornell and MIT, before moving to Harvard in 1962. In 1971, John Rawls published his major work, *A Theory of Justice*, which marked a real turning point for political and moral philosophy. Even if Harsanyian and Rawlsian theories demonstrate a similar initial ambition, i.e. to found a rule of justice on the basis of rational choice theory, the debate that pitted the two authors against each other over the choice of decision criterion under uncertainty in the mid- 1970s was nonetheless very virulent. This debate is well known and has been addressed by secondary literature both from the side of the history of economic thought (Duhamel (2006 and 2012), Dumitru (2006), Kandil (2001 and 2014), Hawi (2016), Igersheim (2022)) and
by social choice theorists (Diamond (1967), Sen (1974), Mongin (1994), Weymark (2005), Fleurbay, Salles and Weymark (2008)).

The aim of our article is to examine the relationship between Rawls and Harsanyi in the light of the archives (Rawls - Harvard and Harsanyi - Berkeley). Beyond the above-mentioned debate between our two authors, we show that a more nuanced and longer narrative is emerging. This can be divided into 3 main moments: 1) the meeting in 1964 at the 2nd conference of the Committee for Non-Market Decision Making; 2) the Rawls-Harsanyi analytical de- bate in 1973/1975; 3) the development of Harsanyi's more assertive philosophical critique of Rawlsian theory from the 1980s onwards.

Thus, revisiting the Rawls-Harsanyi debate in the light of the archives enables us to better understand the nuances and philosophical implications of each thinker's positions. Far from the literature's focus on the 1973-1975 analytical debate, an examination of the archives of these two authors teaches us that Harsanyi's criticism of Rawls spans more than 30 years and addresses various fundamental aspects of their work. While it is true that a central aspect of the (narrow) debate between Rawls and Harsanyi concerns the opposition between the

*Speaker

maximin principle and that of maximizing expected utility, a broader philosophical debate emerges from Harsanyi's pen, raising profound ethical and moral questions as his theory of morality takes shape. From the mid-1980s onwards, Harsanyi's philosophical critique took place on two levels: on the one hand, his critique of the social contract called into question the very foundations of the social structure proposed by Rawls. Secondly, Harsanyi is highly critical of Rawls's non-compatibilist position, which, in his view, denies the effort that individuals can make to develop their talents for the benefit of society. These last two points led Harsanyi to say the following in a 1996 interview (d'Aspremont and Hammond, 2001): " Cl. d'Aspremont : Recently it has been suggested that John Rawls would justify his difference principle by the claim that, in the original position, people would be extremely risk averse and so maximize the minimum possible utility level they could experience. If Rawls were to abandon this hypothesis, and propound a theory based instead on expected utility, would most of your disagreements be resolved? J. Harsanyi: No, I don't think so... And I much more, much more object to his basic moral views than to decision-theoretical views".

Keywords: Rawls, Harsanyi, maximin

Rethinking Adam Smith's Philosophy: A New Foundation for Political Economy

Ibtissem Issaoui*[1]

[1]University of Carthage, LEGI-Tunisia Polytechnic School FSEG Nabeul – Tunisia

Abstract

In recent decades, Adam Smith's moral philosophy has undergone a significant revival, largely driven by efforts to address key societal issues such as the ethical dimensions of capitalism, and the increasing need for interdisciplinary approaches to human behavior and societal organization. This renewed scholarly attention to The Theory of Moral Sentiments (TMS) has produced fresh insights into Smith's complex theory of action, with key contributions from scholars such as Hanley (2009, 2019), Sagar (2017), and Sayre-McCord (2013). These studies have firmly established Smith's work as a critical juncture between moral philosophy, economics, and sociology, revitalizing the study of his thought.

However, despite these advances, certain aspects of Smith's ethical and economic theories remain underexplored, particularly the nuanced interplay between his moral philosophy and economic action. While much has been written about Smith's ideas in relation to social interaction and the role of the "impartial spectator" in moral judgment, this paper offers a different perspective: it focuses on the internal, individual process through which Smith's theory leads to self-discovery. Instead of centering on the social construction of the self, this paper interprets Smith's philosophy as a modraw for understanding how individuals discern the varied dimensions of their pleasure and pain-their varying intensities and values-through a dynamic, three-stage process.

Methodology

This study employs a close textual analysis of Adam Smith's The Theory of Moral Sentiments (TMS), systematically examining how Smith conceptualizes self-awareness and self-love. By carefully analyzing key passages, this paper reconstructs Smith's framework of self-discovery, emphasizing the three-stage process through which individuals refine their understanding of pleasure and pain. This interpretive approach allows for a nuanced reading of Smith's philosophy, highlighting its relevance to contemporary discussions in moral psychology and political economy.

*Speaker

Theoretical Framework

The originality of this approach lies in the reinterpretation of the self not as an entity shaped by external societal standards but as one that emerges through internal reflection and experience, encompassing a process of self-assessment, moral introspection, and the integration of life events and personal values. This includes the individual's capacity to critically evaluate their own emotions, motivations, and moral judgments, ultimately fostering a deeper and more nuanced understanding of their identity and ethical compass. This process of self-discovery, rooted in Smith's stages of moral development, allows individuals to uncover their true self, including the true value of their pleasures and pains, which are often more complex and nuanced than they might appear initially. By engaging with Smith's stages of self-awareness, individuals refine their ability to make decisions that align with their self-love, as defined by their deep understanding of their internal experiences.

This study proposes a redefinition of self-awareness within Smith's framework, focusing on the following three stages:

The First Stage: Individuals independently analyze their emotions, free from societal in- fluence, forming an initial understanding of their experiences of pleasure and pain.

The Second Stage: As individuals engage with society, their self-awareness deepens through interaction and comparison with collective standards, encapsulated in Smith's concept of the impartial spectator. This stage underscores the role of sympathy and shared experiences in refining their evaluation of emotions.

The Third Stage: The highest form of self-awareness is attained when individuals transcend societal norms, achieving a nuanced and autonomous understanding of their emotions and actions.

By framing Smith's work in this way, this paper adds a new layer to the literature by highlighting the process through which individuals' understanding of pain and pleasure can evolve and become more refined, thereby challenging existing theories in Smith scholarship by introducing a dynamic perspective on moral and emotional self-awareness. This internal discovery of the self-offers a more precise basis for making decisions, which can, in turn, enhance the alignment between individual actions and broader societal welfare, resonating with Smith's notion of the "invisible hand."

Implications for Political Economy

This reinterpretation of Adam Smith's moral philosophy-anchored in the three stages of self-awareness-carries significant implications for political economy. By mapping the process through which individuals refine their ability to evaluate their own experiences and accurately gauge the value of pleasure and pain, this framework provides a more sophisticated understanding of decision-making.
This study introduces a novel perspective on how Smith's ideas can contribute to a more human-centered approach to political economy. By highlighting the role of self-awareness and self-love in shaping decision-making, this paper argues that Smith's work opens a pathway for addressing contemporary ethical and economic complexities. It proposes the development of market systems that are more attuned to individual well-being and collective societal prosperity.

Keywords: Theory of Moral Sentiments, Adam Smith, Moral Philosophy, Self, Love, Political Economy, individual behavior, societal interaction.

Standardizing mean differences: A cautionary tale from a rational choice perspective

Ina Jäntgen*†[1]

[1]Ludwig Maximilian University – Germany

Abstract

Decision-makers often want to achieve their goals as effectively as possible. For instance, a policymaker might aim to reduce childhood deaths from diarrhea as much as possible (an example inspired by Bhalotra et al. 2021), or a principal might want to improve students' reading abilities as much as possible (an example inspired by Kulik and Fletcher 2016). Such agents often turn to science for advice. And in applied areas of science – including development economics, education research, or medicine – researchers indeed measure and report the effectiveness of tested interventions, such as a treatment's effectiveness for curing a disease. For continuous outcome variables (such as reading test scores), researchers measure an intervention's effectiveness – its effect size – using the *mean difference*, quantifying how much, on average, the intervention changed the measured outcome. Or they measure the intervention's effect size using the *standardized mean difference*, quantifying by how many standard deviations the intervention, on average, changed the measured outcome. The resulting effect sizes commonly inform evidence-based decision-making, ranging from clinical decisions to policymaking. But these effect size measures provide decision-makers with different information about an intervention's effectiveness. To inform decision-making well, should scientists then report the mean difference, the standardized mean difference or both to decision-makers? In line with recent work on effect sizes for binary outcome variables (e.g., Sprenger and Stegenga 2017), I focus on one important and thus far undiscussed angle on this question: When do interventions' standardized but not unstandardized mean differences provide rational decision-makers with the information they care about learning when deciding between interventions?

This question gets bite once we consider why researchers predominantly report standardized but not unstandardized mean differences in many areas of the biomedical and social sciences. The reason lies in the uncertainty concerning measurement researchers often face in these areas (e.g., Rosnow and Rosenthal 2009): Researchers are often uncertain about a) how the measures used in various studies relate to each other (e.g., how scores from different reading tests compare to each other) and b) whether these measures even measure the same outcome of interest (e.g., whether these tests all measure the same reading skill). Such measurement uncertainty makes interpreting and comparing mean differences difficult; researchers are uncertain about which changes in the outcome of interest observed mean differences represent. To interpret mean differences and compare them across studies, even when facing such measurement uncertainty, researchers typically standardize the mean differences. Standardizing mean differences facilitates meta-analyses (e.g., Schmid, Stijnen, and

*Speaker
†Corresponding author: Ina.Jaentgen@lmu.de

White 2022) and allows researchers to communicate the effectiveness of interventions (e.g., Cummings 2011) despite measurement uncertainty, or so the claim.

I argue that the widespread practice of standardizing mean differences is risky: Even when facing measurement uncertainty, reporting interventions' standardized but not their un- standardized mean differences to decision-makers risks omitting information that rational decision-makers (here: those maximizing expected utility) would care about learning when deciding between these interventions – information these agents could learn from the mean differences and standard deviations researchers could instead report. To derive this result, I propose a novel Bayesian model for how a rational agent decides between interventions after learning from mean differences expressed in measurements of outcomes about which the agent is uncertain (building on Howson and Franklin 1994; Philippi 2022).

This negative result allows us to recognize a hitherto neglected tension between the harm of standardizing mean differences (risking losing information that matters to rational decision- making) and its benefits (facilitating meta-analyses and communication of effectiveness despite measurement uncertainty); reaping these benefits can deprive people of decision- relevant information. But are there scenarios in which we can reap the benefits while avoiding the risk?

There are, as I also prove using my decision model. Even when facing measurement uncertainty, in *some* scenarios, learning the interventions' standardized mean differences informs a rational agent equally well to decide between these interventions than learning the mean differences and standard deviations would inform them. This holds for many rational agents when researchers face either of two scenarios: when the control groups of different studies observed the same standard deviation or when the data in all studies is normally distributed. When researchers face these scenarios, then (for many agents) there is *no* risk of losing in- formation that matters to rational decision-making.

To inform rational decision-making well, should scientists then continue to report standardized but not unstandardized mean differences to people when facing measurement uncertainty? As I discuss, my results allow us to recognize open empirical and methodological questions that we should answer to settle when researchers should standardize mean differences. Moreover, as I also discuss, my results add nuance to a longstanding debate in statistics on the advantages of effect sizes over statistical significance in conveying the practical significance of interventions (e.g., Sprenger 2016; Cohen 1994; Ziliak and McCloskey 2008). Overall, this paper adds normative considerations to methodologists' debates on the merits and downsides of standardized mean differences in evidence-based decision-making (e.g., Baguley 2009; Cummings 2011).

Selected References

Baguley, Thom. 2009. "Standardized or Simple Effect Size: What Should Be Reported?" *British Journal of Psychology (London, England: 1953)* 100 (Pt 3): 603–17.

Bhalotra, Sonia R., Alberto Diaz-Cayeros, Grant Miller, Alfonso Miranda, and Atheendar S. Venkataramani. 2021. "Urban Water Disinfection and Mortality Decline in Lower-Income Countries." *American Economic Journal: Economic Policy* 13 (4): 490–520.

Cohen, Jacob. 1994. "The Earth Is Round (p< .05)." *American Psychologist* 49 (12): 997–1003.

Cummings, Peter. 2011. "Arguments for and Against Standardized Mean Differences (Effect Sizes)." *Archives of Pediatrics & Adolescent Medicine* 165 (7): 592–96.

Howson, Colin, and Allan Franklin. 1994. "Bayesian Conditionalization and Probability Kinematics." *The British Journal for the Philosophy of Science* 45 (2): 451–66.

Kulik, James A., and J. D. Fletcher. 2016. ”Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review.” *Review of Educational Research* 86 (1): 42–78.

Philippi, Cristian Larroulet. 2022. ”Against Prohibition (Or, When Using Ordinal Scales to Compare Groups Is OK).” *The British Journal for the Philosophy of Science*, July.

Rosnow, Ralph L., and Robert Rosenthal. 2009. ”Effect Sizes: Why, When, and How to Use Them.” *Zeitschrift Für Psychologie/Journal of Psychology* 217:6–14.

Schmid, Christopher H., Theo Stijnen, and Ian White, eds. 2022. *Handbook of Meta- Analysis*. 1st edition. Boca Raton: Chapman and Hall/CRC.

Sprenger, Jan. 2016. ”Bayesianism vs. Frequentism in Statistical Inference.” In *The Oxford Handbook of Probability and Philosophy*, edited by Alan Hájek and Christopher Hitchcock, First edition, 382–405. Oxford: Oxford University Press.

Sprenger, Jan, and Jacob Stegenga. 2017. ”Three Arguments for Absolute Outcome Mea- sures.” *Philosophy of Science* 84 (5): 840–52.

Ziliak, Stephen Thomas, and Deirdre N. McCloskey. 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Illustrated edition. Ann Arbor: University of Michigan Press.

Keywords: evidence, based policymaking, effect sizes, standardized mean differences, expected utility theory, measurement uncertainty

Like It or Not - Recommender Systems, Welfarism, and Performativity

Donal Khosrowi*[1] and Lukas Beck*[2,3]
[1]Leibniz University Hannover – Germany
[2]London School of Economics (LSE) – United Kingdom
[3]Leibniz University Hannover – Germany

Abstract

Recommender systems (RS) are widely-used algorithmic systems designed to suggest relevant items or contents to users. To do so, RS typically employ a mix of collaborative filtering and content-based filtering. Collaborative filtering makes suggestions based on a user's past behavior e.g., likes, engagement, or purchases - and choices made by other users, whereas content-based filtering recommends other items and contents with similar pre-defined characteristics (Melville and Sindhwani 2010).The central aim of RS is often characterized as finding good or relevant items for users (see Milano et al. 2020). According to this dominant narrative, the goal in building and employing RS is to "estimate the users' preferences for unseen items or to provide users with item lists ranked in accordance to the estimated preferences" (Jannach and Adomavicius 2016, 7). What is striking about this narrative is that it offers a welfare consequentialist justification for RS (see Sen 1979; Adler 2019): RS should be employed because they make users better off. Importantly, any welfare consequentialist justification is in need of a welfare criterion, i.e. an account that tells us how to rank two outcomes in terms of agents' welfare. In this regard, developers and proponents of RS appeal to users' preferences.

In this paper, we critically examine the welfare consequentialist narrative underlying RS. We argue that the RS literature either 1) lacks an explicit, coherent notion of preferences, or 2) to the extent that it is based on one, it is based on outdated notions that are out of touch with decades of behavioral research. In particular, these notions assume that agents' preferences are a stable target to be learned from, and readily revealed by, choice data.

First, we ask whether it is plausible to think that RS deal in genuine information about preferences. To illustrate this worry, consider RS that mainly rely on content-based filtering. Why should we believe that suggestions based on predefined attributes of items are able to track user preferences? The desires and judgments that form users' preferences may, for instance, concern attributes different from those utilized by RS. Hence, at least in some instances, there is a disconnect between the tools RS employ for making suggestions and the welfare criterion that is meant to justify RS.

A second, more severe worry is that even if RS rest on a conceptually coherent notion of preferences - think of content-based filtering where suggestions are based on past behavior - the question remains whether this notion of preferences can deliver the intended welfare consequentialist justification.

* Speaker

This question is pressing as several decades of research in behavioral economics suggest that preferences are less stable than conventionally assumed and that behavior is often co-determined by complicated interactions between underlying preferences and other context-dependent psychological features (see Bruni and Sugden 2007, Thoma 2021). In light of this, a näive conception of preferences that simply equates preference with past choice behavior threatens to undermine RS' ability to deliver on their welfare consequentialist ambitions.

In particular, we consider three worries that highlight how RS may systematically fail to cater to agents' preferences. First, agents may merely change behavior as a result of their exposure to RS recommendations, while their underlying (welfare-relevant) preferences re- main unsatisfied. In such cases, choice behavior (e.g. engaging with contents) is a poor guide to estimating preference satisfaction. Yet, so far, appeals to the welfare consequentialist narrative in support of RS fail to sufficiently acknowledge and address this possibility. Second, RS may be prone to cater to users' "lower" desires, instead of tracking their (welfare-relevant) all-things-considered evaluations (see Thoma 2021). For example, while an agent may have a myopic desire to watch trashy TV shows, they may all-things-considered prefer not to watch such shows and instead strive to cultivate more refined viewing habits. In such a case, merely satisfying the user's myopic desires would not deliver the welfare consequentialist promises of RS. Third, there are reasons to believe that many predictive systems, including RS, are performative (van Basshuysen 2023; Khosrowi 2023): instead of estimating and catering to users' stable, exogenous preferences, they may change these very preferences through the recommendations they make (Adomavicius et al. 2022; Kalimeris et al. 2021). Such performative effects may be understood as intervening with authentic preference formation and user autonomy, thereby undermining the welfare criterion on which the justification of RS relies.

Together, these three issues raise fundamental concerns about whether the effects RS have on users' behavior are welfare-improving, e.g. as evidenced by user engagement. Rather than enhancing welfare, RS may merely lead to more, or different, engagement and purchases, while leaving unclear whether this makes users better off. In light of this concern, we argue that the welfare consequentialist narrative for RS is currently underdeveloped and requires a more explicit conception of preference.

To address this situation, we highlight some potential avenues for further research that could make RS' welfare consequentialist justification more compelling. In particular, we suggest collaborations at the interface between RS research and behavioral welfare economics that should study the feasibility of behaviorally informed RS; for instance by 1) empirically studying the effects of recommendations on preferences, 2) building more sophisticated conceptual scaffolding that captures behavioral nuances in interaction with RS, and 3) building behaviorally richer models that allow us to identify welfare-relevant preferences. Such re- search, we suggest, is a crucial step in the responsible development and deployment of RS that succeed in delivering on their welfarist ambitions.

References

Adomavicius, G., J. Bockstedt, S. Curley, and J. Zhang. (2022). "Recommender systems, ground truth, and preference pollution". AI Magazine, 43(2):177–189.

Adler, M. D. (2019). Measuring social welfare: An introduction. Oxford University Press.

Bruni, L., and R. Sugden. (2007). "The road not taken: how psychology was removed from economics, and how it might be brought back". The Economic Journal, 117(516), 146- 173.

Jannach, D., and G. Adomavicius. (2016). Recommendations with a purpose. In Proceedings of the 10th ACM conference on recommender systems (pp. 7-10).

Kalimeris, D., S. Bhagat, S. Kalyanaraman, and U. Weinsberg. (2021). "Preference amplification in recommender systems". In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 805–815.

Khosrowi, D. (2023). "Managing Performative Models". Philosophy of the Social Sciences 53(5): 371-395.

Melville, P., and V. Sindhwani (2010). "Recommender systems". Encyclopedia of machine learning, 1, 829-838.

Milano, S., M. Taddeo, and L. Floridi. (2020). "Recommender systems and their ethical challenges". Ai & Society, 35, 957-967.

Sen, A. (1979). ”Utilitarianism and welfarism”. The Journal of Philosophy, 76(9), 463- 489.

Thoma, J. (2021). ”On the possibility of an anti-paternalist behavioural welfare economics”. Journal of Economic Methodology, 28(4), 350-363.

van Basshuysen, P. (2023). ”Austinian Model Evaluation”. Philosophy of Science 90 (5): 1459–68. doi:10.1017/psa.2023.24.

Keywords: Recommender systems, preference, welfare, welfarism, machine learning, performativity, behavioral economics

Model Choice in Economics: Simulation Models, Empirical Evidence and (Normative) Representational Decisions

Anna-Katharina Kothe*†[1]

[1]Europa-Universität Flensburg – Germany

Abstract

When trying to gain a deeper understanding of economics as a scientific discipline, inspecting its models has been a fruitful path in philosophy of economics literature (see for example Heilmann and Reiss 2022; Kincaid and Ross 2021; Knuuttila, Carrillo, and Koskinen 2024; Morgan and Knuuttila 2012). Although many open questions about models in economics remain, it is clear that economics is exposed to and employs a variety of different models: in recent years there has been a growing focus on computerized or *simulation models* in economics (Backhouse and Cherrier 2017; Boumans 2019; Lehtinen and Kuorikoski 2022; Winsberg 2022; Ylikoski 2014). Locating itself within this literature, the paper aims at investigating the (missing) acceptance and application of simulation models in economics. This missing acceptance of simulation model in economics is well documented by Lehtinen and Kuorikoski (2007), who conduct an empirical study on use of simulation models in major economic journals. They attest that 'simulation is still commonly frowned upon in the economics community' (Lehtinen and Kuorikoski 2007:305) and pose the subsequent question: why do economists *'shun'* simulation? Other empirical studies in a similar vein follow suit (Fontana 2006; Lind 2024; Reiss 2011) with varying explanations for the phenomena. Lehtinen and Kuorikoski's explanation is that economists have a certain mode of understanding through the commonly used mathematical models, and that simulation models do not cater to this mode of understanding. Reiss (2011) explains the dearth of simulation models in eco- nomics by 'their characteristics (which) are compared with theoretical models and of course simulations lack mathematical elegance, certainty of results, and analytical transparency' (2011:260).

While these explanations are insightful, the paper aims at extending this debate and amend- ing the work done by Lehtinen and Kuorikoski on simulation in economics – and consequentially model types and model selection in economics – in two ways: first, by reassessing the empirical foundation on the use of simulation models in economics and second, by attempting a new explanation to the updated empirical findings.

In a *first* step, the paper presents an *updated empirical assessment* of the status quo of simulation models in economics. By building a database of journal publications of the last twenty years not only for economics but also for other social and natural sciences, the claim that economists 'shun' simulation can be re-assessed also in comparison to other (social) sciences. The result of this empirical study is that a much more complex and ambiguous picture

*Speaker

†Corresponding author: anna-katherina.kothe@uni-flensburg.de

of simulation in economics presents itself: rather than rejecting simulation altogether, some economists in some specialized fields seem to be reluctant to employ some forms of simulation.

This updated empirical claim is then the basis for the *second* step, namely how the diffusion of some model types, especially simulation models, across economics might be explained. The claim is that model selection is – at least to some extent – value-laden through the channel of *representational decisions*. Representational decisions are decisions about what to represent (i.e., decisions about what entities to include in and exclude from a representation) and a decision how to represent (i.e., decisions about entities already chosen for inclusion in the representation) (Harvard and Winsberg 2022). In making these decisions, the modeler faces representational risk, namely the risk that she makes "wrong" or inadequate representational decisions. By evaluating this representational risk, value judgements come into play and introduce epistemic and non-epistemic values into the modelling process. This means that epistemic and non-epistemic value judgements enter through the channel of representational risk, which are embedded in representational decisions in the modelling process. The crucial point is that some model types prescribe at least some representational decisions: simulation models necessarily prescribe some representational decisions due to their model structure. For example, in the case of agent-based models the modeler is committed to include multiple, autonomous agents with behavioral heuristics into her model, as such agents constitutes an essential element of an agent-based model. In consequence, there is some representational risk – and through channel the according value judgements – inherent to the choice of model type. The argument in this paper is thus that one answer to the peculiar diffusion of simulation models in economics lies in the necessary representational decisions of those model types and the consequential representational risk. In other words, economists shun simulation models because choosing these model types entails representational risk and economists deem this risk too high for them to accept it – at least in comparison to other commonly employed economic models.

References

Backhouse, Roger E., and Béatrice Cherrier. 2017. "'It's Computers, Stupid!' The Spread of Computers and the Changing Roles of Theoretical and Applied Economics." *History of Political Economy* 49(Supplement):103–26. doi: 10.1215/00182702-4166287.

Boumans, Marcel J. 2019. "Simulation and Economic Methodology." Pp. 41–50 in *Models as Epistemic Artefacts: Toward a Non-Representationalist Account of Scientific Representation*, edited by W. Dolfsma, D. W. Hands, and R. McMaster. Abingdon, Oxon; NewYork, NY: Routledge, 2019. — Series: Routledge advances in social economics: Routledge.

Fontana, Magda. 2006. "Simulation in Economics: Evidence on Diffusion and Communication." *Journal of Artificial Societies and Social Simulation* 2(8).

Grüne-Yanoff, Till, and Caterina Marchionni. 2018. "Modeling Model Selection in Model Pluralism." *Journal of Economic Methodology* 25(3):265–75. doi: 10.1080/1350178X.2018.1488572.

Harvard, Stephanie, and Eric Winsberg. 2022. "The Epistemic Risk in Representation." *Kennedy Institute of Ethics Journal* 32(1):1–31. doi: 10.1353/ken.2022.0001.

Heilmann, Conrad, and Julian Reiss, eds. 2022. *The Routledge Handbook of the Philosophy of Economics*. First published. New York London: Routledge, Taylor & Francis Group.

Kincaid, Harold, and Don Ross, eds. 2021. *A Modern Guide to Philosophy of Economics*. Edward Elgar Publishing.

Knuuttila, Tarja, Natalia Carrillo, and Rami Koskinen. 2024. *The Routledge Handbook of Philosophy of Scientific Modeling*. 1st ed. London: Routledge.

Lehtinen, Aki, and Jaakko Kuorikoski. 2007. "Computing the Perfect Model: Why Do

Economists Shun Simulation?" *Philosophy of Science* 74(3):304–29. doi: 10.1086/522359.

Lehtinen, Aki, and Jaakko Kuorikoski. 2022. "Computer Simulations in Economics." Pp. 355–69 in *The Routledge handbook of the philosophy of economics*, edited by C. Heilmann and J. Reiss.

Lind, Hans. 2024. *Theories and Models in Economics: An Empirical Approach to Method- ology*. Edward Elgar Publishing.

Morgan, Mary S., and Tarja Knuuttila. 2012. "Models and Modelling in Economics." Pp. 49–87 in *Philosophy of Economics*. Elsevier.

Reiss, Julian. 2011. "A Plea for (Good) Simulations: Nudging Economics Toward an Experimental Science." *Simulation & Gaming* 42(2):243–64. doi: 10.1177/1046878110393941.

Winsberg, Eric. 2022. "Computer Simulations in Science." *Stanford Encyclopedia of Philosophy*.

Ylikoski, P. 2014. "Agent-Based Simulation and Sociological Understanding." *Perspectives on Science* 22:318–35.

Keywords: simulation models, representational risk, values in science, modelling process, model selection

When all bets fail (to reveal beliefs): ”The Problem of Two Perspectives”

Vita Kudryavtseva*[1]

[1]London School of Economics and Political Science – United Kingdom

Abstract

Betting as a preference-based method of belief measurement is considered fundamentally sound by many economists and decision theorists alike. Many criticisms have been levied against the procedure over the years, however, the proponents of the methodology hold a powerful line of defence —- similar problems affect the measurement in natural sciences and it is up to empirical scientists to improve upon the procedure’s accuracy. What’s of essence from a philosophical viewpoint is having a well-defined measurement procedure that gives the degree of belief its ”precise meaning”. (Ramsey, 1926)
In this paper, I identify a fundamental issue with the idealized betting procedure – ”The problem of two perspectives”. A bet is a special kind of action, a transaction. Two dis- tinct kinds of actors can be a party to a bet: a bettor and a bookie. Hence, two distinct perspectives can be used to evaluate a bet – a bettor’s and a bookie’s. I show that (a) a bookie’s fair betting rate is not equal to her true credence in a proposition at stake, argue that (b) a bookie’s perspective is the one a rational agent ought to entertain and that (c) the empirical scientists do not have the means to rectify the problem on their own. Therefore, a fair betting rate, or equivalently the lowest acceptable odds, cannot be used to infer an agent’s credence in a proposition.

In Part I, I introduce the bettor’s and bookie’s perspectives on an offer of a bet on p. A bettor weighs the merits of a bet on a stand-alone basis and profits from the goodness of her epistemic state with regards to the proposition p. In contrast, a bookie’s essential aim is to turn a profit on the book as a whole regardless of the truth-value of p – after all, that’s what makes her a bona fide bookie. A minimal requirement for a betting agent to be considered a ‘bookie’ is that a bettor must have at least one currently unsettled bet (a ‘book’). I put forth two alternative decision problem representations that a bookie can entertain. I demonstrate that whether a bookie’s fair betting rate can be interpreted as her true credence in p depends on which perspective she adopts. From ”Genuine Bookie”’s perspective, the lowest odds a bookie is willing to accept depend on the composition of the book and her beliefs about other possible bets that can be available to her. Hence, the betting procedure will fail to reveal her true credence in proposition p if this is indeed the bookie’s viewpoint on the decision situation.

Next, I turn to the question of which of the two perspectives a bookie ought to adopt. I make a case that a failure to view the betting problem from the ”Genuine Bookie”’s perspective, or simply put, to consider whether more lucrative bets may be available to her, constitutes a failure of the bookie’s imagination. First, drawing on the literature on the epistemic use of imagination I confirm that a bookie could have possibly

*Speaker

imagined such eventuality. Second, I argue that she would be better off doing this because this would result in the decision problem representation that has a stronger relevance claim. Relevance claim is the claim that the decision problem's description includes all the relevant aspects of the decision situation that an agent cares about. It is the necessary pre-condition of rational choice on a standard view of Subjective Expected Utility (SEU) theory. Hence, a bookie ought to adopt the "Genuine Bookie" representation of the betting situation and her fair betting rate cannot be interpreted as her true credence in p.

Clearly, if an investigator does not know whether she is facing a bookie or a bettor, this is a problem for interpreting the results of the procedure as credences. In Part III, I consider the three avenues open to an experimenter to nip the problem in the bud and ensure the procedure's validity. First, I address the question of whether bookies are exceedingly rare in the general population and/or could be filtered out prior to commencing the betting procedure. I conclude that, on the contrary, we are all 'bookies' by the lights of SEU, the very framework underpinning this measurement procedure. SEU views every action as a lottery over its possible outcomes - a gamble. Hence, to the extent we are still waiting to see how some of our prior actions are going to turn out, we hold a book of unsettled bets. I argue that this entails that betting as a belief measurement procedure can only be applied to Ramsey's "ethically-neutral" propositions – a severe limitation. Second, I consider whether an experimenter can design the experiment to steer the subject away from adopting the "Genuine Bookie" representation. I conclude that no such design modification is available for the procedure conceived as a series of bets. Lastly, I reflect on the nature of the betting procedure as a problem of conjoint measurement of degree of belief and utilities. I entertain and reject the possibility that a careful measurement of a bookie's utilities for a bet's outcomes can solve the 'Two Perspective Problem". Hence, betting as an idealized measurement procedure faces an insurmountable challenge.

Keywords: Belief measurement, Bayesian decision theory, betting.

Who are experts on the sustainability of economic growth?

Teemu Lari*†[1]

[1]Leibniz University Hannover – Germany

Abstract

It is a widely shared ideal that policymaking should be supported by reliable knowledge when possible and that factual claims made and assumed in policymaking should be compatible with the best available knowledge of the world. However, people often need to rely on experts such as researchers to get informed about what the best currently available knowledge about a question is. For example, laypeople should base their beliefs about the efficacy and safety of vaccines on what vaccine researchers tell us about these issues. If there are several experts with mutually conflicting views, much research in social epistemology and philosophy of science advises laypeople to use social indicators of trustworthiness such as trusting that putative expert whose view represents the consensus view in the field (Anderson, 2011; Goldman, 2001). Philosophical research has so far focused on cases in which it is relatively unproblematic to determine the scientific discipline or research field relevant to the question at hand: To get the best knowledge about vaccines, ask vaccine researchers! Also, if possible, determine what the consensus view in the relevant field is, because consensus among experts on a claim is taken to be ”second-order evidence” for that claim. But what if there is more than one discipline or research field in which putative experts have a strong view on the issue at hand? Who are the experts that laypeople should trust? Who are the experts whose possible consensus would count as second-order evidence?

Among the biggest questions of our time is whether economic growth can continue indefinitely despite worries such as climate change and biodiversity loss. It could have catastrophic consequences if around the world policymaking would be based on a mistaken assumption that the pursuit of economic growth will not undermine the bio-physical preconditions of hu- man life. Perplexingly, there seem to exist two opposing consensus views on what we might call the biophysical-constraints-to-growth question: is globally continuing economic growth compatible with the constraints set by the biophysical reality?

The widely shared view among mainstream economists regards economic growth as compatible with environmental boundaries and recommends pursuing ”green growth” enabled by technological advances and environmentally friendly policies which reduce the environ- mental harms of economic activity.

*Speaker
†Corresponding author: teemu.lari@gmail.com

The green growth consensus is reflected, for example, in how all envisioned scenarios in the IPCC reports assume the continuation of economic growth (Keyßer & Lenzen, 2021). But there is a competing consensus – emerging, at least. The research field known as ”post-growth research” – a multidisciplinary epistemic community including for example ecological economists (Kallis et al., 2025) – is converging on the view that the continuation of economic growth across the world cannot be combined with less-than-devastating climate outcomes under any realistic assumptions. This is due to the stringency of environmental boundaries such as the remaining carbon budget compatible with keeping global warming at tolerable levels (Hickel & Kallis, 2020).

Unfortunately, the current tools of social epistemology cannot tell which experts one should trust in this case, as researchers on both sides seem to display the conventional markers of expertise, such as convincing academic credentials.

To solve the problem, I take a social epistemology perspective to compare the abilities of the two epistemic communities, economics and post-growth research, to reliably answer the biophysical-constraints-to-growth question. Generally, epistemic communities such as disciplines and their subfields are characterized by distinctive norms, conventions, values and conceptions of epistemic goals. These affect what kind of research the community members conduct and consequently what kind of research questions they can answer. For example, if a community had a methodological norm that prescribes using only quantitative and formal research methods, it would not be well positioned to answer research questions about the meanings and interpretations that people give to societal phenomena. In other words, the elements belonging to the social level of epistemic communities (the norms, conventions, values, etc.) are fit for the purpose of answering some questions but not others. To determine who are the most reliable experts in a divisive question, we should examine whether the social elements of seemingly relevant disciplines and research fields are conducive towards answering the question reliably. Consensus position in that field which is best positioned to answer the question can be taken as best available knowledge.

I argue that the epistemic community of mainstream economics is not well aligned with the biophysical-constraints-to-growth question. The methodological norms and conventions regarding what kind of idealizations it is appropriate to make, how research questions should be framed and what the epistemic goals of the discipline are stand in tension with what it takes to answer the question in a reliable way. In contrast, the epistemic community of post- growth research takes as its explicit goal to discover to what extent there are bio-physical limits to economic growth and accordingly has settled on norms, conventions and values that are more appropriate for answering the question. With regard to this question, post-growth researchers should therefore be seen as the experts that laypeople, including policymakers, should rely on.

Regardless of this conclusion, there are other important questions about economic growth and economic policy that post-growth researchers arguably are not as well placed to answer as mainstream economists are. These include questions concerning the consequences that discontinued growth would have under current economic institutions to macroeconomic variables like employment and public debt, which in turn are relevant for the space of eco- nomically and politically feasible policies. This means that mainstream economists have an important role in answering policy-relevant questions around sustainability policies – we just need to think more carefully about what the exact questions are in which they are the best experts.

References

Anderson, E. (2011). Democracy, Public Policy, and Lay Assessments of Scientific Testimony. Episteme, 8(2), 144–164. https://doi.org/10.3366/epi.2011.0013

Goldman, A. I. (2001). Experts: Which Ones Should You Trust? Philosophy and Phenomenological Research, 63(1), 85–110.

Hickel, J., & Kallis, G. (2020). Is Green Growth Possible? New Political Economy, 25(4), 469–486. https://doi.org/10.1080/13563467.2019.1598964

Kallis, G., Hickel, J., O'Neill, D. W., Jackson, T., Victor, P. A., Raworth, K., Schor, J. B., Steinberger, J. K., & Ürge-Vorsatz, D. (2025). Post-growth: the science of wellbeing within planetary boundaries. The Lancet Planetary Health, 9 (1), e62–e78. https://doi.org/10.1016/S2542- 5196(24)00310-3

Keyßer, L. T., & Lenzen, M. (2021). 1.5◦C degrowth scenarios suggest the need for new mitigation pathways. Nature Communications, 12(1), 2676. https://doi.org/10.1038/s41467- 021-22884-9

Keywords: economic growth, degrowth, post, growth, sustainability, ecological economics, expertise, policy

The Economics of Coercion: Kolm's Liberal Social Contract and Public Goods

Kevin Leportier *[1]

[1]University of Caen Normandie - CREM – Université de Caen, CREM UMR CNRS 6211 – France

Abstract

The public good argument has a long history in economics, with contributions from Adam Smith and John Stuart Mill, among many others. Mill famously argued that "no one would build lighthouses from motives of personal interest unless indemnified and rewarded from a compulsory levy made by the state" (Mill 1848). This reflects the core issue of free riding, which is often seen as justifying state intervention to ensure the optimal provision of public goods. The public good argument (Schmidtz 1991) states that, due to the free rider problem (and the symmetric assurance problem), the status quo equilibrium-where no one contributes to the production of the public good-is expected to persist when everyone is free to choose. However, since forcing a certain number of individuals to contribute constitutes a Pareto improvement over the status quo, imposing this constraint would be justified as it is necessary from the perspective of welfare improvement. The economic debate primarily revolves around whether state intervention is truly warranted by the existence of public goods (Varian 1993). However, from a normative economics perspective, an equally important aspect of the argument is the role of coercion. Implicitly, the public good argument presents a trade-off between welfare and freedom: in Mill's example, the collective benefit derived from the erection of the lighthouse is supposed to more than offset the harm caused by the imposition of a "compulsory levy". Even if such a levy would prove to be consensual, would that really be the case for enforcing constraints in other, more sensitive contexts? Is the judgment that welfare improvement more than offsets the harm caused by coercion really something most people could agree on? In any case, traditional welfare economics is ill-equipped to examine or justify this kind of judgment, as there is no widely agreed-upon way to frame and analyze this trade-off.

Among classical solutions to public good problems, private provision of public goods through "assurance contracts" (Palfrey and Rosenthal 1984, Bagnoli and Lipman 1989, Tabarrok 1996) designed to mitigate the free-rider and assurance problems suggests that coercion may not be necessary, which would simply refute the public good argument. This line of response to the public good argument was also developed from a libertarian perspective by some philosophers (Schmidtz 1994): there would be no need to face such a difficult trade-off between freedom and welfare. Assurance contracts enable people to coordinate to produce public goods by making the contribution of someone conditional on the contribution of others, and (in Tabarrok's version) by giving people incentives to sign the contract even if they expect it to fail. However, these contracts would only be viable for small and connected groups of beneficiaries, as transaction costs increase significantly with scale.

*Speaker

In parallel, Serge-Christophe Kolm's work, at the intersection of philosophy and economics, informed by Kolm's work in public economics and his reading of Nozick, offers an alternative response to the public good argument through the concept of the 'liberal social contract' (1985). This hypothetical contract, made by actual individuals in actual societies, differs from traditional social contracts in two key ways: it directly binds specific individuals rather than remaining abstract, and it commits them to providing specific public goods rather than merely justifying the existence of a broader and more encompassing set of institutions.

Since all individuals facing a public good problem have an interest in coordinating their commitments to contributing to its production, a social contract concerning a public good is expected to result in its actual provision. The outcome of such a contract represent what individuals would have agreed to under appropriate conditions, making it enforceable even if, due to high transaction costs or time constraints, it was never formally signed. It provides an alternative response to the public good argument by suggesting how the trade-off between welfare and freedom can be negotiated through the act of contracting. Kolm's liberal social contract may be deemed "liberal" for this reason, but it is not quite libertarian as it legitimizes coercion and the existence of a "more than minimal" state needed to enforce it.

This article compares Kolm's approach with the literature on assurance contracts, highlighting the libertarian influences underlying both perspectives. It also aims to underscore the originality of Kolm's economic philosophy, particularly regarding his concept of the liberal social contract, which fruitfully connects public economics and the philosophical tradition of the social contract, yet remains little analyzed.

Keywords: Public Goods, Coercion, Assurance Contracts, Liberal Social Contract, Normative Economics, Libertarianism, Serge, Christophe Kolm

The Value Judgments of Economists and of Economic Agents: The Case for a Distinction

Nestor Lovera*[1]

[1]Recherches en Économie Gestion Agro Ressources Durabilité Santé- EA 6292 – Université de Reims Champagne-Ardenne, Universit´e de Reims Champagne-Ardenne : EA6292 – France

Abstract

The distinction between economists' value judgments and those of economic agents remains unclear in the welfare economics literature. I argue that clarifying this distinction is essential for determining whether economists may incorporate value judgments into their professional work and how such judgments play a crucial role in justifying economic agents' preferences. Rather than treating these aspects separately, I examine their interconnected- ness. My aim is not only to draw the distinction but also to highlight its significance and practical implications for normative economics. Accordingly, this article seeks to answer two key questions: How should economists account for the value judgments of economic agents? And, in doing so, to what extent should they refrain from introducing their own when offering advice or otherwise engaging with agents' value judgments?

Welfare economics has been at the center of this debate because it provides a theoretical framework that allows assessing the consequences of individual actions and public decisions on social states (Baujard 2017). Behind individuals' actions, there are the value judgments of individuals. Counseling economic agents is economists' concern, and this counseling is meaningful if it is based on values of economic agents (Bergson 1954). A subject that relates the value judgments of economists to those of economic agents is interpersonal comparisons of utility (ICU's). Accepting or rejecting ICU's has been crucial in the history of welfare eco- nomics and therefore in normative economics' progress.(1) The rejection of ICU's has been a key element in the transition from the "old" to the "new" welfare economics. However, ICU's will be reintroduced in welfare economics in order to avoid Arrow's impossibility theorem (Backhouse, Baujard, and Nishizawa 2021, 5–6). In the "old" welfare economics, ICU's can play an essential role in determining economic policies to improve people's well-being. The way ICU's are made reveals a normative conception. In the "new" welfare economics, ICU's are rejected because these comparisons are based on value judgments and not on empirical facts and also because the economist cannot match them with possible experience (Mongin 1999).

Although these two positions on interpersonal comparisons of utility seem irreconcilable, Harsanyi (1955; 1977) argues that it is possible to make ICU's, without resorting to value judgments. The analysis of Harsanyi's approach to ICU's allows one to study how an economist can claim that ICU's are not based on ethical or political value judgments. This is because Harsanyi assumes that individual preferences are based on impersonal attitudes (extended preferences) guaranteed by the existence of an impartial observer. I use Harsanyi's attempt to 'objectify' ICU's to analyze the distinction between the value judgments of economists and

*Speaker

those of economic agents in welfare economics. This distinction has not been explicitly established in welfare economics, which has concentrated on the debate on economics as a neutral social science. As a result, economists have to step back and consider not only their own value judgments but also those of economic agents when they study eco- nomic phenomena. So, my contribution is to bring this distinction to light and propose an extended theoretical framework that allows it to be better studied.

This paper is structured as follows. First, I use interpersonal comparisons of utility as a case study to examine the distinction between the value judgments of economists and those of economic agents, focusing on Harsanyi's impartial observer theorem. Next, I discuss two key criticisms of this theorem: (1) Broome's (1993) well-known critique, which highlights the role of individual value judgments that Harsanyi sought to bypass through his assumption that the same psychological laws govern all individuals, and (2) MacKay's (1986) argument, which emphasizes the need to properly distinguish between the value judgments of economists and economic agents to prevent the risk of "experimental contamination." Finally, examining these criticisms will lead me to explore the extent to which subjective attitudes can be "objectified" when making interpersonal comparisons of utility. To shed light on this discussion, I draw on Dietrich and List 's (2013) framework, which introduces a reason-based theory of rational choice to explain how economic agents' preferences are determined. This theory provides insight into how the psychological laws referenced by Harsanyi are grounded "a priori" in motivating reasons. A motivating reason, as Dietrich and List (2013, 6) describe, "may affect the agent's actual preference for the given alternative vis-à - vis others."

(1) Mongin(2006) studies the progress in normative economics. He proposes a provisional progress definition, which will be his main tool for analyzing what he calls the intertheoretic progress in normative economics. Thus, Mongin suggests the following definition of progress:

"A shift from theory T to theory T' is progressive if: (1) T' provides a solution to at least one unresolved problem of T; (2) T' provides a solution to the main problems that T had already addressed and resolved in its own way; (3) T' raises new problems and manages to solve at least one of them; (4) T does not satisfy the previous conditions with respect to T'" (Mongin 2006, 25).

References

Backhouse, Roger E, Antoinette Baujard, and Tamotsu Nishizawa. 2021. "Introduction: Revisiting the History of Welfare Economics." In *Welfare Theory, Public Action, and Ethical Values: Revisiting the History of Welfare Economics*, edited by Roger E Backhouse, Antoinette Baujard, and Tamotsu Nishizawa, 1–18. Cambridge: Cambridge University Press.

Baujard, Antoinette. 2017. "L'´economie Du Bien-Être Est Morte . Vive l'´economie Du Bien -Être!" In *Philosophie Économique. Un État de Lieux.*, edited by Gilles Campagnolo and Jean-Sèbastien Gharbi, 77–128. Paris: Editions Mat´eriologiques.

Bergson, Abram. 1954. "On the Concept of Social Welfare." *The Quarterly Journal of Economics* 68 (2): 233–52.

Broome, John. 1993. "A Cause of Preference Is Not an Object of Preference." *Social Choice and Welfare* 10 (2): 57–68.

Dietrich, Franz, and Christian List. 2013. "A Reason-Based Theory of Rational Choice." *Noûs* 47 (1): 104–34. https://doi.org/10.1111/j.1468-0068.2011.00840.

Harsanyi, John. 1955. "Cardinal Welfare , Individualistic Ethics , and Interpersonal Com- parisons of Utility." *Journal of Political Economy* 63 (4): 309–21.

—. 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

MacKay, Alfred. 1986. ”Extended Sympathy and Interpersonal Utility Comparisons.” *Journal of Philosophy* 83 (6): 305–22.

Mongin, Philippe. 1999. ”Normes et Jugements de Valeur En Économie Normative.” *Information Sur Les Sciences Sociales* 38 (4): 521–53. https://doi.org/0803973233.

—. 2006. ”A Concept of Progress for Normative Economics.” *Economics and Philosophy* 22 (1): 19–54. https://doi.org/10.1017/S0266267105000696.

Keywords: Welfare economics, Value judgments, Interpersonal comparisons of utility, Impartial observer, Value neutrality, Economists, Economic agents

How to be (robustly) fair and just

Roland Luttens*[1], Marina Uzunova†[1], Stefan Wintein*[2], and Ruth Korte*[2 1]

Vrije Universiteit, Amsterdam – Netherlands
[2]Erasmus University Rotterdam – Netherlands

Abstract

This symposium discusses issues of fairness and justice in the context of distributive, and related, problems. Three running themes are particularly worth pointing out. First, all three papers in the symposium are concerned with the evaluation of principles for the fair or just distribution of resources-often in cases where there isn't enough. Second, a pair of the papers in the symposium is concerned with the meaning of 'equal treatment' as a requirement of fairness. Finally, another pair of the papers in the symposium is concerned with evaluating principles of fairness or justice on the basis of how robust such principles are.
In "Consistency and Aristotelian Fairness", Stefan Wintein formalises the idea of robustness with a 'consistency' axiom for a class of weighted bankruptcy problems called 'Broomean problems'. The paper then axiomatically characterises a type of a priority rule using 'consistency' and four other axioms, all of which, Wintein argues, can be justified on the basis of an Aristotelian conception of fairness.

Ruth Korte's "Not Making Anyone Worse Off in Taurek Cases and Other Allocation Problems" discusses fair distribution in the context of so called 'Taurek cases'. Motivated by the idea that, in such cases, it is not unfair to assign to an agent a higher chance of being saved when this does not make anyone worse off, Korte develops an account of fairness that can be justified on the basis of a specific 'equal treatment' principle.

In "Justice in an Inverted World", Roland Luttens and Marina Uzunova evaluate rules for resolving three types of Aristotelian justice problems on the basis of two robustness principles called 'inversion invariance' and 'universal invariance'. The paper formulates a framework for this analysis and puts six popular rules to the invariance test, only two of which, surprisingly, successfully pass it.

List of contributors:

Stefan Wintein (Erasmus University Rotterdam, The Netherlands): "Consistency and Aristotelian Fairness"

Ruth Korte (Erasmus University Rotterdam, The Netherlands): "Not Making Anyone Worse Off in Taurek Cases and Other Allocation Problems"

*Speaker
†Corresponding author: m.g.uzunova@vu.nl

Roland Iwan Luttens and Marina Uzunova (Vrije Universiteit Amsterdam, The Netherlands): ”Justice in an Inverted World”

Keywords: fairness, justice, allocation, bankruptcy problems, Aristotelian justice, Taurek cases

From Theory to Practice (and Back Again): Measuring the Value of Life and the Science of Preferences

Yam Maayan*†[1,2]

[1]Eitan Berglas School of Economics, Tel Aviv University – Israel
[2]Center for the History of Poitical Economy, Duke University – United States

Abstract

This paper explores economists' methodological debates surrounding the estimation of the value of a statistical life (VSL) in the 1970s and 1980s. Instead of addressing the ethical controversies that surrounded this practice, it examines the methodological challenges it posed: What do individuals' market choices actually reveal? What assumptions under- pin these measurements? And what were the major obstacles to establishing a rigorous approach? Drawing on published materials and archival sources, the paper investigates how economists grappled with these challenges in their efforts to establish a scientific approach to measuring preferences, using the concept of willingness to pay and market data.
While extensive methodological research has examined preference measurement in public policy, its historical development remains largely unexplored. This study argues that the debate over measuring the VSL marked a key moment in the engagement of mathematical economists with the practical challenges of quantifying subjective values for policy purposes. This new engagement shaped both the applied dimension of economics and the theoretical understanding of preferences. The paper adopts the view that economic concepts, such as preferences, are constructed within specific contexts and that the VSL debate represented a critical case where methodological and conceptual questions were deeply interconnected.

Starting in the 1960s, several economists began arguing that the proper way to measure the value of saving lives-particularly in the context of regulatory policy-was through the concept of willingness to pay. This approach was intended to replace the traditional human capital method, which estimated the cost of lost life based on earnings potential. The most influential proposal, and the one that gained the most attention at the time, was Thomas Schelling's 1966 argument for shifting the discussion from the value of an actual life to the value of risk reductions, emphasizing how individuals make decisions based on trade-offs between money and risk.

This framework became the foundation for cost-benefit analyses that evaluated life-saving regulations by quantifying the monetary value of risk reductions. By the late 1970s, agencies such as the Environmental Protection Agency had adopted the VSL model to assess regulations aimed at reducing pollution and traffic accidents.

*Speaker
†Corresponding author: yamaayan@gmail.com

Although controversial-particularly due to ethical concerns about assigning monetary values to human life-using market data to estimate the value of reducing mortality risks was widely accepted within the economics community. In fact, many economists viewed it as a more ethical alternative to the traditional human capital approach, as it relied on individuals' own valuations of human lives rather than external calculations based on earnings potential. However, the challenge of how to systematically use market data for this purpose remained largely unresolved. Even today, estimates of the VSL vary widely.

Economists' methodological debates over measuring the value of life marked a pivotal moment in the broader effort to establish preferences as measurable through market data. This debate represents a critical phase in the ongoing history of attempts to quantify preferences and utility. While historical research has examined preference measurability from various perspectives, the specific context of cost-benefit analysis remains largely unaddressed. At the same time, within the economics and philosophy communities, this context is widely recognized as highly relevant to the broader question of preference measurement. This study aims to bridge that gap by examining a key moment when mathematical economists, associated with 'high theory' approaches to economics, actively engaged with the challenge of measuring preferences for cost-benefit analysis.

This development can be seen as part of the broader empirical and applied turn in eco- nomics. In this context, the valuation of human life stands out as particularly significant. Unlike other pioneering attempts to estimate preferences in a theoretically grounded (i.e., scientific) manner-such as Daniel McFadden's work on predicting public transportation usage based on observed choice patterns-this approach explicitly used *willingness to pay* as a measure of the *monetary value individuals assign to an object.* In this case, however, the object was not just any commodity but human life itself-a uniquely complex and ethically charged subject. Over time, this practice extended beyond safety regulation to areas such as environ- mental litigation, health economics, and cultural heritage valuation, applying market-based valuation techniques to a broader range of non-market goods.

Moreover, estimating the value of life through market data-by assessing how much individuals are willing to pay to reduce their own or others' mortality risks-introduced an additional layer of complexity. It incorporated the notion of risk, raising intricate methodological questions about how to interpret the relationship between individuals' choices, their preferences, and their risk perceptions.

The paper examines the methodological debates surrounding the VSL measurement by identifying the different sub-communities involved in the discussion. It distinguishes among three groups: public finance-oriented economists, econometricians, and decision theorists, each of whom approached the issue differently, focused on distinct aspects of the problem, and relied on specific assumptions to enable measurement.

Public finance-oriented economists primarily focused on regulatory impact and policy de- sign, considering the estimation of VSL in terms of *aggregate welfare* analysis. Econometricians concentrated on the technical challenges of deriving reliable estimates from market data, particularly in determining *shadow prices*. Meanwhile, decision theorists stood apart by explicitly modeling decision-making processes to isolate individual preferences, making *preference measurability* a core methodological concern.

Each of these approaches framed different methodological concerns as the most critical in making willingness to pay a meaningful and reliable concept for policy analysis. Decision theorists, in particular, developed explicit models of rational decision-making, focusing on how individuals assess and respond to uncertainty.

Within this context, the measurement of individuals' willingness to pay to reduce risk high- lights a key theoretical distinction in decision theory-between people's preferences and their risk estimations. I argue that this distinction, in relation to value measurement, played a crucial role in the emergence of the '*cognitive critique'* of this practice. Drawing on the findings of Kahneman and Tversky, some economists began to

argue that individuals systematically misestimate risks, casting doubt on the reliability of market choices as reflections of valid preferences.

Thus, the VSL debate not only shaped cost-benefit analysis but also marked a key moment in the introduction of cognitive biases into economic discussions on choice and preferences. The paper reconstructs the early phases of this debate using archival materials from Kenneth Arrow's papers at Duke University, including records from an interdisciplinary seminar on decision theory held at Stanford in the late 1970s. This seminar, which brought together economists such as Arrow and Richard Thaler with cognitive psychologists like Amos Tversky, served as a key forum for discussions on the reliability of market-based preference measurement.

By tracing the methodological debates surrounding VSL estimation and its implications for preference measurement, this paper highlights a pivotal moment in the history of economics when theoretical constructs were tested against real-world policy applications. As debates over preference measurability persist, revisiting the VSL case provides valuable insights into the evolving relationship between economic theory, empirical practice, and public policy.

Keywords: Value of a Statistical Life, Willingness to Pay, Preferences, Cost, Benefit Analysis, Empirical Turn in Economics, Applied Economics, Economic Methodology, Decision Theory, Revealed Preferences

On the normative authority of economics

Uskali Mäki*†[1]

[1]University of Helsinki – Finland

Abstract

What I submit is something different from the traditional ways of addressing the issue of how the positive and the normative are to be defined and interrelated in the case of economics. These customary ways are mainly put in terms of the syntax and semantics of language (is-ought) or the metaphysics of what this language is about (facts-values). My proposal supplements such perspectives by focusing on the social and institutional aspects of how economics functions in society. I suggest economics plays ”unofficial” normative roles in society by acquiring and exercising what I call normative authority. Authority is an institutional attribute of the epistemic institution of economics. Normative authority is authority over normative issues that society, for example by means of economic policy, is to settle in one way or another. Yet it is not reduced to narrow normative authority merely on ”technical norms” that advise policymakers about the effective means for attaining given ends. The claim I set out to examine is that economics enjoys broad normative authority in contemporary society, on matters ranging from world view shaping to policymaking. I ask and answer questions such as What is economics' broad normative authority? What is it based on? By what pathways is it exercised?

The first question riddled by ambiguities of 'economics' that can be taken to denote things such as the academic discipline of economics as a whole, some of its ”schools” or fields or key theories; and the economics profession or its elite members, academic economists or economists of practice. Authority is a matter of influence or power to have effects (on beliefs or behaviour) that lies somewhere between coercion and mere persuasion. Broad normative authority is a power to do many kinds of thing, such as setting conceptual framings and dominant agendas, pointing out advisable policy and effective institutional design, determining what facts are relevant and what goals are feasible and worth pursuing, shaping prudent ways of thinking and acting in contemporary society, and other such things.

One might presume that were economics to enjoy high normative authority, this derives from economics enjoying supreme epistemic authority concerning matters of economic fact, but it is doubtful that this would be the case, at least in any straightforward manner; we know that economics has a mixed epistemic record.

I sketch some pathways by which the normative authority of economics is originated, enabled, exercised, and constrained. Important roles are played by framing issues in economic terms, thereby delimiting the conceptual space within which goals are set and means are chosen. This restricts the range of normative issues and normative viewpoints that can be appropriately expressed and effectively acted upon. Typical elements in such restrictive frameworks include GDP, money, market, incentives, efficiency, and other such items.

*Speaker

†Corresponding author: uskali.maki@helsinki.fi

They contribute to setting the agenda for normative considerations. The frameworks are spread and fortified by means such as economics education and interdisciplinary ”imperialism” whereby economic ways of thinking intrude into other disciplines and the larger culture or adopt a hegemonic position in relation to them. Economics has a further advantage of what I call policy tractability as its framings simplify otherwise immensely complex policy issues.

Many notions of authority imply that their exercise requires ordering, commanding or legislating someone to do something. But I pursue a weaker idea here and suggest it is enough for having normative effects to be in a position to contribute to delimiting and conceptualizing the space of options. This prepares the stage for a limited range of normative concerns and normative moves that require economic expertise for dealing with them; and this creates a self-reinforcing mechanism by which exercising authority also produces and reproduces it. To be in such a position is a function of professional status and disciplinary standing as well as institutional position in powerful policy organisations.

Finally, I point out how the normative authority of economics is potentially constrained by commonsense conceptions (both normative and descriptive) regarding economic and related issues. There are known tensions here, concerning matters of causal direction as well as justice and other moral issues. It is likely that these tensions remain chronic in that no amount of economics education will helpfully eradicate them.

Keywords: normative authority, epistemic authority, conceptual space of economics

Carroll's Tortoise Against Meta-Decision Theory

Nicholas Makins*[1]

[1]University of Leeds – United Kingdom

Abstract

Decision theories are typically considered to be theories of instrumental rationality. In particular, they are supposed to tell us which patterns of belief, preference, and choice are instrumentally rational in conditions of uncertainty. Given the range of alternative theories, and the wealth of arguments for and against each, it would be reasonable to remain uncertain about which gives the correct account of instrumental rationality. How, then, should we make choices under uncertainty about decision theory itself? In this paper, I draw a distinction between two approaches that are implicit in the literature, and argue that one of these is untenable.

One approach to decision-theoretic uncertainty is to view decision theories themselves as options between which one must choose. Because different decision theories present different strategies for making first-order choices, decision-theoretic uncertainty generates an additional, higher-order choice about which theory to follow. Before I choose whether to walk or catch the bus, I choose how to make this choice. There are different ways to approach this higher-order choice, which, again, we must choose between. The observation that there are different ways of making first-order choices leads to a proliferation of higher-order choices. Call those who think about decision-theoretic uncertainty in this way "proliferators" (Petti- grew, ms; Smith, 1991; Thorstad, 2022).

An alternative approach is to reframe first-order choices in order to incorporate one's decision theoretic uncertainty. This involves treating decision theories like states of the world and assigning them probabilistic credences. With these credences in place, we can calculate a new expected value for each of the original options: a weighted average of the values assigned to an option by each theory, where the weights are the agent's credences in the competing decision-theories. When choosing whether to walk or catch the bus, I take into account each decision theory's evaluation of each option, in a reframing of the original choice. Call those who think about decision-theoretic uncertainty in this way "reframers" (MacAskill, 2016; MacAskill et al., 2021; Nozick, 1993; Sepielli, 2014; Tarsney, 2024; Trammell, 2021).

This paper presents an argument against the reframers' strategy. The argument proceeds in two steps. The first step is to show that this strategy involves adopting instrumental rationality as an end in itself. This is established by demonstrating that, in order for decision- theoretic uncertainty to have any bearing on an agent's choice, they must assign different values to propositions that differ only with respect to the nature of instrumental rationality. The best explanation for this is that the agent being modelled non-instrumentally values instrumental rationality itself. This claim is supported by an analogy with decision-theoretic approaches to choice under moral uncertainty. It is widely accepted that these approaches involve a de dicto desire to do what is morally right, whatever that may be (Aboodi, 2017; Hedden, 2016; MacAskill et al., 2020; Weatherson, 2014). Reframers propose an analogous approach, replacing moral theories with decision theories, and thereby model agents with a de dicto desire to do what is instrumentally rational, whatever that may be.
The second step is to show that holding instrumental rationality as an end in itself is necessarily self-defeating. The argument for this claim builds on an analogue of Lewis Carroll's story of Achilles and the tortoise (Carroll,

1895). The original story demonstrates that logical reasoners fall into an infinite regress if they attempt to incorporate a rule of inference, such as modus ponens, as a premise in a deductive argument. A number of philosophers have presented modified versions of this story to demonstrate that taking instrumental rationality as an end in itself faces an equivalent regress problem (Blackburn, 1995; Brunero, 2005; Dreier, 1994, 1997; Railton, 1997; Wieland, 2013). Here, I flesh out the story in a decision-theoretic framework to show that, once an agent adopts instrumental rationality as an end in itself, they cannot arrive at a stable evaluation of the possible outcomes of a choice under uncertainty about decision theory. This precludes them from identifying which of their options maximises expected value, understood as a measure of which option is the best means to their ends. Therefore, the reframers' strategy is unable to provide an answer to the question of which option is instrumentally rational in choices under decision-theoretic uncertainty.

With this conclusion established, we return to a comparison between proliferators and reframers. Despite the above problem for the reframers' approach, proliferators face a well- known regress problem of their own (Lipman, 1991; Mongin, 2000; Pettigrew, ms.; Smith, 1991; Thorstad, 2022). However, building on the idea that not all regresses are equally problematic (Wieland, 2014), I conclude the paper by arguing that the problem facing the proliferators is the lesser of two evils.

References

Aboodi, R. (2017). One Thought Too Few: Where De Dicto Moral Motivation is Necessary. Ethical Theory and Moral Practice, 20(2), 223–237.

Blackburn, S. (1995). Practical Tortoise Raising. Mind, 104(416), 695–711.

Brunero, J. (2005). Instrumental Rationality and Carroll's Tortoise. Ethical Theory and Moral Practice, 8(5), 557–569.

Carroll, L. (1895). What the Tortoise Said to Achilles. Mind, IV (14), 278–280.

Dreier, J. (1994). Perspectives on the Normativity of Ethics. Nouˆs, 28(4), 514.

Dreier, J. (1997). Humean Doubts about Categorical Imperatives. In G. Cullity & B. Gaut (Eds.), Ethics and Practical Reason (pp. 81–99). Clarendon Press.

Hedden, B. (2016). Does MITE Make Right? On Decision-Making under Normative Uncertainty. In R. Schafer-Landau (Ed.), Oxford Studies in Metaethics Volume 11 (pp. 102–128). Oxford University Press.

Lipman, B. L. (1991). How to Decide How to Decide How to ...: Modeling Limited Rationality. Econometrica, 59(4), 1105.

MacAskill, W. (2016). Smokers, Psychos, and Decision-Theoretic Uncertainty. Journal of Philosophy, 113(9), 425–445.

MacAskill, W., Bykvist, K., & Ord, T. (2020). Moral Uncertainty. Oxford University Press.

MacAskill, W., Vallinder, A., Oesterheld, C., Shulman, C., & Treutlein, J. (2021). The Evidentialist's Wager. The Journal of Philosophy, 118(6), 320–342.

Mongin, P. (2000). Does Optimization Imply Rationality? . Synthese, 124(1/2), 73–111. Nozick, R. (1993). The Nature of Rationality. Princeton University Press.

Pettigrew, R. (unpublished manuscript). On choosing how to choose. 1–28.

Railton, P. (1997). On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action. In G. Cullity & B. N. Gaut (Eds.), Ethics and Practical Reason (pp. 53–79). Clarendon Press.

Sepielli, A. (2014). What to Do When You Don’t Know What to Do When You Don’t Know What to Do ... Nouˆs, 48(3), 521–544.

Smith, H. (1991). Deciding How to Decide: Is there a regress problem? . In M. Bacharach & S. Hurley (Eds.), Foundations of Decision Theory (pp. 194–219). Wiley–Blackwell.

Tarsney, C. (2024). Metanormative regress: an escape plan. Philosophical Studies, 181(5), 1001–1023.

Thorstad, D. (2022). Two paradoxes of bounded rationality. Philosophers’ Imprint, 22(15), 1–16.

Trammell, P. (2021). Fixed-point solutions to the regress problem in normative uncertainty. Synthese, 198(2), 1177–1199.

Weatherson, B. (2014). Running Risks Morally. Philosophical Studies, 167(1), 141–163.

Wieland, J. W. (2013). What Carroll’s Tortoise Actually Proves. Ethical Theory and Moral Practice, 16(5), 983–997.

Wieland, J. W. (2014). Infinite Regress Arguments. Springer.

Keywords: Decision theory, normative uncertainty, instrumental rationality.

Which values in modern economics? Analytical and methodological strategies for identifying non-epistemic values in economic research

Magdalena Małecka *[1]

[1]University of Copenhagen – Denmark

Abstract

The paper intends to contribute to, and expand into a novel direction, the recent discussion about the ambiguity of the concept of value in philosophical analyses of epistemic and non-epistemic values in science. In the discussion, it has been noticed that philosophers of science often use the term 'value' to refer to social and political interests, ideologies, preferences, moral and ethical commitments, and so on (Elliot 2022, Solomon 2012, Biddle 2013, Hilligardt 2022). Recently, Elliott & Korf (2024) have identified four major concepts of value in the philosophical analyses and have shown that arguments against the value-free ideal presume various concepts. While these contributions are important and they offer an illuminating nuance to the previous arguments about the role of epistemic and non-epistemic values in science (e.g., Douglas 2009, Intemann 2001, Elliott 2017), we will argue that there is yet another neglected, and related to the point about ambiguity of the concept of value, aspect of the philosophical analyses of valueladenness of scientific research.
Philosophy of science lacks a philosophical account of how exactly values are, and can be, identified in scientific research. In other words, there is no systematic analysis yet of how we can reliably state *which* values *precisely* have shaped methods, concepts, or background assumption of research. We will provide arguments for why such an account is needed and how it can be developed. Philosophical discussions mostly focus on the implications of the observation *that* values enter science. Furthermore, too often philosophers of science rely on their pre-existing intuitions or informed guesses when making claims about the exact values which may have entered science. We illustrate the point by analyzing the claims about the non-epistemic values in science. In particular, we refer to the research in economics as an example and case for our analysis.

Neoclassical approach in modern economics has been criticized by many scholars, including philosophers of science. Apart from numerous flaws in explanatory and predictive power of the neoclassical approach (see e.g., Boumans & Davis 2017 for overview), critics often point out as problematic values embedded in its formal and axiomatic frameworks: 'sexist', 'racist', and/or 'capitalist'. In particular, the so-called rational choice theory has been characterized as sexist and androgenic – due to the notion of agency it presumes (self-interested, selfish, autonomous, separate, detached agent) (e.g., Ferber & Nelson 2009, England 1989). Yet, how have the critics identified these values? Can we know for sure which values exactly have entered, for instance, the expected utility theory (EUT)? We will scrutinize the analyses of the decision theory provided by feminist economists. Based on our earlier work, we will critically assess the way in which non-epistemic values of sexism are ascribed by feminist economists to the so-called rational choice theory by pointing out that they do not

*Speaker

convincingly and accurately analyze the formal framework of theories such as EUT (Authors 2022, informed by Ross 2010, Herfeld 2020). It does not mean that non-epistemic values may not have shaped EUT, but that we need to develop a more philosophically and methodologically nuanced analysis to identify them. In the next step we will suggest how to do it.

We aim to show that some analyses of feminist philosophers of science offer us guidance on which analytical strategies are needed to reliably claim that *precise* non-epistemic values have shaped scientific research. We will provide an analysis of the work of Longino (1990), Anderson (2004), and Haraway (1989) to reconstruct their philosophical and methodological approach to the identification of *which* non-epistemic values enter science. We will show that their approaches can lead to developing a systematic philosophical account of how to single out precise types of non-epistemic values in research. We focus on non-epistemic values, as their presence in science is considered more controversial than the role of epistemic values. It is also more difficult to identify non-epistemic values in science: they are rarely stated by researchers, as contrasted with epistemic values which are occasionally mentioned explicitly as desiderata for research agendas (for instance, simplicity or broad scope). Finally, identification of *exact* non-epistemic values in scientific research allows us to identify which social and political interests shape the direction of research – these are high stakes in attempts to properly understand how and in which way *precisely* society and power structures influence science, including modern economics.

References

Anderson, E. (2004). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. Hypatia, 19(1), 1-24.

Biddle, J. (2013). State of the field: Transient underdetermination and values in science. Studies in History and Philosophy of Science Part A, 44(1), 124-133.

Boumans, M., & Davis, J. (2017). Economic methodology: Understanding economics as a science. Bloomsbury Publishing.

Douglas, H. (2009). Science, policy, and the value-free ideal. University of Pittsburgh Press.

Elliott, K. C. (2017). A tapestry of values: An introduction to values in science. Ox- ford University Press

Elliott, K. C. (2022). Values in science. Cambridge University Press.

Elliott, K. C., & Korf, R. (2024). Values in science: what are values, anyway?. Euro- pean Journal for Philosophy of Science, 14(4), 53.

England, P. (1989). A feminist critique of rational-choice theories: Implications for soci- ology. The American Sociologist, 20(1), 14-28.

Ferber, M. A., & Nelson, J. A. (Eds.). (2009). Beyond economic man: Feminist theory and economics. University of Chicago Press.

Haraway, D. J. (1989). Primate visions: Gender, race, and nature in the world of mod- ern science. Routledge.

Herfeld, C. (2020). The diversity of rational choice theory: A review note. Topoi, 39(2), 329-347.

Hilligardt, H. (2022). Looking beyond values: The legitimacy of social perspectives, opinions and interests in science. European Journal for Philosophy of Science, 12(4), 58.

Intemann, K. (2001). Science and values: Are value judgments always irrelevant to the justification of scientific claims? Philosophy of Science, 68(S3), S506–S518.

Longino, H. E. (1990). Science as social knowledge: Values and objectivity in scientific inquiry. Princeton University Press

Ross, D. (2010). The economic agent: not human, but important, in: Handbook of the Philosophy of Science, Philosophy of economics, U. Maki (ed.), Elsevier, 691-736

Solomon, M. (2012). Socially responsible science and the unity of values. Perspectives on Science, 20(3), 331-338.

Keywords: values in science, values in economics, feminist philosophy of science

When is a replication attempt successful?

Mariusz Maziarz*[1]

[1]Institute of Philosophy, Jagiellonian University [Krakow] – Poland

Abstract

Short Abstract

The replicability crisis is one of the significant problems troubling contemporary research. It started in psychology, but other researchers, including economists conducting and attempting to replicate randomized field experiments, discovered that a large proportion of such studies do not replicate. However, even though researchers across social and biomedical sciences agree that some considerably large proportion of empirical findings do not replicate and hence are purportedly false, several conflicting concepts of replication and successful replication are used in the debate. We develop existing literature concerning these questions by building on the Resampling Account of replication (Machery 2020) and Fletcher's (2021; 2022) criticism of the measures of replication success used by the Open Science Collaboration (2015) to estimate the replicability of psychological experiments. We argue that confidence intervals considered in assessing replication success should consider not only random error ( ), but also the different experimental components between the original study and replication attempts ( ). On this ground, we argue that successful replications are actually more common than is currently reported in the metascientific literature.

Fully-fledged abstract

Replication crisis denotes the problem (crisis) experienced by virtually all empirical disciplines, such as psychology, medicine, biology, and economics (Goldfarb 1997; Page et al. 2021).In the latter discipline, the replication crisis primarily affects randomized field experiments, with certain controversies, such as the deworming controversy (Ozier 2019), being widely debated in both scientific journals and the popular press. But in some cases, such as the Reinhart-Rogoff controversy, econometric studies of observational data do not replicate too (Maziarz 2017). Romero (2019) defined the replication crisis as attempts at independent repeats of the original experiment that report results that differ from the original. Replication attempts differ in the degree of similarity to the original study. Reproducibility of results means "obtaining the same numerical results when repeating the analysis using the original data and the same computer code". Reproducibility failures are usually a sign of calculation or coding error committed by researchers either conducting the original study or the replication attempt. Direct replications are identical (concerning the causally relevant factors) repetitions of the original experiment. Replication failures indicate that some un- known factors confound outcomes, the original study includes errors or questionable research practices, or random differences between trial arms have led to false-positive results. Finally, conceptual replications use modified research methods or outcome measures to test the generalizability of the original results to other contexts (Romero, 2019). However, Machery (2020) recently argued against the distinction between these two types of replications.

*Speaker

According to his Resampling Account of replication, the experimental components of the original studies are treated as random factors; a resample of which is included in a replication attempt. This account defuses the distinction between direct and conceptual replication. Similarly, Feest (2019) suggested that failed conceptual replications are not necessarily a sign of the original result emerging from mistakes or errors but result from replication attempts differing from the original studies. She argued that the differences between studies account for the heterogeneity of reported results.

The rate of failed replication attempts varies across disciplines, but even though researchers across social and biomedical sciences agree that some considerably large proportion of empirical findings do not replicate and hence are purportedly false, several conflicting concepts of replication and successful replication populate the debate. Fletcher (2021) criticized ways of assessing whether a replication attempt reports a result similar to the original study used by Open Science Collaboration: subjective assessment, null hypothesis significance testing, comparing effect sizes, comparing the original effect size with the replication confidence interval, and meta-analysis. He later (Fletcher 2022) argued that the replicability should be tackled with meta-analysis, which only allows for correctly analyzing result reliability and confirmation. While we agree that meta-analytic estimates are (at least in principle) the best estimates of the actual treatment effects, Fletcher's (2022) approach remains uninformative concerning whether a replication attempt of the same research question reports an average treatment effect that agrees with the original study. To solve this problem and allow for a meaningful discussion of replication success, we develop this concept in a way inspired by the random effects model of meta-analysis.

We start by developing the successful replication concept for an ideal case of an exact replication. In such a case, the difference in results (assuming both studies being free from errors, questionable research practices, or outright frauds) emerges only due to random error (). However, considering that the two studies report only estimates of the true average treatment effect ( ), while the true ATE may be different and falls within the confidence interval with a certain probability (e.g., the standard 95% confidence interval), it suffices for a successful replication study to report confidence interval that overlaps with the confidence interval reported by the original study. In other words, only disjunctive confidence intervals of the original study and replication exclude (with 95% probability) the possibility of the intervention tested in the two studies having the same effect size.

The Resampling Account of replication (Machery 2020) treats various aspects of experimental design as random. Assuming an additive model of interactions between these components (which is a conservative assumption given our purpose and the plausibility of multiplicative interactions), we argue that the distribution of the joint effect of different research designs and statistical analysis pipelines can be modeled as having normal distribution and null average. Such an assumption resembles the random effects model of meta-analysis, where the heterogeneity of reported average treatment effects results not only from random errors but also from differences among studies. We argue that accounting not only for the sampling error () but also the differences in study design and statistical analyses () leading to (much) wider 95% confidence intervals that should be used to assess replication success correctly.

Based on this, we conclude that replication success is actually more common than it is believed. The crisis claims result from underappreciating the heterogeneity of reported average treatment effect estimates resulting from differences in design and statistical analysis between original studies and their replications.

References

Feest, U. (2019). Why replication is overrated. Philosophy of Science, 86(5), 895-905.

Fletcher, S. C. (2021). The role of replication in psychological science. European Journal for Philosophy of Science, 11(1), 23.

Fletcher, S. C. (2022). Replication is for meta-analysis. Philosophy of Science, 89(5), 960- 969.

Fletcher, S. C., Jones, G., & Rothman, A. (2021). Discussion: What Is a Replication?. Goldfarb, R. S. (1997). Now you see it, now you don't: emerging contrary results in economics. Journal of Economic Methodology, 4(2), 221-244.

Machery, E. (2020). What is a replication?. Philosophy of Science, 87(4), 545-567.

Maziarz, M. (2017). The Reinhart-Rogoff controversy as an instance of the 'emerging contrary result'phenomenon. Journal of Economic Methodology, 24(3), 213-225.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.

Ozier, O. (2019). Replication Redux: The Reproducibility Crisis and the Case of Deworming. World Bank Policy Research Working Paper, (8835).

Page, L., Noussair, C. N., & Slonim, R. (2021). The replication crisis, the rise of new research practices and what it means for experimental economics. Journal of the Economic Science Association, 7(2), 210-225.

Romero, F. (2019). Philosophy of science and the replicability crisis. Philosophy Compass, 14(11), e12633

Keywords: replication crisis, randomized field experiments, replicability, econometrics

Causal presuppositions in econometrics and the science-policy gap

Luis Mireles-Flores*[1]

[1]University of Helsinki (TINT) – Finland

Abstract

During the last three decades, economics has seen an increasing emphasis on the use of empirical methods of causal inference as the most suitable and reliable basis to answering policy related questions. This trend has been called ”empirical turn” or ”credibility turn” in economics (Angrist & Pischke 2010), because the scientific results obtained are said to be more reliable and ”credible” than the abstract and theory-based reasoning in most mainstream economics of the late 20th century.

The specific set of methods of causal inference promoted by the credibility turn in eco- nomics are ‘design-based empirical methods’, which try to emulate the inferential properties of an ideal experimental design (Holland 1986; Morgan & Winship 2007; Cohen & Easterly 2009; Banerjee & Duflo 2011). Examples of these methods are: RCTs, field experiments, instrumental variables, difference-in-differences, regression discontinuity, matching techniques, synthetic controls, and the like (Angrist & Pischke 2015; Huntington-Klein 2022).

The proliferation of design-based techniques within empirical economics is often defended as the best means to producing research of high policy value (e.g., Duflo 2017). Since the research results are said to be grounded on the most reliable methods of causal inference, then the policy prescriptions based on such scientifically well-established causal evidence must allegedly be much more reliable and effective as well.

The increasing dominance of design-based empirical methods in economics has received several criticisms from different fronts (e.g., Cartwright & Hardie 2012; B´ed´ecarrats, et al. 2017; Kvangraven 2020). The criticisms suggest that design-based methods produce evidence that does not necessarily match the kind of evidence that policy making actually requires to be effective. I refer to this mismatch as the ”science-policy evidential gap”.

In this article, I dig into the foundations of the design-based methods of causal inference to uncover the background assumptions that, I argue, can explain the science-policy evidential gap. I suggest that understanding the presuppositions in the empirical research that lead to the evidential gap will shed some light on possible ways to bridge it.

The main claim of the article is that the methods of causal inference employed in empirical economics always come with built-in causal commitments and presuppositions (for short, background assumptions), which can have significant and varied influence on the meaning, reliability, and policy relevance of the estimated causal effects.

More specifically, I argue that uncovering the background assumptions of these methods contributes to understanding at least three ways in which the causal results are limited, restricted, or insufficient to properly fulfil the evidential requirements of policy making:

First, the focus of the framework is meant, by design, to investigate the effects of alter- native causes, and not to discover what the causes of given effects are. But the effectiveness of policies and the policy process requires at various stages information about the causes of phenomena.

Second, the characterization of the estimators of causal effects involves statistical, methodological, and ontological presuppositions which restrict the kinds of ”things” that can be considered as causal relata, as well as the kinds of relations that can be taken as causal.

Third, even when the empirical research includes robustness analyses of the causal effects relative to variations of relevant contextual factors (such as sensitivity or mediating analyses), qualitative/narrative evidence about confounding factors that can be relevant to the policy process is typically ignored and left out of the studies.

These three aspects concerning the background presuppositions of design-based methods enable and constrain the inferences that can and cannot be made exclusively based on the posited empirical evidence. Moreover, bringing to the forefront these background assumptions will make clear how the causal questions that end up being researched tend to be those that can be framed in accordance to the specifications of the design-based causal frame- work. Consequently, the causal effects that are investigated are selected more in line with the background assumptions of the methods than with the actual questions and problems that policy makers confront. I will illustrate the three issues mentioned above with examples from empirical research in economics that have been presented as reliable evidence for public policy prescription.

References

Angrist, J. D. & J.-S. Pischke. 2010. The credibility revolution in empirical economics. Journal of Economic Perspectives, 24 (2): 3-30.

Angrist, J. D. & J.-S. Pischke. 2015. Mastering ’metrics. Princeton University Press. Banerjee, A. V. & E. Duflo. 2011. Poor economics. Public Affairs.

Béd écarrats, F., I. Guérin & F. Roubaud. 2017. All that Glitters is not Gold: The Political Economy of Randomized Evaluations. Development and Change 50(3): 735–762. Cartwright, N., and J. Hardie. 2012. Evidence-based policy: a practical guide to doing it better. Oxford University Press.

Cohen, J., and W. Easterly. 2009. What works in development? Brookings Institution.

Duflo, E. 2017. The Economist as Plumber. American Economic Review, 107 (5): 1-26. Holland, P. W. 1986. Statistics and causal inference. Journal of the American Statistical Association, 81(396), 945-960 Huntington-Klein, N. 2022. The Effect. CRC Press.

Kvangraven, I. H. 2020. Nobel Rebels in Disguise: Assessing the Rise and Rule of the Randomistas, Review of Political Economy, 32 (3): 305-341.

Morgan, S. L., and C. Winship. 2007. Counterfactuals and causal inference. CUP.

Keywords: causal inference, policy relevance of economics, potential outcomes, econometrics, evidence, public policy

Original Positions and Decision Theory

Michael Moehler*[1]

[1]Virginia Tech – United States

Abstract

Significant renewed interest in Rawls's (1971, 2001) original position has emerged, in particular regarding the choice of rational agents behind the veil of ignorance and defense of the difference principle.(1) Specifically, new arguments have been presented that rational agents in the original position would choose (i) utilitarianism (Chung 2023), (ii) weighted- rank utilitarianism (Buchak 2017), or (iii) distribution-sensitive utilitarianism (Stefánsson 2021) as a distributive principle.
The newly presented arguments either show increased sophistication regarding the decision- theoretic considerations behind the veil of ignorance, including subtle changes to the assumed rationality and preferences of agents, or they attempt to bypass such considerations. In sub- stance, however, no real progress seems to have been made in this debate since the seminal dispute between Rawls and Harsanyi,(2) because the newly presented arguments all reach different conclusions. This plurality of moral outcomes seems especially problematic because, according to common understanding, the main reason for the introduction of the original position as an analytic device was precisely to justify, uniquely and unambiguously, one set of principles of justice.

In this article, I clarify that the newly presented arguments suffer from the same problem as the seminal discussion, which quickly focused on consideration of formal methods rather than moral theory. If this focus is shifted, then it becomes evident that original position arguments, old and new, implicitly or explicitly, simply make different moral assumptions regarding the choice of rational agents in the original position. The different moral assumptions captured by different formalizations define different versions of the original position that constrain the choice, or the choice of the normative decision rule, behind the veil of ignorance, and consequently lead to different moral conclusions. From a moral point of view, the arguments refer to different original positions that are valid for different moral worlds.

Ultimately, moral assumptions, not formal methods, play the essential role for the justification of principles of justice in the original position. As such, in an important way, the arguments have it backwards. The significance of original position arguments, especially for contemporary diverse societies, lies not in defending a unique moral outcome as a result of the application of normative decision theory, but rather in justifying the moral assumptions that inform the design of the original position, and their best formalization.(3) Understood in this way, original position arguments are highly valuable and have the potential to advance moral discourse. Thus understood, original position arguments allow agents to understand more clearly and precisely where their moral disagreements lie and the implications of such disagreement for the demands of justice.

*Speaker

My argument proceeds as follows. In Section 2, I discuss the features of Rawls's original position that are relevant for the Rawls–Harsanyi dispute. This discussion reveals that, from a moral perspective, Rawls and Harsanyi refer to different original positions. In Section 3, I focus on Chung's (2023) argument that defends utilitarianism. I argue that Chung's argument does not contest Rawls's conclusion for the difference principle. Rather, it makes different moral assumptions and consequently justifies a different moral outcome. In Section 4, I make a similar point regarding Buchak's (2017) and Stefánsson's (2021) original position arguments to which Chung refers in his discussion.

Section 5 concludes. It stresses the nature of my argument that aims to highlight how the works discussed can be productively embedded in a broader discussion, if they are not narrowly conceived to be about Rawls's original position argument and its correctness. The section highlights the importance of original position arguments for morally diverse societies and the need for genuine interdisciplinary analysis that combines considerations of decision theory with moral philosophy.

References

Bradley, Richard. 2022. "Impartial Evaluation under Ambiguity." *Ethics* 132 (3): 541-569.

Bradley, Richard, and H. Orri Stefánsson. 2023. "Fairness and Risk Attitudes." *Philosophical Studies* 180: 3179-3204.

Buchak, Laura. 2017. "Taking Risks behind the Veil of Ignorance." *Ethics* 127 (3): 610-644.

Chung, Hun. 2023. "When Utilitarianism Dominates Justice As Fairness: An Economic Defence of Utilitarianism From the Original Position." *Economics & Philosophy* 39 (2): 308- 333.

De Coninck, Thijs, and Frederik Van de Putte. forthcoming. "Original Position Arguments: An Axiomatic Characterization." *Economics & Philosophy*.

Gaus, Gerald, and John Thrasher. 2016. "Rational Choice and the Original Position: The (Many) Models of Rawls and Harsanyi." In *The Original Position*, edited by Timothy Hin- ton, 39-58. Cambridge: Cambridge University Press.

Gläser, Micha. 2023. "Nozick on the Difference Principle." *Politics, Philosophy & Eco- nomics* 22 (2): 126-159.

Gustafsson, Johan E. 2018. "The Difference Principle Would Not Be Chosen behind the Veil of Ignorance." *Journal of Philosophy* 115 (11): 588-604.

Harsanyi, John. 1953. "Cardinal Utility in Welfare Economics and in the Theory of Risk- taking." *Journal of Political Economy* 61 (5): 434-435.

Harsanyi, John. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (4): 309-321.

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

Karni, Edi, and John Weymark. 2024. "Impartiality and Relative Utilitarianism." *Social Choice and Welfare* 63: 1-18.

Muldoon, Ryan, Chiara Lisciandra, Mark Colyvan, Carlo Martini, Giacomo Sillari, and Jan Sprenger. 2014. "Disagreement Behind the Veil of Ignorance." *Philosophical Studies* 170 (3): 377-394.

Rawls, John. 1971. *A Theory of Justice*. 1999 revised ed. Cambridge: Harvard University Press.

Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge: Harvard University Press.

Stefánsson, H. Orri. 2021. "Ambiguity Aversion Behind the Veil of Ignorance." *Synthese* 198: 6159-6182.

Zhao, Michael. 2023. "Ignore Risk; Maximize Expected Moral Value." *Noûs* 57 (1): 144-161.

Footnotes

(1) See Buchak (2017), Gustafsson (2018), Stefánsson (2021), and Chung (2023) for examples. See Bradley (2022) and Bradley and Stefánsson (2023) for a more general discussion of rational choice under risk, ambiguity, and fairness considerations associated with the veil of ignorance and/or an impartial observer. See Karni and Weymark (2024) for an axiomatization of relative utilitarianism that supplements Harsanyi's social aggregation theorem with an impartiality axiom, and De Coninck and Van De Putte (forthcoming) for a general axiomatic analysis of original position arguments from a social choice perspective. See Zhao (2023) for discussion of moral choice under uncertainty across time. See Gläser (2023) for discussion of the difference principle.

(2) For the seminal dispute, see Rawls (1971, 2001) and Harsanyi (1953, 1955, 1975). For a summary of central aspects of the dispute, see Gaus and Thrasher (2016).

(3) For a productive, formal discussion of moral disagreement behind the veil of ignorance, see Muldoon et al. (2014).

Keywords: Original positions, decision theory, veil of ignorance, impartiality, impersonality, equality, difference principle, utilitarianism

Psychological narratives in decision theory: what they are and what they are good for

Ivan Moscati*†[1]

[1]University of Insubria, Department of Economics – Italy

Abstract

Decision theory is a branch of economics that investigates how individuals make decisions in situations of uncertainty. Broadly speaking, these are situations in which the outcome of a decision depends on whether some event occurs. For instance, the outcome of the decision to buy a lottery ticket depends on whether the purchased ticket is drawn, and the outcome of the decision to undergo surgery depends on whether the operation is successful.
Theories of decision-making are typically accompanied by narratives about the psychological determinants of economic behavior in situations of uncertainty. These include the desire to gain money or the fear of losing it; the aversion, or attraction, to risk; perceptions of event likelihood; and concerns about the potential disappointment if a decision proves mistaken. Intuitions about these psychological determinants are formalized using technical concepts and tools such as utility functions, preference relations, probability measures, and decision weights. A second layer of narratives accompanying decision theories concerns the way in which the single psychological determinants are combined into a decision mechanism. Typically, this happens via mathematical operations, such as multiplication, summation, maximization, or integration.

Consider, for example, the canonical model in decision theory, Expected Utility theory (EU). According to the psychological narrative associated with EU, two main factors drive economic behavior under uncertainty: the intensity of the individual's desires for outcomes and her beliefs about the likelihood of those outcomes. EU represents the intensity of desires by a utility function $u(x)$ and beliefs by a probability function $\pi(x)$. The two are then combined via multiplication and summation: the individual weighs the desirability $u(x)$ of each outcome by multiplying it by its probability $\pi(x)$, and chooses the option for which the weighted sum $\sum u(x)\pi(x)$ is highest.

This paper discusses the psychological narratives that typically accompany decision theories. Such narratives refer to both the single psychological determinants of economic behavior and the manner in which these determinants are combined into a decision mechanism to produce a choice. The paper offers two contributions.

First, beginning in the late 1970s, the accumulation of robust experimental evidence against EU prompted decision theorists to advance a number of alternative models. This paper reviews the psychological narratives accompanying ten influential decision theories proposed to supersede EU. Specifically, it focuses on the two most prominent strands of research in the field: decision-making under risk and decision-making under ambiguity. The post-EU theories considered here include Prospect Theory, Rank-Dependent Utility theory, Regret Theory, Disappointment Theory, Cumulative Prospect Theory, the Three-Moments model, Salience Theory, the Choquet Expected Utility theory, the Maxmin Expected Utility theory, and the Smooth Model of ambiguity. While several good reviews of these decision models exist, they primarily emphasize the mathematical aspects of the models, such as their axioms and theorems. In contrast, this paper focuses on the psychological narratives embedded in these models, making them accessible even to readers unfamiliar with the post-EU decision theory literature.

The second contribution of the paper is methodological. It clarifies the status and functions of psychological narratives in decision models, addressing the following questions: Do these narratives draw from scientific psychology, or are they primarily based on introspection and folk psychology? Should they be understood as

realistic representations of actual psychological processes, or are they better interpreted as “as-if” stories in the sense of Friedman's well-known methodological perspective?

This paper argues that the psychological narratives in decision theory – both in EU and post-EU models – primarily draw from introspection and folk psychology rather than from scientific psychology. Furthermore, it defends the view that these narratives are best understood as “as-if” narratives.

If this interpretation is correct, however, an apparent paradox arises: Why are psychological narratives so prevalent in decision theory? Why do decision theorists not simply put forward mathematical decision models, deduce their empirically observable implications, and test their descriptive validity without invoking psychological explanations? In other words, what epistemic functions do psychological narratives perform in decision-theoretic modeling?

Answering this question requires recognizing that, despite their potentially fictional nature, psychological narratives have two primary functions in decision theory. First, in a situation where experimental studies have failed to narrow the set of competing post-EU models, these narratives serve as rhetorical devices that enhance a model's perceived validity. Second, and more importantly, psychological narratives provide explanations for choice-behavior phenomena, and this despite their fictional nature. Recognizing this explanatory function requires reconsidering the traditional notion of scientific explanation and, in particular, abandoning the assumption that truth is a necessary condition for a valid explanation.
Acknowledging these two functions of psychological narratives does not license freestanding storytelling within decision theory. On the contrary, it allows us to move beyond a psychologically realistic but naive interpretation of these narratives, and to recognize that the psychological drives and mechanisms posited by post-EU models are often far less plausible and straightforward than they initially appear. This awareness, in turn, serves as a caution against the unchecked proliferation of decision models witnessed over recent decades.

Keywords: Choice under risk, Choice under ambiguity, Non, expected utility theories, Psychological narratives, Scientific explanation

Design principles for the IPCC emission scenario ensemble

Francesco Nappo*[1]

[1]Politecnico di Milano – Italy

Abstract

An important part of Working Group III's (WGIII) role in the cyclical assessments by the Intergovernmental Panel on Climate Change (IPCC) is the collection and evaluation of quantitative emission scenarios produced by research institutions worldwide (Riahi et al. 2022). Assessed scenarios link projections about socio-economic systems and technology roll- out with the changing climate. As such, they aim to address crucial questions about the space of intervention for public policies as well as the intensity of efforts needed to achieve mitigation goals: are current regulations and pledges likely to limit global warming in this century to 1.5◦ above baseline, and in any case below 2◦, as set by the Paris Agreement? How would stricter policies affect not only energy supply and demand, but also land use, food security, transportation, and the provision of other services? Conversely, how would climate inaction affect the same variables in the medium and long term?

Over time, WGIII has developed its own practices for assembling and reflecting on collected scenarios. They differ in important ways from the ensemble methods adopted by Working Group I (WGI) for physical climate science, which have received philosophical coverage (Parker 2010; Katzav et al. 2012; Jebeile & Crucifix 2020; Harris & Frigg 2023). WGIII's latest assessments adopt an *unstructured* ensemble approach, whereby scenarios from the scientific community are collected with no pre-determined design (Guivarch et al. 2021). Besides their distinctiveness, WGIII's assessments are a worthy object of philosophical interest for several reasons. From a theoretical viewpoint, they represent an unprecedented effort at synthesizing outputs of models at a crucial juncture of the science-policy interface. They are, accordingly, a case study of high relevance for philosophical discussions on objectivity and science's role in policy. On a more practical level, the evolving demands of policymakers paired with constraints on model-making generate the conditions for expectation mismatch and the need for critical evaluation. As preparations for the next cycle of IPCC assessment continue, it is important to build capacity to critically assess the legitimacy of demands on ensemble construction and the uses of scenario assessment information in public policy.

Echoing IPCC's recent call for a "science of scenario assessment" (IPCC Workshop in Bangkok, IPCC 2023:4), this presentation will provide a philosophical analysis of WGIII's ensemble methods. An important outcome of the discussion is to connect the evolution of WGIII's unstructured ensemble approach to the distinctive requirements placed upon the IPCC's mandate to synthesize 'policy-relevant but policy-neutral' (IPCC 2024) information, distinguishing WGI and WGIII's ways of interpreting that mandate. Whereas WGI sets up a structured ensemble of climate simulations to control the number and type of simulations

*Speaker

considered in its ensemble exercise, thereby mitigating the problem of over-representation of modeling frameworks, WGIII has gradually turned the main function of its ensemble to that of offering a vetted collection of scenario exercises produced by the scientific community, thereby reducing the involvement of WGIII authors in the ensemble design process. As will be discussed in the course of the presentation, embracing a minimal form of what Parker (2010) calls 'ownership' of the ensemble and included scenarios bears on the way WGIII experts assess their own database of emission scenarios. In particular, the focus will be on the implications of WGIII's unstructured ensemble approach for the type and epistemological significance of agreement across models, which differs from the conceptualizations of 'con- vergence' and 'robustness' that one finds in the epistemological literature on the physical climate sciences.

Having noted the distinct features of WGIII's ensemble methods, a brief argument will be sketched to the extent that WGIII should move past the unstructured ensemble approach currently in use. The argument is based on three interlocking considerations. The first, 'missed learning', is that recent advances in formal assessment techniques allow for the identification of robust policy options beyond what is allowed by the descriptive statistics in use in AR6 (Dekker et al. 2023; Al Khourdajie et al. 2024; Giarola et al. 2024). The second, 'missed linkages', is that a more active role by WGIII in the design of the ensemble would also help connecting the long-term scenarios produced by the integrated assessment modeling (IAM) community with the national energy system modeling and other scenario approaches (Koberle et al. 2022). The third, 'shared ownership', is that a more structured scenario ensemble can be achieved without the IPCC's taking ownership of it, e.g., by opening up the ensemble database to the community through a live database approach or by sharing recommendations about what may be missing in the ensemble after initial submissions. Having laid out the case for moving past an unstructured ensemble approach, we conclude with some notes of warning with regards to the significance of formal analyses of scenario robustness for policy.

References

Al Khourdajie, A., Skea, J., Green, R. 2024. Climate ambition, background scenario, or the model? Attribution of the variance of energy-related indicators in global scenarios. *Energy and Climate Change*. 5: 100126.

Dekker, M. Andries, H., van den Berg, M., Daioglou, V., van Heerden, R., van der Wijst, K. van Vuuren, D. 2023. Spread in climate policy scenarios unravelled. *Nature*. 624:309-316.

Giarola, S., Chiani, L., Drouet, L., Marangoni, G., Nappo, F., Muttarak. R., and Tavoni, M. forthcoming. Underestimating demographic uncertainties in the IPCC syn- thesis process. *npj Climate Action*. (Pre-print available on ResearchGate)

Guivarch, C., Le Gallic, T., Bauer, N., Fragkos, P., Huppmann, D. et al.. Using large ensembles of climate change mitigation scenarios for robust insights. *Nature Climate Change*. 12: 428-435.

Harris, M, and Frigg, R. 2023. Climate models and robustness analysis – part I: Core concepts and premises. In: Pellegrino, G., Di Paola M. (eds.). *Handbook of Philosophy of Climate Change*. Cham: Springer.

IPCC 2023. IPCC Workshop on the Use of Scenarios in the Sixth Assessment Report and Subsequent Assessments. Bangkok, Thailand. 25-27 April.

IPCC 2024. *The Intergovernmental Panel on Climate Change*. Retrieved from the IPCC website: https://www.ipcc.ch

Jebeile, J. and Crucifix, M. 2020. Multi-model ensembles in climate science: Mathematical structures and expert judgments. *Studies in History and Philosophy of Science*. 83: 44-52.

Katzav, J., Dijkstra, H., de Laat, A. 2012. Assessing climate model projections: state of the art and philosophical reflections. *Studies in History and Philosophy of Modern Physics*. 43(4): 258-276. Koberle, A., Daioglou, V., Rochedo, P., Lucena, A., Szklo, A.,

Fujimori, S. et al. 2022. Can global models provide insights into regional mitigation strategies? A diagnostic model comparison study of bioenergy in Brazil. *Climatic Change*. 170:2.

Parker, W. 2010. Whose probabilities? Predicting climate change with ensembles of models. *Philosophy of Science*. 77: 985-997.

Riahi, K., Schaeffer, R., Arango, J., Calvin, K., Guivarch. C. et al. 2022. Mitigation pathways compatible with long-term goals. In IPCC, 2022: *Climate Change 2022: Mit- igation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Inter-governmental Panel on Climate Change* (eds. Shukla, P. R. et al.).

Cambridge University Press. Zimm, C., Mintz-Woo, K., Brutschin, E., Hanger-Kopp, S., Hoffmann, R., Kikstra, J., Kuhn, M., Min, J., Muttarak, R., Pachauri, S., Patange, O., Riahi, K., and Shinko, T., 2024. Justice considerations in climate change research. *Nature Climate Change*. 14: 22-30.

Keywords: climate economics, emission scenarios, multi, model ensembles, science, policy

Should we really worry about Arrow? On using the axiomatic method in normative inquiry

Ittay Nissan-Rozen*[1] and David Enoch

[1]Hebrew University of Jerusalem (HUJ) – Jerusalem 91905, Israel

Abstract

The axiomatic method is widely used in normative economics (as well as in other disciplines, such as philosophy). In this paper we discuss a methodological question having to do with the proper use of this method. The question is best framed in the context of impossibility theorems. However, it applies also to possibility theorems because every surprising possibility theorem can be reformulated as an impossibility theorem, by adding an axiom that rules out whatever it is which is surprising about the theorem.
Most impossibility theorems show that a set of normative conditions cannot all be jointly satisfied in some cases that belong to a family of cases in which one would (presumably) expect all the conditions to hold. The theorems do not usually show that the conditions always conflict; only that they can conflict (sometimes an "unrestricted domain" axiom is added to make this explicit, and sometimes the commitment to the axioms in all cases in the relevant family of cases stays implicit). The question we explore is whether - and in what sense - the mere possibility of a conflict is problematic in cases in which the conflict does not actually arise?

We start with the story of Charles Chiharas's (1979) Sec-Lib Paradox, the case of a club that admits as members all and only club secretaries whose clubs won't accept them as members. We take it to show – exclusively – that sometimes incoherence (i.e., the *possibility* of a conflict between a set of conditions) is just not a problem at all.

We then move to consider Arrow's (1950) theorem (under different interpretations), in the context of which we argue for three points: First, the mere incoherence need not pose a worrying problem, if we are unlikely to be in circumstances in which it arises (to argue for that we use the example of Black's (1948) Median Voter Theorem). Second, we note ways – normative and epistemic – in which the incoherence may, nevertheless, be worrying even if we are unlikely to be in the circumstances in which it's directly relevant. Lastly, we note how, as a result of the previous points, merely showing an incoherence result is never sufficient by itself to generate a serious problem, so that the job of someone putting it forward is not done unless they also show that we should be worried for one of the reasons we point to (or another reason).

From a normative point of view, in some contexts, it may perhaps be argued that if there are any circumstances in which a set of conditions conflict then some of these conditions cannot be normatively significant in other circumstances as well. Perhaps, for instance, there's an important distinction between merely rules, where an incoherence is only a problem if the

*Speaker

circumstances that give rise to it are likely to occur, and *ideals*, which run deeper, and which must, in order to be ideals at all, be coherent across all possible circumstances, even those that are unlikely to arise (think, for instance, of Arrow-style worries about the idea of "the will of the people", or other group attitudes). We do not rule out the possibility that *in some cases* such considerations apply. Our point is that this should be argued for, in the context of those specific cases, and not merely assumed.

From an epistemic point of view, an impossibility result may require, in some cases, a reevaluation of what, before learning the result, seemed so clearly plausible and perhaps this reevaluation should reduce one's confidence in one's relevant pre-theoretical intuitions (those that led one to accept the axioms in the first place). The intuitions, after all, seem to support the axioms in all cases, not only in those cases in which the axioms do not conflict. Thus, it might be that one of the intuitions is misleading (also in cases in which the axioms do not conflict).

In such cases one can either outright desert one of the axioms, or replace it with a qualified version thereof, one that is especially designed to avoid the impossibility result. Sometimes the latter option may be a legitimate one, and when it is, this will be a way of containing the impossibility result. But often it won't be a reasonable response, but rather a paradigm of ad-hoc reasoning, motivated by the need to save a theory more than by relevant substantive and evidential concerns.

In the last part of the paper, we discuss, in light of our conclusions, the case of Harsanyi's (1955) Utilitarian Theorem. Harsanyi's theorem is not an impossibility result, but rather a possibility result (it shows that under certain conditions social preferences can be represented in a particular way). However, the reason that many scholars found it interesting (and – in many cases – disturbing and shocking) is the set of possibilities that it *rules out:* it rules out a very minimal commitment to egalitarianism. Thus, the theorem can easily be formulated as an impossibility result. This makes it possible for us to apply the lessons learnt from our discussion of Arrow's theorem, to Harsanyi's argument. We consider different ways, some normative and some epistemic, in which Harsanyi's argument can be understood and show that each one of them is based on some substantial normative or epistemological assumptions that may be rejected.

References

Arrow, K.J. (1950), A Difficulty in the Concept of Social Welfare, *Journal of Political Economy*, 58: 328–346.

Black, D. (1948), On the Rationale of group decision making, *Journal of Political Economy*, 56:1, 23-34.

Chihara, C. (1979), The Semantic Paradoxes: A Diagnostic Investigation, *The Philosophical Review,* 88:4, 590-618.

Harsanyi, C. J. (1955), Cardinal Welfare, Individualistic Ethics, and Interpersonal Com- parisons of Utility, *Journal of Political Economy,* 63:4, 309-321.

Keywords: Normative Economics, Impossibility Theorems, Arrow's Theorem, Harsanyi

Competing Claims with Uncertainty

Makan Nojoumian*[1]

[1]University of Edinburgh – United Kingdom

Abstract

Suppose we had to choose between helping one or the other of two patients with severe health conditions, but could not help both. Is it overall better to significantly improve the well-being of the patient with a severe condition, or to moderately improve the well-being of the patient with a very severe condition instead?
The answer to this question naturally depends on the exact well-being levels and improvements referred to, but it also depends on our broader views about the value of total well-being improvements, priority to the worse off, and relative equality. Which view we adopt in the background ('utilitarian', 'prioritarian' or 'egalitarian') has serious implications for deciding in this problem as well as many real-world moral questions and policy decisions that can be characterised in this way.

One intuitive and popular approach to the problem above in the philosophical literature is to think in terms of the 'competing claims' of the two patients and ask: whose claim on our aid is stronger? (See, for example: Nagel 1979, Parfit 1997, Otsuka and Voorhoeve 2009, Voorhoeve and Fleurbaey 2012, Adler 2019, and Adler 2022.) Does the patient with a severe condition have a stronger claim because they will receive a significant improvement in well-being? Or does the patient with the very severe condition have a stronger claim because they are worse off and have a lower well-being level? Beside its intuitive appeal, one attraction of this approach is that it does not compare the outcomes in a so-called 'imper- sonal' way, purely in terms of intrinsic properties of the resulting distributions of well-being. Instead, it is a 'person-affecting' approach that looks at how each person is affected by the two options, what their competing claims are, and weighs these against each other. And such a person-affecting approach in turn provides a powerful way to justify priority to the worse off and relative equality.

This paper addresses two sets of questions about competing claims. Firstly, how should we formalise the claims of individuals? Does the framework for defining claims and weighing them against each other depend on our background views (on utility, priority or equality)? And how should we decide what claims individuals have under conditions of uncertainty where their well-being levels under each prospect are not known? Should their claims be based on their expected well-being levels under each prospect so that they have one over- all claim ('expected claims')? Or should they be based on their final well-being levels in each outcome of the prospect so that they have multiple claims, one in each possible state of affairs ('final claims')? Or should they have both kinds of claims or a hybrid version of them?

In answer to this first set of questions, I first present a framework under conditions of certainty in which each individual has a claim in favour of one or both prospects with a numerical strength, and in which one prospect

*Speaker

is better than its alternative if the sum of claims in its favour is higher than the others'. In the philosophical literature, the underlying framework for discussing claims and comparing prospects often depends on views about utility, priority or equality. A virtue of the unified framework presented here is that even though the strength and valence of claims depends on our background views, the underlying framework itself does not. So the debate about utility, priority or equality can be rephrased as a debate about the strength and valence of claims.

I then explore extensions of this framework to conditions of uncertainty and present new arguments against using expected claims: Insensitivity to Final Outcomes, Dominance, and Expectational Levelling Down. I also show that hybrid views suffer from similar problems. Since final claims do not face these disadvantages, I do not agree with the widely held view that competing claims under uncertainty require expected or at least hybrid claims. Rather, I argue that under conditions of uncertainty, each individual has a final claim for one prospect or both in each possible state of affairs; the strength and valence of the claims depend on our views about utility, priority or equality; and one prospect is better than its alternative iff it has the higher sum of claims in its favour when weighted by the probability of the states in which the claims arise.

The second set of questions is: how do we decide the strength and valence of claims? Should we adopt a utilitarian, prioritarian or egalitarian approach? And in particular, is priority for the worse off a non-starter? These questions are interestingly related to the previous ones. Under a utilitarian approach, the final and expected claims of individuals happen to coincide. A result of Harsanyi's (1955) famous interpersonal aggregation theorem is that only utilitarianism has this property, and that prioritarianism and egalitarianism each result in different final and expected claims. This has in turn led to influential arguments against priority and equality under uncertainty. (See, for example: Broome 1991, McCarthy 2008.) Interestingly, other influential arguments claim that whereas equality does not face any problems under uncertainty, priority to the worse off is a non-starter. In the final section of this paper, I address two highly influential and widely discussed arguments against priority un- der conditions of uncertainty – Expected Well-Being Maximization and the Separateness of Persons – and give new responses to each. (For the objections, see Otsuka and Voorhoeve 2009, Voorhoeve and Fleurbaey 2012, Otsuka 2015, and Otsuka and Voorhoeve 2018.) I conclude that priority and equality with final claims are both highly plausible and attractive approaches to weighing the competing claims of individuals under conditions of uncertainty.

References

Adler, Matthew D. 2019. *Measuring Social Welfare: An Introduction* (Oxford University Press)

Adler, Matthew D. 2022. 'Claims Across Outcomes and Population Ethics', in Gustaf Arrhenius, and others (eds), *The Oxford Handbook of Population Ethics*, Oxford Handbooks (2022; online edn, Oxford Academic, 20 Apr. 2022)

Broome, John (1991). *Weighing Goods: Equality, Uncertainty and Time*. Wiley-Blackwell.

Harsanyi, John C. (1955). Cardinal Welfare, Individualistic Ethics, and Interpersonal Com- parisons of Utility. *Journal of Political Economy* 63 (4): 309–21.

McCarthy, David (2008). Utilitarianism and prioritarianism II. Economics and Philosophy 24 (1):1-33.

Nagel, Thomas (1979). Equality. In Mortal questions. New York: Cambridge University Press.

Otsuka, Michael (2015). Prioritarianism and the Measure of Utility. Journal of Political Philosophy 23 (1):1-22.

Otsuka, Michael & Voorhoeve, Alex (2009). Why it matters that some are worse off than others: An argument against the priority view. Philosophy and Public Affairs 37 (2):171-199.

Otsuka, Michael & Voorhoeve, Alex (2018). Equality versus Priority. In Serena Olsaretti (ed.), The Oxford Handbook of Distributive Justice. New York, NY: Oxford University Press. pp. 65-85.

Parfit, Derek. 1997. 'Equality and Priority', *Ratio,* 10.3, pp. 202–221

Voorhoeve, Alex & Fleurbaey, Marc (2012). Egalitarianism and the Separateness of Persons. Utilitas 24 (3):381-398.

Keywords: distributive ethics, decision theory, prioritarianism, egalitarianism, expected value theory, competing claims.

The Relevance of Internalization for Social Norm Interventions

Tatu Nuotio*[1]

[1]Helsingin yliopisto = Helsingfors universitet = University of Helsinki – Finland

Abstract

Social norms are increasingly being targeted for policy purposes. Norm-based intervertions have used a variety of methods in different behavioral domains, both for changing harmful norms and for leveraging the influence of positive ones. Despite their diversity, one thing these interventions have in common is that they focus mainly on external forms of social norms' influence. Internalization, the process that makes conformity internally motivated and less dependent on external enforcement, is rarely viewed as relevant for inducing norm-related behavior change.

Echoing recent calls in the empirical literature, I argue that internalization may be more relevant for norm-based interventions than what is currently acknowledged (e.g., Helferich et al., 2023). To properly appreciate the role of internalization in norm-based policy, however, a theoretical reappraisal of internalization itself is first needed. Currently, interventions are informed by theories that conceptualize internalization as a dichotomy between externally and internally motivated conformity, a view that has been criticized for oversimplification. In contrast, I argue that internalization should be seen as a continuum characterized by qualitative shifts in motivational processes. To demonstrate the utility of this continuum view, I consider three ways in which internalization matters for social norm policy: for having a moderating role in norm-based interventions, for being a potential pathway of behavior change in itself, and for having distinct ethical implications.

The current neglect of internalization in norm-based policy largely reflects the diminished role of internalization in the contemporary social norm theorizing generally. While internalization was once viewed as the main form of social norms' influence, the most prominent theories now focus on motivation based on the perceived behavior and expectations of others, and even view internalized norms as conceptually distinct from social norms proper (e.g. Bicchieri, 2006; Cialdini et al., 1991). Often, this is conceptualized as a distinction between social norms, defined by external motivation and social interdependence, and personal norms, defined by internal motivation and socially unconditional conformity. Although influential, this all-or-nothing view of internalization has received criticism for being implausibly strict and simplistic (Thøgersen, 2006). If personal norms are formed by internalizing social norms in the first place, it seems odd to suppose that they would make behavior insulated from further social influence, as the idea of unconditional conformity suggests. There also seems little reason to think that internalization should typically determine conformity fully, and not act alongside other, possibly more social motivations (Kelly & Davis, 2018).

*Speaker

A more fruitful alternative is to understand internalization as a continuum where ”fully internalized” norms represent only one extreme (Thøgersen, 2006). This view is espoused notably in the Social Determination Theory (SDT), a highly influential psychological theory of motivation. According to SDT, internalization happens gradually: first from external to ”introjected” regulation, then to identification, and finally to full integration with personal values (Ryan & Deci, 2017). An important tenet of SDT is that different degrees of internalization are associated with different types of internal motivations, with internal sanctions and guilt being specific to the introjected stage.

In addition to being more plausible generally as an account of internalization, the continuum view offers a richer and more nuanced picture of how internalization matters for norm-based policy:

First, since internalized norms are derived from social norms, they are likely to be present in many contexts where we intervene on the latter, which might moderate the processes induced in current norm-based interventions. The dichotomous view implies that when social norms are internalized, interventions aiming to leverage their influence would essentially be ”preaching to the converted”, so the extent of internalization should simply make them less effective. In contrast, the continuum view suggests that norms can also be internalized partially, in which case the degree of internalization rather informs us which motivational processes underlie conformity and thus which contextual factors should be paid attention to in interventions, e.g., whether the target behavior needs to be publicly observable, or whether normative appeals should be framed in terms of identity.

Second, internalization is deemed by all accounts as making conformity more robust. Inducing internalization could therefore be a way to make social norm interventions more effective, especially since many of them suffer from mainly working on observable behaviors and short-term. However, while achieving ”full internalization” and unconditional conformity may not be feasible for many intervention programs, the continuum view suggests that internalization could also be advanced more gradually. Norm-based interventions should thus seek to establish conditions that facilitate internalization, e.g., when people experience autonomy and relatedness, as some otherwise promising strategies, such as encouraging social sanctioning, could actually be detrimental to the development of internal motivation.

Finally, internalization is often conceived of in terms of internal sanctions, especially guilt. It might thus pose additional ethical problems for norm-based interventions, for example if they end up working by essentially guilt tripping people. However, the continuum view suggests that this is not necessarily a problem of internalization as a whole, but more specifically of insufficient internalization. If guilt is characteristic of a relatively early stage of internalization only, moving people ”up the internalization ladder” should help not just in- crease the desired behavior, but also decrease guilt and even increase overall wellbeing by promoting autonomy and helping align the behavior of individuals with their personal values.

References

Bicchieri, C. (2006). The Grammar of Society. Cambridge University Press.

Cialdini, R., Kallgren, C., & Reno, R. (1991). ”A Focus Theory of Normative Conduct: a Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior”. Ad- vances in Experimental Social Psychology, 24, 201–234.

Helferich, M., Thøgersen, J., & Bergquist, M. (2023). ”Direct and mediated impacts of social norms on pro-environmental behavior”. Global Environmental Change, 80, 102680.

Kelly, D., & Davis, T. (2018). ”Social Norms and Human Normative Psychology”. Social Philosophy and Policy, 35(1), 54–76.

Ryan, R., & Deci, E. (2017). Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. Guilford Press.

Thøgersen, J. (2006). "Norms for environmentally responsible behaviour: An extended taxonomy". Journal of Environmental Psychology, 26(4), 247–261.

Keywords: social norms, internalization, behavioral public policy, behavioral interventions

Identity, Social Comparison, and Institutional Categorization: A Tripartite Model of Institutional Dynamics

Noriaki Okamoto*†[1]

[1]Rikkyo University – Japan

Abstract

Although we know that individual and organizational behaviors are heavily influenced by various institutions, the nature of those institutions remains controversial, and this is partly due to the lack of a systematic examination of institutions. As such, a better understanding of the institutional framework is essential for the development of social science research.

This cross-disciplinary study thus draws on the concepts of identity, social comparison, and categorization to propose a theory of institutional dynamism. First, by reconsidering recent works to re-examine the economy and society through the lens of identity as the fundamental element that triggers human behavior (Akerlof and Kranton, 2010; Burke, and Stets, 2009; Davis, 2024; Fukuyama, 2018; Jenkins, 2014), it highlights one's identity-based self- realization as the engine driving individual actions. With limited cognitive abilities, our life represents a journey, the purpose of which is to gain an understanding of our surroundings and discover who we are and what we should do, which is, essentially, one's identity. In this sense, while previous research on identity has focused on the individual, one's identity can be seen as a hybrid of individual and collective elements. After all, whether we like it or not, we are all to some extent, involved with other people and participate in social activities-a process in which specific individuals are influenced by the identities of others and reconstruct or strengthen their own identities. In some cases, they may accept the identities of others through their interactions, and as a result, groups with similar identities are formed. In this respect, this study takes a position that differs from individual reductionism and assumes that a collective identity also exists.

Second, this study argues that our involvement in social comparison is necessary for us to update our identities. Although identity itself is a fuzzy concept, the social comparison perspective serves to complement this fuzziness. It is our innate tendency to compare our- selves to others and categorize elements of our personality, as without a continuous flow of information such as other people's benchmarks and rankings, we cannot understand who we are or decide on the optimal course of action in many cases.

*Speaker

†Corresponding author: nokamoto@rikkyo.ac.jp

In this way, one's self-identity and social comparison activities are two sides of the same coin (Akerlof and Kranton, 2010). As we progress through life, we become members of numerous communities and organizations, within which social comparison is a constant phenomenon (Matthews and Kelemen, 2024): Individuals belonging to the same community identify with each other through this phenomenon and feel a sense of homogeneity. On the other hand, in the competitive environment outside the community, we naturally feel compelled to differentiate ourselves from others (Zuckerman, 2017, p. 35), creating hierarchies through vertical social comparisons and simultaneously construct heterarchies through horizontal comparisons. In any case, the concept of social comparison requires a certain framework as both methods of comparison are influenced by institutions.

Thirdly, the present study emphasizes the significance of institutional categorization: As social comparison often relies on the establishment of benchmarks or metrics by vertical and horizontal methods, categorization is necessary (Àsta, 2018; Murphy, 2024). Institutions take a variety of forms (either formal or informal) and facilitate social comparison through categorization (Durand et al., 2017, p. 13). For example, while informal moral rules tell us what is right and wrong, formal accounting rules facilitate the comparison of companies' profits and the creation of categories based on their financial performance (i.e. whether they are in the red or black). From the perspective of the two types of social comparison outlined above (vertical and horizontal), institutions can be vertically objective in a transparent way in which various actors with significant knowledge and experience can participate in the process of deliberations to build institutions. For instance, the establishment of formal laws through parliamentary debate is often a prolonged process involving a designated procedure since this is essential for ensuring the law's objectivity and binding force. On the other hand, if a particular institution is accepted collectively by a wide range of people across various boundaries, the degree of abstraction of that institution tends to be generally high, making it horizontally comprehensive.

These three factors (identity, social comparison, and institutional categorization) interact with each other, aptly explaining the dynamism of social reality (see the Figure below). Each of us is inevitably drawn into society and as such, cannot escape social comparison- our identity is fundamentally constructed through it and requires continuous strengthening. Moreover, to better facilitate social comparison and identity formation, individuals primarily seek categorization through collective procedures, which in turn constitute institutional reality. Although there are various approaches to exploring these kinds of relationships, social scientists cannot ignore the significance of institutional categorization. Finally, this study examines a new category of growing sustainable finance and impact investing with reference to changes in the relevant actors' collective identity and social comparison.

References

Akerlof, G. A. and Kranton, R. E. (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press.

Ásta (2018). *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories*. Oxford University Press.

Burke, P. J. and Stets, J. E. (2009). *Identity Theory*. Oxford University Press.

Davis, J. B. (2024). *Identity, Capabilities, and Changing Economics*. Cambridge University Press.

Durand, R., Ganqvist, N., and Tyllstrm, A. (2017). From categories to categorization: A social perspective on market categorization. In Durand, R., Ganqvist, N., and Tyllstrm,

A., *From Categories to Categorization*, pp. 3-30.

Fukuyama, F. (2019). *Identity: Contemporary Identity Politics and the Struggle for Recognition*. Profile Books.

Jenkins, R. (2014). *Social Identity* 4th edition. Routledge.

Matthews, M. J. and Kelemen, T. K. (2024). To compare is human: A review of social comparison theory in organizational settings. *Journal of Management*, forthcoming.

Murphy, G. (2024). *Categories We Live By: How We Classify Everyone and Everything*. The MIT Press.

Zuckerman, E. W. (2017). The categorical imperative revisited. In Durand, R., Ganqvist, N., and Tyllstrm, A., *From Categories to Categorization*, pp. 31-68.

Keywords: Identity, Social comparison, Categorization, Collective norms, Sustainable finance, Im- pact investing

Cost-Benefit Analysis, Incommensurability, and Imprecise Utilities

Malvina Ongaro*†[1] and Luca Zanetti[12]

[1]Politecnico di Milano– Italy

School of Advanced Studies Studies (Scuola Universitaria Superiore) (IUSS) – Italy

Abstract

Cost-benefit analysis (CBA) is one of the most important tools to assess alternative options under scarcity of resources. In the last decades, CBA has enjoyed a rising success in public decision-making, with applications ranging over most aspects of public policy all over the world. In light of its importance, CBA has been the object of intense philosophical scrutiny, investigating its methodological and ethical assumptions (Adler & Posner 2006; Hansson 2007).

Among these, particular attention has been devoted to the assumption that all costs and benefits of an intervention are commensurable, i.e. that they can be meaningfully compared under a common metric, which is typically expressed in monetary terms (Hirsch Hadorn 2022). Without commensurability, different costs and benefits could not be traded off and aggregated in the standard CBA results (e.g. net present value, benefit-cost ratio, internal rate of return). However, some critics contend that many policy-relevant dimensions are incommensurable, perhaps because they are intrinsecally invaluable or non-substitutable. Nonetheless, in the face of incommensurable goods and services we still need to make decisions involving important trade-offs. In this paper, we aim to provide a formal setting that can help CBA include some degree of consideration of incommensurability, which could make it a valuable instrument to render these trade-offs explicit and thus more open to scrutiny.

First, we note that the notion of incommensurability can be interpreted in two ways. Under the mainstream interpretation, incommensurability denotes a quality of the external reality such that the things under consideration cannot, even in principle, be traded off. Under a different interpretation, however, incommensurability is an epistemic quality of the agent that refers to their ignorance of the precise value of the two aspects to be compared, and can therefore be reduced: if only they had enough information of the relevant kind, they would be able to meaningfully compare them.

Then, we present an approach that implements imprecise costs and benefits into CBA. Our approach adopts the machinery of imprecise Bayesianism in formal epistemology (Bradley 2017). Standard Bayesianism represents the agent's cognitive and conative states by a prob- ability function and a utility function. In the imprecise setting, credences and desirabilities are represented by sets of probability and utility functions respectively (Weirich 2021).

*Speaker

†Corresponding author: malvina.ongaro@polimi.it

We suggest that, in the context of CBA, incommensurability can be captured by making costs and benefits imprecise with the use of sets of real-valued utility functions (Weirich 2021).

Incommensurability between costs and benefits can be formally captured in this setting. To this aim, we draw on an analogy with incomplete preferences: if two quantities are in- commensurable, then any preference relation over them will be incomplete. Assuming that the agent's probability function is unique, if an agent has preferences $\alpha \beta$ and $\alpha \gamma$, but neither $\beta \gamma$ nor $\gamma \beta$, these can be represented with a set of complete utility functions such that according to some, the utility of β is greater than the utility of γ, and according to others the utility of γ is greater than the utility of β. The more opinionated the agent, the smaller the set of utility functions that are compatible with their preferences. Thus, a CBA that aimed to include considerations of incommensurability could employ imprecise costs and benefits.

We then turn to the interpretation of imprecise costs and benefits. Similarly to incommensurability, these can be interpreted either epistemically or non-epistemically. Under the epistemic interpretation, imprecise costs and benefits track our uncertainty over the actual and definite value of some aspect of the options under evaluation. Under a non-epistemic interpretation, these costs and benefits track instead an intrinsic feature of some aspects that do not have a matter-of-fact correct value that we need to discover. Consequently, in the first case imprecise costs and benefits can be used to capture incommensurability as ignorance of actual values, while in the second case they can represent it as an intrinsic feature of the world.

One can assume that all imprecise costs and benefits are epistemic; in that case, they are always the sign of incommensurability. Indeed, if they are always due to our ignorance over the actual costs and benefits, then they all represent epistemic incommensurability. Alter- natively, one can assume that no uncertainty is possible over values, or that if there is it should be captured with other (e.g. probabilistic) instruments. In that case, imprecise costs and benefits could only be interpreted non-epistemically, and thus always be the sign of non-epistemic incommensurability. However, the two interpretations are not necessary in opposition: one could also assume that both are possible, and that the same instrument can be used to track both epistemic and non-epistemic incommensurability. Which one it tracks will depend on use.

We conclude that the use of imprecise costs and benefits in CBA allows for the application in that context of some interesting decision-making tools that have been developed in the literature on imprecise Bayesianism. These may increase the degree to which CBA can take into account issues of incommensurability, thus improving its applicability to complex decisions.

References

Adler, M. D., & Posner, E. A. (2006). New foundations of cost-benefit analysis. Harvard University Press.

Bradley, R. (2017). Decision Theory With a Human Face. Cambridge: Cambridge University Press.

Hansson, S. O. (2007). Philosophical problems in cost–benefit analysis. Economics and Philosophy 23.2, 163–183.

Hirsch Hadorn, G. (2022). Which methods are useful to justify public policies? An analysis of cost–benefit analysis, multi-criteria decision analysis, and non-aggregate indicator systems. Journal for General Philosophy of Science, 53 (2), 123-141.

Weirich, P. (2021). Rational Choice Using Imprecise Probabilities and Utilities. Cambridge: Cambridge University Press.

Keywords: Incommensurability, Imprecision, Cost, benefit analysis, Policymaking, Bayesianism

Big M 2.0. The art of war for philosophers of economics

Aleksander Ostapiuk*
[2]Wroclaw University of Economics and Business – Poland

Abstract

Problem

Over the years, there were two special issues of JEM (Journal of Economic Methodology) where philosophers of economics critically reflect on their field (Blaug et al. 2001; Davis and Hands 2021). The two questions capture the philosophers of economics' imagination. The first question is descriptive. Philosophers of economics ask about the state of philosophy of economics (PoE) and analyze how it has changed over years. The second question is normative and concerns the future of PoE. The two normative issues are debated. Firstly, should PoE be more about philosophy or economics? Secondly, should philosophers of economics focus on small m or Big M methodology? (McCloskey 1998).

The descriptive discussion concerning PoE is well-analyzed. Over time PoE had moved from economics first to philosophy first. It means that now PoE has a more philosophical perspective. Secondly, philosophers of economics move away from Big M and focus more on small m. It means they now focus on description (small-m) not prescription (Big M). Moreover, the focus is on daily practice of economics (small m), not big debates on economics as science (Big M). Philosophers of economics agree that PoE has changed over the years (Davis and Hands 2021; Mireles-Flores and Nagatsu 2022; Truc et al. 2021). The existing debates concern a normative question on PoE's future.

Although philosophers of economics accept that philosophical turn happened, there is a shared feeling among them that PoE should influence economics more (Davis and Hands 2021). My paper's contribution starts here. I analyze how PoE can make a greater impact on economics discourse which is underlined by the paper's subtitle: "The art of war for philosophers of economics".

My contribution

Philosophers of economics propose two conflicting roads to influence economists. Most argue that the focus should be on small m (e.g., Boland 2001; Boumans 2021; Colander 2013; Guala 2021; Hands 2001; Kincaid and Ross 2021; Lehtinen 2021). The second, smaller group argues that we must return to some new version of Big M (Alexandrova et al. 2021; Dow 2021; Vromen 2021).

I agree with the second camp and argue that we should get back to some version of Big M if we want to

*Speaker

have more influence on economics discourse. My argument is as follows:

1. Small m does not attract normal economists and does lead to PoE's greater impact on economics.

2. Instead of normal economists who do not care about methodology, philosophers of eco- nomics should focus on other audiences like economists (big shots, heterodox economists, students) and the educated public who are interested in Big M.

3. I propose "Big M 2.0" as a strategy to influence economics discourse. It is an adjusted version of Big M used in the 80s.

My paper provides three arguments to the discussion on the status of PoE (e.g., Davis and Hands 2021, Mireles-Flores and Nagatsu 2022). Firstly, I show the limits of small m as a strategy to make an impact on economists. Secondly, I turn the philosophers of economics' attention from normal economists to other underappreciated groups interested in economics discourse (big shots, heterodoxy, students, and educated public). Thirdly, I present my ac- count of "Big M 2.0" as a strategy to have more impact on economics discourse.

Structure

In 1 section, I analyze the changes in PoE from the 80s. I focus on two movements: from Big M to small m and from economics to philosophy. I analyze the disenchantment with Big M and the process of PoE's specialization.

In 2 section, I analyze small m as the strategy proposed by some philosophers of economics to influence economists. I show why this strategy does not attract normal economists and impedes communication with heterodoxy and educated public ("mainstream bias" in Mireles- Flores and Nagatsu 2022).

In 3 section, I show that other audiences: 1. economists (big shots, heterodoxy, students) 2. educated public are more interested in Big M than small m. I presented the crisis of 2008 as a PoE's wasted opportunity in capturing the demand for the philosophy of economics. I distinguish three areas of special interest: 1. What is economics right now?, 2. How values influence economics' methodology? 3. Which is better: Neoclassical or heterodox economics? I argue that by focusing on these questions, philosophers of economics can capture the attention of various audiences interested in PoE.

In 4 section, I present my account of "Big M 2.0". I show how it differs from the un- successful Big M used in the 80s. Firstly, in Big M. 2.0 philosophers of economics are not philosophers kings that preach from the high horse of philosophy of science. Secondly, Big
M. 2.0 focuses on audiences other than normal economists (big shots, heterodoxy, students, educated public). I present how philosophers of economics can contribute to the discussions on each Big M question 1. What is economics right now? 2. How values influence economics' methodology? 3. Which is better: Neoclassical or heterodox economics? I argue that by contributing to these debates, PoE can have more impact on economics discourse.

Conclusions

In this paper, I argue that philosophers of economics should return to big M questions and focus on different audiences (big shots, heterodoxy, students, educated public). I pro- pose Big M 2.0 as a framework that omits some problems with old Big M and shows how to communicate with various audiences efficiently. Consequently, PoE could influence eco- nomics discourse more.

”Big M 2.0” is open to discussion and can be questioned as a feasible strategy to have more impact on economics discourse. Nonetheless, the paper’s overarching goal is to wake philosophers of economics from their dogmatic slumber where it is accepted that PoE cannot influence economics discourse.
If we want philosophers of economics to influence economic discourse, how do we see our role? As plumbers who analyze the discussions within mainstream economics (small m) or generalists who guide discussion on economics (Big M 2.0)? In the paper I argue for the latter.

Keywords: philosophy of economics, Big M, small m, philosophy of science

Of the Vices Behind Biases: Reassessing the Contribution of Nudges as Welfare Policy Tools

Daniel Zarama Rojas*[1]
[2]Bureau d' Économie Théorique et Appliquée (BETA) – Université de Lorraine – France

Abstract

The present text attempts to answer by proposing a new way of conceptualizing cognitive bi- ases, which have often been presumed as the justification for nudge interventions. I introduce a novel conceptual framework into the philosophy of economics, virtue epistemology (VE, hereafter) to the study of biases. VE is a recently born branch of academic epistemology that studies the traits that characterize excellent and deficient knowers. For this reason, VE allows to study how the differences between the capacity to judge (epistemic capacity) of different agents can be related to their individual attributes. It started by identifying the faculties that sustain adequate judgment (Sosa 1980): someone might adequately judge the sky to be blue by virtue of her paired vision. Recent studies have included the study of other types of individual traits and have considered also traits who block or hinder judgment (Za- gzebski 1996, Battaly 2014). For instance, it could be possible that the same individual with paired vision might offer unreliable financial advice, because she lacks the required skills to process financial problems. Virtue epistemologists offer an outlook into the individual traits that sustain and harm epistemic capacity, what they call epistemic virtues and vices.(2) In this paper, I mobilize their conceptualization of epistemic virtue and vice to look upon and distinguish among the different types of vices that may be associated with cognitive biases.

Accordingly, biases could be comprehended as the effects of underlying epistemic vices. Thus, I propose to discuss a reformulated policy problem in terms of VE. What may be the contribution of nudges to the suppression of epistemic vices? To answer this question, I first discuss Thaler and Sunstein's (2008, (2021) 2022) proposal of nudges in section II.

Therein, I show how nudges bundled a collection of diverse cognitive errors (biases) together with Kahneman's (2003) dual-process theory of the mind, and the how the latter cannot adequately account for the emerge of biases. Section III brings forth a VE framework to look into the problem. First, I propose a simple conceptual model with which to categorize epistemic virtues and vices. Then, I discuss two forms in which it is possible to specify the relation between traits and misjudgment. Reliabilist epistemic vices, on the one side, regroup the traits that 'consistently' lead to false judgments and those that 'consistently' block the acquisition of true judgments. Obstructivist epistemic vices, on the other side, comprehend the traits that might lead to such epistemic bads "overall" (Cassam 2019, p. 5). Section IV distinguishes between biases and epistemic vices, which might be behind their emergence. Finally, section V discusses the contribution that nudges may offer to the welfare concerns that would be associated with the suppression of epistemic vices. I advance that nudges would, at best, a partial solution to the problem of cognitive biases. Nudges are incapable of altering any vicious individual traits, generalizing the case of character traits (Meehan 2020) to all kinds of epistemic attributes. Then, their contribution would depend upon the kind of vices at play: while nudges could be useful to affront 'mild' obstructivist epistemic vices, they are structurally disproportionate to the problematic represented by more 'consistent' reliabilist epistemic vices. Section VI concludes and suggest lines for future research.

References

Battaly, Heather. 2008. ”Virtue Epistemology.” Philosophy Compass, 3(4): 639-663.

Battaly, Heather. 2014. ”Varieties of Epistemic Vice.” In Jonathan Matheson and Rico Vitz (eds.). The Ethics of Belief: Individual and Social. Oxford: Oxford University Press.

Beshears, John, and Harry Kosowsky. 2020. ”Nudging: Progress to date and future di- rections.” Organizational Behavior and Human Decision Processes, 161(Supplement): 3-19.

Cassam, Quassim. 2019. Vices of the Mind: From the Intellectual to the Political. Ox- ford: Oxford University Press.

De Ridder, Denise, Floor Kroese, and Laurens van Gestel. 2022. ”Nudgeability: Map- ping conditions of susceptibility to nudge influence.” Perspectives on Psychological Science, 17(2): 346–359.

Grüne-Yanoff, Till. 2016. ”Why Behavioural Policy Needs Mechanistic Evidence.” Eco- nomics and Philosophy, 32: 463-483.

Kahneman, Daniel. 2003. ”Maps of Bounded Rationality: Psychology for Behavioral Eco- nomics.” American Economic Review, 95(3): 1449-1475.

Kidd, Ian James, Heather Battaly, and Quassim Cassam (eds.) 2021. Vice Epistemol- ogy. London and New York: Routledge.

Meehan, Daniella. 2020. ”Epistemic Vice and Epistemic Nudging: A Solution?” In Guy Axtell and Amiel Bernal (eds.), Epistemic Paternalism: Conceptions, Justifications and Im- plications. Lanham: Collective Studies in Knowledge and Society: 249-261.

Mertens, Stephanie, Mario Herberz, Ulf J. J. Hahnel, and Tobias Brosch. 2022. ”The effec- tiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains.” Proceedings of the National Academy of Sciences of the United States of America, 119(1): e2107346118. https://doi.org/10.1073/pnas.2107346118

Sosa, Ernest. 1980. ”The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge.” Midwest Studies in Philosophy, 5(1): 3-26.

Thaler, Richard H., and Cass R. Sunstein. 2008. Nudge: Improving Decisions About Health, Wealth, and Happiness. New Haven and London: Yale University Press.

Thaler, Richard H., and Cass R. Sunstein. (2021) 2022. Nudge: The Final Edition. Penguin Random House.

Turri, John, Mark Alfano, and John Greco. 2021. ”Virtue Epistemology,” in Edward N. Zalta (ed.). The Stanford Encyclopedia of Philosophy.

Zagzebski, Linda T. 1996. Virtues of the Mind: An inquiry into the nature of virtue and the ethical foundations of knowledge. Cambridge: Cambridge University Press.

(1) Their review studies the changes in the effectiveness of nudges under particular condi- tions. They argue that people’s response to nudges does not change when the transparency of policy mechanisms and objectives increases. Besides, individuals become insensitive to nudges when policy objectives differ from their own. They conclude ”(m)ore systematic re- search on which conditions make people more (or less) responsive to nudges is warranted” (De Ridder, Kroese, and van Gestel 2022, p. 354).

(2) For a more detailed introduction to Virtue Epistemology, to the main questions that guide associated research, and to the main differences between VE approaches, see Battaly 2008. A more comprehensive version may be found in Turri, Alfano and Greco (2021). For a brief history of vice epistemology and an outlook on its recent developments, see Kidd, Battaly and Cassam (2021).

Keywords: nudges, biases, virtue epistemology, epistemic vices, public policy

Modeling in Economic Policy: Rethinking Models' Epistemic Limitations

Nadia Ruiz*†[1]

[1]Leibniz Universität Hannover=Leibniz University Hannover – Germany

Abstract

In the last decade, philosophers of science have examined the relationship between scientific models and policy. This interest ranges from deepening our understanding of the role of scientific models in making predictions and providing causal explanations-used for the development of policies such as climate mitigation and economic interventionist policies-to a more recent interest in exploring the trade-off between models' epistemological and normative implications for policy (Douglas, 2000, 2009; Lloyd, 2009, 2015; Parker, 2011; Parker & Lusk, 2019; Contessa, 2020; van Basshuysen et al., 2021; Winsberg & Harvard, 2022a, 2022b; Douven, 2023).
Beyond the attempt to establish an appropriate balance between epistemic and normative values-under the assumption that better policies will emerge from such a balance-there re- mains a largely overlooked aspect of modeling and model-based policymaking: the assessment of models' epistemic limitations in relation to policy. One reason for this oversight is the excessively embellished evaluation of models through their characterization as "adequate for a purpose" and the way that purpose is defined (whether epistemic or practical). An overemphasis on adequacy for a purpose-as the justification for using a model for a specific purpose-has prevented philosophers of science from addressing important questions such as: Are all purposes able to find an adequate model? Do the needs and perspectives of policymakers influence a model's adequacy for a purpose? How do these factors shape the development of new models? Finally, and most importantly, how do modelers, in their role as policy advisors, address "new" purposes that cannot (yet) be analyzed using available models?

To assess these questions, I will focus on the *Central Planbureu* (CPB Netherlands Bureau for Economic Policy Analysis). The CPB, founded in 1945 and instituted by the Law of April 21, 1947, is a government-independent institution that functions as a central source of economic information within the government and as a center for applied economic analysis. Among the tasks of the CPB are to develop macroeconometric models for predicting major variables such as percentage change in GDP–i.e., forecast modeling. And study the impact of certain policies such as budget deficit, inflation, and unemploymenyt in the Ductch economy–i.e., scenario analysis modeling. Notably, the CPB not only serves the government but also receives policy analysis and analysis of election program requests from opposition parties, trade unions, and employers' organizations. These external engagements reinforce the CPB's independence (Bos and Teuling, 2011), and it could be argued that, for this rea- son, its models maintain a certain degree of government-value neutrality.

*Speaker
†Corresponding author: nadia.ruiz@cells.uni-hannover.de

By focusing on the CPB's role as a center for applied economic analysis-whose main function is to provide general socio-economic insights on how certain economic policies could impact the Dutch economy-this study separates itself from debates regarding non-epistemic values in scientific practices and allows for a concentration solely on the epistemic status of models. More specifically, I am interested in the CPB's practices as a knowledge-producing entity. Although the CPB's reputation had to be earned and repeatedly defended, it is undeniable that its policy analyses are generally influential. One reason is that CPB's public trust in its independence. The second reason is that macroeconomists working at the CPB treat models as major informational instruments that help them and policymakers structure information in a coherent way, in other words, information that is clear and connected (van Els, 2003), somewhat contrary to academic economists' skepticism about economic modeling in general, and macroeconomic modeling in particular.

So, what have macroeconomists at the CPB been doing to maintain a cohesive narrative of the Netherlands' economy? I propose that this involves two possible tactics:

1. Using models' epistemic limitations (i.e., when a purpose lacks an adequate model) to guide the development of new models.
2. Including and explicitly acknowledging a model's epistemic limitations in policy analysis.

In this paper, thus, I argue that in cases in which a policy question (i.e., purpose) does not have an adequate model, both economists and policymakers turn to the epistemic limitations of available models, which subsequently influence the development of future models. To demonstrate this, I will focus on CPB models and employment policy between 1950 and 1997. This period allows for an assessment of how CPB's modeling focus shifted from a Keynesian macroeconomic approach (demand-side) to the VINTAF model (which incorporated supply-side effects) and eventually to the development of MIMIC (Micro Macro Model to Analyze the Institutional Context), a fully-fledged general equilibrium model that includes microfoundations.

References

Bos, F., & Teulings, C. N. (2010). CPB and Dutch fiscal policy in view of the financial crisis and ageing. *Available at SSRN 1742181*.

Contessa, G. (2021). Inductive risk in macroeconomics: Natural Rate Theory, monetary policy, and the Great Canadian Slump. *Economics & Philosophy*, *37*(3), 353-375.

Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Pre. Douven, I. (2023).

Pandemics and flexible lockdowns: In praise of agent-based modeling.
European Journal for Philosophy of Science, *13*(3), 35.

Harvard, S., & Winsberg, E. (2022). The epistemic risk in representation. *Kennedy Institute of Ethics Journal*, *32*(1), 1-31.

Lloyd, E. A. 2009. "I-Elisabeth A. Lloyd: Varieties of support and confirmation of cli- mate models." Aristotelian Society Supplementary Volume 83 (1):213–32.

Lloyd, E. A. 2015. "Model Robustness as a Confirmatory Virtue: The Case of Climate Science." Studies in History and Philosophy of Science Part A 49:58–68.

Parker, W. S. 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." Philosophy of Science 78 (4):579–600.

Parker, W. S., & Lusk, G. (2019). Incorporating user values into climate services. *Bulletin of the American Meteorological Society*, *100*(9), 1643-1650.

Van Basshuysen, P. (2023). Austinian model evaluation. *Philosophy of Science*, *90*(5), 1459-1468.

van Basshuysen, Philippe, Lucie White, Donal Khosrowi, and Mathias Frisch. 2021. "Three Ways in Which Pandemic Models May Perform a Pandemic." Erasmus Journal forlosophy and Economics 14 (1):110–27.

van Els, P. J. (2003). Policy-making and model development: the case of the Nederlandsche Bank's model MORKMON. In *Empirical Models and Policy Making* (pp. 94-99). Routledge. Winsberg, E., & Harvard, S. (2022). Purposes and duties in scientific modelling. *J Epidemiol Community Health*, *76*(5), 512-517.

Keywords: Emperical models, policy making, epistemic limitations

Scratching the Surface: The Use of Participatory Methods in the Lab to Uncover Wicked Problems

Massimo Rusconi*[1] and Lorenzo Gagliardi*[2]

[1]InsIDE Lab, Department of Economics, University of Insubria – Italy
[2]Department of Developmental Psychology and Socialization, University of Padua – Italy

Abstract

Scientific experimentation in the social sciences has long relied on controlled laboratory settings to investigate human decision-making, cooperation, and preference formation (Roth, 1995, Guala, 2005). While these experiments have yielded critical insights into how people behave under varied incentive structures, they typically limit the scope of interpersonal interaction to pre-set, highly regulated channels. This constraint makes it difficult to investigate ”wicked” topics, such as abnormal psychologies, political polarization or conspiracism, in an open-ended manner (Bonetto et al., 2018, Charness et al., 2022). Although field experiments and focus groups have been employed to tackle complex issues (Tadajewski, 2016), these approaches pose unique challenges when it comes to capturing both the nuanced interactions due to ”wicked” topics and the need for rigorous causal identification. In recent years, participatory methods have gained traction in sustainability and develop mental research, where they have proven effective for enacting practical solutions and enhancing the understanding of local and regional issues (Moallemi et al., 2021). These methods, which encompass a broad range of strategies including participatory design (Namioka and Schuler, 2009) and participatory enaction, invite participants into a more active, co-creative role. By leveraging open-ended discussion, negotiation, and iterative reflection, participatory methods can reveal interpersonal dynamics and encourage self-reflection processes often obscured by traditional lab experiments. In what can be seen as a dynamic ”safe space,” participants are free to revisit their own assumptions, reconcile contradictory views, and articulate their reasoning in response to peers’ perspectives. Such iterative feedback loops have the potential to reduce the gap between stated and revealed preferences, a key issue in experimental economics, and produce unique insights into how individuals revise their attitudes in real time. Despite their widespread implementation in sustainability and community-engagement contexts, participatory methods have not been systematically integrated into lab experimentation, leaving unexploited the possibility of a hybrid methodology that combines the control of laboratory protocols with the openness and reflexivity characteristic of participatory sessions. To address this gap, the present paper explores how participatory methods can be deployed alongside, or in lieu of, standard lab-based experimental protocols, inspired by sociological perspectives on ”serious games” (Barreteau et al., 2021) and the notion of ”safe space” engagement (Pereira et al., 2015).

*Speaker

We argue that these approaches can be operationalized in a manner that preserves methodological consistency and replicability, provided a few key innovations are introduced. One key element is individuated in the need of splitting roles among the designer, the enactor, and the analyzer. The designer shapes the overall research structure, thus identifying core questions, determining how participatory elements will be integrated, and ensuring alignment with the study's objectives. The enactor, in turn, facilitates the sessions: guiding discussions, ensuring inclusive participation, and maintaining the session's flow without biasing outcomes. Finally, the analyzer collates and interprets the resultant data, which may include transcripts, video recordings, and participants' self-reported reflections. By segregating these roles, biases introduced by facilitation or participant self-selection can be mitigated, helping sustain the reliability of the study's findings.

Further, to ensure the potential for causal inference (Angrist and Pischke, 2009), we propose multiple strategies for embedding participatory methods in experimental or quasi- experimental designs. A straightforward option is a randomized controlled trial (RCT) at the group level: participants are randomly assigned to either a participatory session or a traditional lab session, enabling direct comparisons of outcomes across the two conditions. Another approach is difference-indifferences (DiD), particularly suited if data can be collected across multiple time points, thereby tracing changes in preferences or behaviors that unfold over the course of participatory interventions. In settings where randomization is not feasible, instrumental variables may help address self-selection bias, so long as a suitable instrument, i.e. one that influences the likelihood of participation but does not directly affect outcomes, can be identified. A persistent methodological challenge lies in analyzing the qualitative data produced by open-ended discussions, group reflections, and non-verbal cues. Lab experiments traditionally yield discrete, numerical outcome measures (e.g., contribution levels in a public goods game), whereas participatory sessions generate multifaceted, context-rich information. Managing such data calls for mixed-methods approaches (Braun and Clarke, 2006) that blend qualitative techniques (thematic analysis, discourse analysis) with quantitative proxies of interaction (participation rates, sentiment coding, or decision outcomes). Rigorous protocols for data recording, that is backed by transparent coding procedures and, where feasible, blind evaluation, are essential to preserving replicability and minimizing interpretive bias.
Notwithstanding these challenges, integrating participatory methods with lab experimentation promises significant advantages. First, it bolsters ecological validity by reflecting the fluid, conversational nature of everyday decision-making. Second, it narrows the distance between stated and revealed preferences, affording participants room to clarify or adjust their perspectives in light of peer feedback. Lastly, it allows researchers to observe emergent phenomena, ranging from leadership dynamics to consensus-building, which would typically remain hidden in more rigid laboratory protocols.

The paper is organized into three main sections. Section 1 introduces participatory methods, briefly comparing them with other techniques such as focus groups, and surveys the contexts in which they are most commonly employed. Section 2 outlines the minimal methodological requirements for using participatory methods in an economic laboratory, discussing the role division among designer, enactor, and analyzer, as well as the protocols needed to maintain experimental rigor. Section 3 delves into the challenges of implementing participatory approaches in the lab, emphasizing design complexity, data richness, and alignment with causal inference frame-works. Finally, we offer perspectives on how this hybrid methodology can deepen our understanding of human decision-making and address some of the limitations found in conventional lab experiments.

Keywords: Economic Methodology, Economics and Psychology, Experimental Economics, Serious Games, Participatory Methods

On the Utility of Models: Bridging the Theoretical and Analytical Sides of Model Selection

Massimo Rusconi*[1] and Raffaello Seri[1]

[1]InsIDE Lab, Department of Economics, University of Insubria – Italy

Abstract

Scientific inquiry often hinges on selecting models that adequately capture the phenomena under investigation, by employing different strategies and toolkits based on modeling purposes and contexts of analysis. Within this landscape, the theory of model selection has been broadly constructed around two primary focuses. On the one hand, philosophers of science emphasize the conceptual integrity of models (e.g., Grüne-Yanoff and Marchionni, 2018, Tieleman, 2022), that is how well a model's structure aligns with theoretical frameworks or explanatory aims. On the other hand, econometricians focus on output adequacy measured through criteria such as predictive performance, hypothesis testing, and information criteria (e.g., Hotelling, 1958, Box, 1979, Myung, 2000, Spanos, 2021). Striking a balance between these perspectives remains a central challenge, with both practical and theoretical implications. In fact, while most models are usually scrutinized separately in their theoretical consistency and empirical performance, the prerequisite for enacting the latter vis a vis the former, that is the unambigous distinction among different families of models on the same phenomenon, remains unfulfilled. This paper aims at elucidating such necessary prerequisites and proposes a unified framework for model selection that leverages Utility Theory as the central criterion, integrating both conceptual and statistical evaluations into a single decision-making process. In doing so, the framework relies and enhances the analogical understanding within the field of model theory (Bartha, 2010, Knuuttila and Loettgers, 2017). In this framework, each candidate model m from a set M is evaluated by a total utility function U (m) composed of two distinct components: a conceptual utility UC (m) and a statistical utility US (m j D), where D represents the observed data. The overall model selection process then amounts to choosing the model that maximizes the sum of these utilities:

m⋆ = arg max
m2M
fUC(m) + US(m j D)g :

At the conceptual level, model selection is set as guided by analogical reasoning that compares the origin structure of a model with a target phenomenon. This evaluation is performed relatively to a benchmark B, which may be a model template or an ideal theoretical reference. Distinct utility functions may capture this analogical reasoning.

*Speaker

First, the Global Similarity Utility UG(m) assesses the overall resemblance between the structure S(m) and that of the benchmark B, considering all the features of the model. In a hypothetical quantitative formulation, one might assign ranks and weights to each feature of S (m), yielding an expression such as
UG(m) = X
i2F
wi ρi(m; B);
where F is the full set of features, ρi represents a qualitative rank, and wi is the weight assigned to feature i. However, in practice, this evaluation remains largely qualitative, allowing for considerable degrees of freedom. The second function, the Relevance-Based Utility UR(m), is confined to a subset C ⊂ F of core features deemed essential for capturing the phenomenon. Here, only the ”core” elements are evaluated:
UR(m) = X
i2C
vi σi(m; B);
with σi representing a qualitative rank for the core feature i and vi its corresponding weight.
It is important to note that UR(m) and UG(m) are not aggregated into a single metric because they serve distinct evaluative purposes: global similarity considers all aspects of a model's structure, while relevance-based similarity is restricted to those components critical to the theoretical framework. Moreover, models are intrinsically ranked according to the provisioned utility function, which may include further properties such as simplicity and parsimony (Falk and Muthukrishna, 2021, Bargagli Stoffi et al., 2022).

These conceptual evaluations also allow the identification of families of models. Models that differ only in minimal, distinct ways, that is where the differences are both distinct and countable, are grouped into the same model family. This grouping clarifies the conceptual space and permits systematic comparison among closely related models by emphasizing their direct differences. At the statistical level, model selection is based on the model outputs derived from observed data D. The statistical utility US (m j D) reflects empirical performance, often approximated through measures such as likelihood, predictive accuracy, and information criteria (e.g., AIC or BIC). Techniques such as hypothesis testing are also employed; however, it is critical to recognize that hypothesis testing serves as a condition for pairwise selection rather than providing a complete ranking across multiple models. Provided that, US(m j D) is expressed as the utility based on the data, US(m j D).

By combining these two facets-the qualitative, analogical conceptual evaluation and the quantitative, data-driven statistical evaluation-we arrive at an integrated utility function that guides the selection of an optimal model. This approach unifies the conceptual and empirical dimensions of model selection under a clearly defined two-stage evaluative framework, ensuring that the trade-offs between theoretical coherence and empirical adequacy are made explicit.

The paper is structured as follows. Sect. 1 reviews the existing philosophical and econometric
frameworks of model selection and it makes a case for the use of the concept of utility as a comprehensive theoretical framework. Sect. 2 provides a detailed description of the frame- work as well as the necessary conditions for its deployment. Sect. 3 delves into critical issues, such as stability and univocity in defining models and model families, and the potential extension of the model to accommodate the comparisons across fundamentally different families. Moreover, it deals with the limitations in providing a quantitative expression for conceptual evaluations, and the consequences of the unchecked use of qualitative evaluations with multiple degrees of freedom.

Keywords: Model Selection, Economic Methodology, Epistemology, Information Criteria, Analogical Inference

Welfarist Meritocracy

Enrico Mattia Salonia*[1]

[1]Toulouse School of Economics – université Toulouse 1 Capitole, École des Hautes Études en Sciences Sociales, Centre National de la Recherche Scientifique, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement – France

Abstract

Meritocracy has recently been at the centre of debate (Markovits, 2019; Sandel, 2020; Tirole, 2022). However, as Sen (2000, p. 5) suggested, "*the idea of meritocracy may have many virtues, but clarity is not one of them.*" This paper constitutes an attempt to bring some clarity. A society is meritocratic if individuals' behaviour is rewarded according to merit considerations. Unpacking this informal definition requires dissecting the institutional components that make society a meritocracy and study each of them separately. I develop a theory in which institutions are characterised by what I refer to as a **basis of merit** and a **nature of reward**. Meritocracy is then defined as an institution satisfying specific properties in its two components.

There is a society of individuals choosing one of several actions. Each of them has a preference defined over collective action profiles. I interpret preferences as attitudes toward choice, in line with revealed preference theory. I only employ ordinal information, avoiding inter- personal comparisons or cardinal measures of satisfaction.

The first component of the institution is the **basis of merit**. To determine whether an individual deserves more or less, a metric must be introduced. I assume that an individual's action is more deserving than another when it leads to greater social welfare. As is common in economics, I assume that welfare coincides with preference satisfaction. One action is more deserving than another if it leads to a Pareto improvement in social welfare. Consider one individual and fix the profile of others' actions. One of her actions is more deserving than another, keeping others' behaviour fixed, if it leads to a profile of actions which is weakly preferred by all individuals and strictly preferred by some, compared to the profile induced by the latter action. This basis for merit does not make welfare comparisons between individuals, which would require more information about preferences.

The second component of the institution is the **nature of reward**. More deserving individuals must be rewarded with more of something. I assume that welfare is the reward for merit. If an action is more deserving than another, one must be better off choosing the first than the second. As before, better off in this framework means preference satisfaction. Choosing an action that is more deserving than another must lead to a profile preferred by the individual compared to the one induced by the latter action.

A **basis of merit** and a **nature of reward** induce a game among individuals. I study the equilibria of the game, assuming that the basis of merit is increments of social welfare

*Speaker

and the nature of reward is increments of individual welfare. I assume there is complete information on individuals' preferences and no feasibility constraints on rewards. The model represents the most ideal setting for introducing institutions with desirable properties.

I show that, by properly setting the reward system, any action profile can be implemented. I interpret this result as a hint that meritocracy is a vacuous concept under the welfarist economic approach. The nature of both the basis of merit and reward completely overlaps with individuals' motivation for behaviour, as the metric of all three is the satisfaction of preferences. The notion of merit becomes meaningless, as in an ideal world with no lack of information or feasibility constraints individuals can be led to behave in any way.

In the language of the paper, I obtain the following result.

Proposition 1. *If the basis of merit satisfies Pareto, if the reward system is welfarist, then for any action profile, there exists ameritocratic rewardsystemunder which it is pareto efficient and implemented in a dominant strategy equilibrium.*

Proposition 1 shows that if the relevant currency for choice, merit and reward is welfare, then one action profile can be selected arbitrarily to be the most deserving and all individuals will choose their action from that profile. Choosing the most rewarding action increases both in- dividual and collective rewards. Since reward is welfare, this leads to a Pareto improvement and a higher score on the basis of merit. Because the system is meritocratic, higher merit induces higher welfare. Individuals choose to increase their welfare and will therefore opt for the most rewarded action. This loop is because of the overlap of the nature of reasons for choice, basis of merit and rewards. Opting for a meritocratic system is equivalent to selecting which actions the individual will choose, which defeats the purpose of rewarding people for choosing more contra less worthy actions. The intuition behind Proposition 1 was noticed by Arneson (2017), where it is argued that such circularity is a problem for desert-based theories of distributive justice.

Keywords: Meritocracy, welfarism, Desert Based Justice

Reality bites: Bringing “policy” back to philosophy of evidence-informed policy

Sofia Blanco Sequeiros*[1]

[1]University of Helsinki – Finland

Abstract

During the past two decades, evidence-informed and evidence-based policy have gained increasing attention in philosophy of social science and philosophy of economics. Philosophers of social science have focused particularly on the ”evidence” in evidence-informed policy, discussing the intricacies of producing (policy-relevant) scientific evidence at length (e.g. Douglas 2009, Cartwright and Hardie 2012, Melo-Mart ın and Intemann 2018). Per- haps understandably, the ”policy” in evidence-informed policy has received less attention from philosophers of science. Particularly in philosophy of economics, little attention has been paid to the details of the actual policy processes in which scientific evidence is ideally used, or the way that those policy processes relate to scientific evidence and its production as policy processes.

This paper discusses the use of scientific evidence, especially the use of uncertain evidence, in support of policy. I discuss these specifically from the policymaking perspective. I show that philosophers of social science, including philosophers of economics, have overlooked the heterogeneity of both policymakers and policy processes in evidence-informed policy. Instead of a single group of ”policymakers” who uniformly decide or decline to use evidence in their work, there is a diverse group of government officials employed at different levels of the government hierarchy, with varying opportunities and resources for using evidence in their work. Likewise, instead of a single ”policy process”, there is a multitude policy processes within government, ranging from research by government-employed scientists to drafting laws, and from policy interventions implemented by government bureaus to explicitly political decisions by politicians trying to appease their voters.

I argue that overlooking this complexity has led to philosophers of economics neglecting key issues in the use of scientific evidence and research for policymaking, which in turn has led to misguided normative arguments about using scientific evidence in policy. Most importantly, these issues include the ”unideal” situations in and constraints under which policy processes and various policymakers often operate, such as lack of time and money and political pressure and feedback. For instance, Thoma (2022) argues for value pluralism in the science used to support democratic policymaking. She highlights three challenges for value pluralism: the accessibility of pluralist frameworks, strategic exploitation of pluralism, and partisan science, and provides solutions for these challenges. While she is correct in arguing that value pluralism supports epistemically equal democratic decision making, I argue that Thoma does not fully appreciate the magnitude of these and other challenges that truly adopting value pluralism in government would face. As a result, the normative force of her arguments is diminished. Adopting value pluralism in government without truly addressing the challenges would likely turn against itself, and result in policymaking that would further contribute to epistemic inequality.

*Speaker

To paint a better picture of the realities of policymaking, I discuss three challenges in evidence-informed policy: lack of resources, political pressure, and the different norms and standards under which scientists and various policy officials operate. The first is a daily reality of policymaking, the second a challenge to policymakers updating their beliefs with novel evidence, and the third a challenge to researchers being able to effectively communicate their findings to policymakers. Using an example from economics of crime on the effective- ness of various punishments on crime, I show that all of these challenges are exacerbated if the scientific evidence at hand is uncertain. As all scientific evidence is uncertain to a degree, these challenges are present in the everyday of evidence-informed policy.

The solutions to these challenges, I argue, lie in better grasping uncertainty, discordance and disagreement in scientific evidence (Reijula and Blanco Sequeiros 2024; Maziarz 2023; Roussos 2020), as well as disagreement and political representation in policymaking (de Melo- Mart´ın and Intemann 2018, Hilligardt 2023). Uncertainty and lack of consensus in science are easily exploitable in policymaking, which a better understanding and communication of uncertainty, emphasizing why uncertain evidence can nevertheless be reliable evidence, can ameliorate. This leads to questions about the political role of scientists: I argue, along the lines of Hilligardt (2023), that researchers should be seen as a type of political representative, and policymakers as stakeholders in the research process. Positioning researchers as political representatives and policymakers as stakeholders of scientific research bridges the gap between researchers and policy officials. Finally, I argue that the interplay of dissent and consensus in democratic discussion and deliberation must be better appreciated in philosophy of social science, including philosophy of economics. Democratic governance is premised on the idea that a shared notion of political good can be found even between disagreeing parties, while the deliberative process remains open to and tolerant of dissent; only through dissent can a shared political goal be reached.

In summary, I show that in evidence-informed policy, reality bites: unless philosophers of economics, and philosophers of social science more broadly, more carefully consider the intricacies of various policymaking processes and the incentives and interests of various policymakers and policy institutions, the normative force of philosophical conclusions about using scientific evidence in policymaking is diminished. To press this point, I outline (a non-exhaustive set of) three challenges central to evidence-informed policymaking, and pro- pose solutions to them. Throughout the paper, I use an example from economics of crime to illustrate my arguments, as the challenges become particularly clear when dealing with uncertain evidence.

Cartwright, N., & Hardie, J. (2012). Evidence-based policy: A practical guide to doing it better. Oxford University Press.

de Melo-Mart´ın, I., & Intemann, K. (2018). The fight against doubt: How to bridge the gap between scientists and the public. Oxford University Press.

Douglas, H. E. (2009). Science, policy, and the value-free ideal. University of Pittsburgh Pre. Hilligardt, H. (2023). Partisan science and the democratic legitimacy ideal. Synthese, 202(5), 135.

Maziarz, M. (2024). ”Conflicting results and statistical malleability: Embracing pluralism of empirical results.” Perspectives on Science, 32(6), 701-728.

Reijula, S., & Blanco Sequeiros, S. (2024). ”Persistent evidential discordance”. British Jour- nal for the Philosophy of Science.

Roussos, J. (2020). Policymaking under scientific uncertainty (Doctoral dissertation, London School of Economics and Political Science).

Keywords: evidence, informed policy, policymaking, political philosophy of science

Diversity and strategic beliefs

Milosz Slepowronski*[1]

[1]Uniwersytet Jagielloński w Krakowie = Jagiellonian University

Abstract

Gerald Gaus' New Diversity Theory (2016, 2021) extends classical liberalism to address modern pluralism and social complexity. It seeks to answer how individuals with divergent moral, religious, and philosophical commitments can co-exist peacefully under shared norms. Rejecting a unitary vision of the social contract, Gaus examines how diverse societies can reconcile on a common moral framework without compromising individual moral integrity. Diversity, in this view, is not a problem, but a resource: differing perspectives yield novel insights and solutions to moral and social challenges. Suppressing this diversity through rigid laws or enforced consensus impedes social learning and undermines a flourishing social order.

The project contains an implicit vision of human psychology, which stresses how norms should be justifiable to the people following them, not merely functional in preventing chaos. If norms are justifiable, that is, individuals see a good reason to endorse them from their own perspective, they are more likely to internalize and maintain them.

In his theory, Gaus incorporates ideas from decision theory (like rational choice and optimization models), but does so in a qualified way, adapting their assumptions to fit the reality of moral decision making. He explicitly contrasts his approach with what he calls the 'simple optimization model' associated with neoclassical economics and rational choice theory. Under that classic model, an ideal decision maker is faced with a well-defined set of options and simply assigns a utility or "score" to each, then chooses the option with the highest score (maximizing their objective). Gaus points out that this framework, while useful in some domains, rests on strong assumptions that rarely hold in complex moral-social situations. He acknowledges the elegance of the rational choice model: Essentially, it treats decision-making as picking the best option according to one's goals. He even notes that we can represent a variety of moral doctrines in a formal decision framework (for example, even the principles of a deontologist can be mapped to a utility-like ranking of outcomes).

However, Gaus significantly modifies the classical model to make it more realistic for hu- man morality. Drawing on Herbert Simon, Gaus emphasizes that real people have limited cognitive resources and seldom find the best solution, and thus the theoretical focus is the collective process of search and discovery. The result is a decision-making framework more akin to an evolutionary search or complex adaptive system than a straightforward utility maximization. In addition, individuals have a method of revising their values, which means that what a person cares about can evolve with new arguments or experiences. Standard eco- nomic models almost never allow an agent's utility function to change endogenously; Gaus explicitly builds in the possibility of moral learning and change. However, even given all these modifications, this might not be enough given recent updates on our understanding of the sources of moral and political beliefs.

*Speaker

The Gausian approach to the problem of justifiability of norms might be in danger in light of the Alliance Theory of Political Belief Systems by Pinsof et al. (2023). The Alliance Theory challenges the view that political beliefs arise from deep-rooted ideological principles. Instead, it argues that such beliefs are shaped primarily by political alliance structures, which vary across societies and historical contexts. Individuals adopt political positions aligning with their group's interests rather than intrinsic ideological commitments. Since alliance structures differ between societies and eras, the specific content of political belief systems is fluid and depends on the dynamics of the prevailing alliance. People's moral judgments and political stances are influenced by strategic considerations related to their alliances, suggesting that moral reasoning is often a tool for alliance maintenance rather than a reflection of objective principles. At first glance, alliance theory undermines Gaus' project because public reason assumes agents are persuadable by justifications, the system relies on people being able to see reasons that transcend their tribal commitments. If beliefs are mostly post hoc rationalizations for alliances, public reason looks performative: If people adopt moral stances because of group loyalty, not because they are reasoning impartially, then public reason becomes just another arena for signaling in-group allegiance.

In the paper, I will analyze how much of a danger the Alliance Theory really is for the New Diversity Theory. It is not a coup-de-grace because Gaus understood that individuals' moral beliefs are often shaped by self-interest, social alliances, and in-group loyalty. He didn't see humans as pure reasoners, but as boundedly rational agents, shaped by history, institutions, and social pressures. However, in order to accommodate alliance-based belief systems, New Diversity requires a shift in emphasis. Instead of relying on individuals being fully persuaded by moral reasons, we might instead see public reason as a constraint on alliance dynamics. Alliances will always exist, but norms of public reason might act as a soft check on how far alliances can go in undermining cooperation; e.g., even if people are mostly tribal, norms of public reason could set boundaries on how alliances operate (like forbidding outright scapegoating of minorities).).

If all political reasoning is just post hoc rationalization of alliances, Gaus' system would be in trouble. Public reason would collapse into a mask for power struggles. But even Pinsof et al. do not claim that everything is alliance driven; they argue that it is dominant, not total. As long as there is some space where reasons and arguments can matter, we have a room for maneuver.

Secondly, Alliance Theory can itself be a resource for further research. Gaus saw moral orders as evolving systems, where norms emerge and stabilize not because people agreed on them rationally, but because they worked in maintaining social order. Alliance dynamics could be part of how these norms evolve and research into that can be fruitfully integrated into Gausian framework.

Keywords: Diversity, decision making, beliefs, Gaus

Rethinking the Rules of the Game: Synthetic Data for Institutional Economics

Petr Špecián *[1]

[1]Charles University, Institute of Political Science – Czech Republic

Abstract

Introduction

In the marketplace of ideas, the free exchange of thoughts allows the best concepts to emerge through open debate and memetic competition. Yet can this ideal be separated from the technological and institutional infrastructure supporting the discourse? Today's epistemic cost reducers-technologies lowering barriers to information dissemination and processing- have already transformed how ideas are generated and evaluated (Sunstein 2017). With generative AI (GenAI) emerging, we stand before another phase shift.

This paper hypothesizes that the pending AI-induced epistemic shock could translate into decisive improvements in institutional design in both political and economic contexts. Particularly, GenAI-enabled synthetic data offers a promising methodological frontier, potentially breaking the longstanding "experimentation bottleneck" in institutional research (Specian 2024). Seizing this promise, however, requires methodological flexibility and creativity.

The Strenuous Crawl of Institutional Research

Institutional economics faces daunting challenges in providing reliable evidence on novel institutional mechanisms. Traditional research methods struggle against constraints that impede epistemic progress: institutional mechanisms involve non-linear interactions and emergent properties complicating theoretical modeling (Beinhocker 2006); identical mechanisms may induce different behaviors at varying scales, making small-scale experiments potentially mis- leading (Ostrom 2005); institutional performance can be highly contingent on historical and cultural contexts (Acemoglu and Robinson 2013); and agents adapt strategically to new institutions, creating evolving dynamics difficult to capture in typical experiments (Basu 2018).

Theoretical models rely on simplifying assumptions about behavior and information, limiting their real-world relevance. Laboratory experiments offer high internal validity but struggle with external validity. Field trials provide ecological validity but remain costly, time-consuming, and ethically challenging, especially when testing untried mechanisms (cf. List 2020).

These methodological constraints create a persistent gap between institutional innovation conceptually and real-world institutional reform. Any substantive reform carries risk, and today's social science offers only a limited evidence base to pursue it.

*Speaker

Synthetic Data: A New Methodological Frontier

The limitations noted above are not due to fundamental impossibilities; these constraints can be eased. GenAI-enabled synthetic data offers a promising path forward. Researchers have demonstrated GenAI models' ability to create 'generative agents' capable of simulating complex social interactions (Park et al. 2023, 2024). Early explorations show that these agents can mimic complex behaviors like cooperation (Kaiya et al. 2023), suggesting syn- thetic data could effectively simulate human behavior relevant to institutional analysis.

The potential of these techniques is profound: synthetic data can be gathered at minimal cost compared to current methods, enabling more extensive exploration of institutional design space; researchers can easily manipulate agent characteristics, institutional parameters, and environmental conditions; using digital agents avoids ethical concerns associated with human subject experiments; and simulations have no obvious scale limits, potentially enabling large agent populations that approximate real-world sizes and capture emergent properties (Chan et al. 2024).

In essence, synthetic data could help realize a dream of many institutional economists: rapid iteration and testing of institutional mechanisms in conditions realistic enough to deliver reliable predictions of real-world performance. This approach could unblock avenues to progress in institutional design sorely needed against the background of fast technological change and mounting global issues.

Epistemic Humility Required

We currently stand at an early stage of the necessary concerted effort to ascertain the promise-and perils-of synthetic data in institutional research. Synthetic data cannot be taken at face value without understanding its specific limitations and risks (Grossman et al. 2023). Methodological adaptations, technical improvements, and epistemic humility will be necessary for fruitful and responsible use.

This methodological shift raises fundamental epistemological questions. Reliance on AI- generated simulations forces a reevaluation of how we interpret and validate findings. Ensuring model validity is crucial; simulations must be rigorously tested against empirical data to confirm they accurately capture essential aspects of human behavior and institutional dynamics (Specian 2024). Researchers must also remain vigilant about potential biases encoded in AI algorithms, which could inadvertently perpetuate or exacerbate social inequities (Buolamwini and Gebru 2018).

Moreover, these methodological innovations will need to be coupled with technical improvements and likely a renegotiation of research community norms. This may include embracing prediction as a key criterion for evaluating model success, developing hybrid approaches combining synthetic and human-subject data, and implementing adversarial testing to stress-test proposed institutional mechanisms. The academic community may need to recalibrate its focus towards epistemic progress, potentially updating norms around authorship, peer review, and quality control to better accommodate the novel challenges of GenAI-assisted research.

Conclusion

The integration of synthetic data into institutional economics may transform the field, helping overcome longstanding challenges in studying and designing institutions. By bridging the gap between theoretical models and empirical reality, it could inform more effective institutional designs capable of addressing key coordination problems of modern societies.

While this promise may ultimately prove a mirage, given the stakes we face and the paucity of alternative avenues to swift epistemic progress in institutional research, few methodological challenges deserve more attention from reform-minded institutional economists.

References

Acemoglu, Daron, and James Robinson. 2013. *Why Nations Fail*. New York: Crown Business.

Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. 2023. "Out of One, Many: Using Language Models to Simulate Human Samples." *Political Analysis* 31 (3): 337–51. https://doi.org/10.1017/pan.2023.2.

Basu, Kaushik. 2018. *The Republic of Beliefs*. Princeton ; Oxford: Princeton University Press.

Beinhocker, Eric D. 2006. *Origin of Wealth*. Boston, Mass: Harvard Business Review Press.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Dis- parities in Commercial Gender Classification." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html.

Chan, Xin, Xiaoyang Wang, Dian Yu, Haitao Mi, and Dong Yu. 2024. "Scaling Synthetic Data Creation with 1,000,000,000 Personas." arXiv. https://doi.org/10.48550/arXiv.2406.20094.

Grossmann, Igor, Matthew Feinberg, Dawn C. Parker, Nicholas A. Christakis, Philip E. Tetlock, and William A. Cunningham. 2023. "AI and the Transformation of Social Science Research." *Science* 380 (6650): 1108–9. https://doi.org/10.1126/science.adi1778.

Kaiya, Zhao, et al. 2023. "Lyfe Agents: Generative Agents for Low-Cost Real-Time Social Interactions." arXiv. https://doi.org/10.48550/arXiv.2310.02172.

List, John A. 2020. "Non Est Disputandum de Generalizability? A Glimpse into The External Validity Trial." Working Paper. *National Bureau of Economic Research*. https://doi.org/10.3386/w27535.

Ostrom, Elinor. 2005. *Understanding Institutional Diversity*. Princeton: Princeton University Press.

Park, Joon Sung, et al. 2023. "Generative Agents: Interactive Simulacra of Human Behavior." In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 1–22. San Francisco CA USA: ACM. https://doi.org/10.1145/3586183.3606763.

Park, Joon Sung, et al. 2024. "Generative Agent Simulations of 1,000 People." arXiv. https://doi.org/10.48550/arXiv.2411.10109.

Specian, Petr. 2024. "Digital Homunculi: Reimagining Democracy Research with Generative Agents." arXiv. https://doi.org/10.48550/arXiv.2409.00826.

Sunstein, Cass R. 2017. *#Republic: Divided Democracy in the Age of Social Media*. Prince- ton ; Oxford: Princeton University Press

Keywords: Synthetic Data, Institutional Economics, Economic Methodology, Generative AI, Marketplace of Ideas

Should economics as a science be impartial?

Hanna Trojanowska*[1,2]

[1]Hanna Trojanowska, Warsaw University of Technology – Poland
[2]Warsaw University of Technology – Poland

Abstract

One of the tasks of science is to seek truth. One of the tests we can perform to see whether our scientific research has really led us to objective truth is to perform tests using different methodologies. If the results of research conducted using different methods are the same, or at least similar within the limits of statistical error, then we can assume that our pursuit of truth has indeed reached its happy end.

It can be assumed that such an approach in the exact sciences is fully justified. After all, it is impossible to deny the facts that are the results of well-conducted scientific research. However, what are we to do if we are dealing with social sciences, where proving a thesis may depend to a significant extent on the school to which the scientist belongs, and therefore also on the way he or she thinks and proves the truth.
Such dilemmas are not alien to economics, which, as a science about economic man, is to some extent determined by his or her humanism and individualism. The interpretation of economic phenomena can lead us to various conclusions, which do not necessarily have to be wrong. Then the question arises: are ”erroneous” theories leading us to social profit and prosperity ”better” or ”worse” than ”true” theories that will lead us to losses?
The question then arises whether social effects or the impartiality of economic theories are more important in economics. And if impartiality, what does it really mean? What does impartiality and objectivity mean in science, which assumes that different schools may have diametrically different views on the same phenomena? A good example of such a dilemma is the predatory pricing strategy and the discussion on this topic that the two schools had with each other. On the basis of game theory, it was argued that the predatory pricing strategy is a very effective method of defending access to the market and eliminating competitors from it. Meanwhile, economists from the so-called Chicago School argued that this strategy does not really exist, and therefore cannot be ap- plied in practice. On the basis of game theory, behavior consistent with this strategy could lead to the ethical elimination of a competitor from the market. Meanwhile, the Chicago School argued that the predatory pricing strategy is merely the result of theoretical speculations that have no basis in real economic life. The dispute was not only an academic pursuit of truth and had serious consequences. A law was formulated on its basis. The result of the philosophical question of whether economics can be impartial has become an element of the strategic to be or not to be for many companies.

In this example, we can see that academic disputes are not just an exchange of opinions and lead to specific actions in practice.

*Speaker

So what should be the criterion of truth in economists' searches? Is convincing the opponent on theoretical grounds enough? Should economics be impartial or should it primarily take into account the criteria of social well-being? In the case of, for example, a company seriously polluting the natural environment, should the application of the predatory pricing strategy be assessed according to the Chicago School and in the case of the collapse of the only free hospital in the city according to the school using game theory? Or should we value game theory more for its mathematical approach and ignore the arguments of the Chicago School, which uses non-thematic argumentation? Should theory evaluation be impartial or ethically engaged?

Keywords: truth, making decisions

Decision Models and Tainting Beliefs

Oyku Ulusoy*[1]

[1]. – Turkey

Abstract

Decision theory typically assumes that individuals form preferences based on their subjective beliefs and wants (Savage, 1954). In this context, value assignments correspond to the comparative betterness of alternative acts or outcomes as perceived by the decision-maker, while beliefs are subjective probability judgments about relevant possible occurrences or states of the world (Thoma, 2019). More recently, Williams (2020) have argued that beliefs influence more than just the structuring of state likelihoods; they can shape how individuals describe options and can themselves become objects of preference. For example, as conceptualized by neophobia (Collins, 2015), the perceived risks of extremely negative outcomes can determine the salient features of major life changes for someone, leading to a strong aversion to them. Individuals may even assign values to beliefs (Williams, 2023)- for instance, prefer- ring to believe that humans are inherently good rather than inherently bad. In this sense, subjective probabilities do not merely reflect a judgment of accuracy. Rather, they reveal that beliefs do not diverge and contradict solely due to misinformation, as suggested earlier in the literature (Kahan, 2017; Mercier, 2020). This highlights the challenge of coherently and fully integrating one's beliefs when contemplating one's decisions. In this paper, I further examine how a preference for beliefs complicates modelling decisions for decision theory, as allowing for alternative framings means that alternative perspectives to entertain a choice problem may not always be compatible, each involving distinct belief sets that shape the problem's salient features. I first analyse this issue through the lens of transformative choice (Paul, 2014) and as a framing problem (Bermúdez, 2020).

There are some propositions that can occupy human thought for a lifetime-sensitive to evidence yet resistant to definitive resolution. Are people ultimately good or bad? Is there life after death? Will I be a good parent? Will there be another pandemic like COVID-19 in our lifetimes? Did my supervisor act in my best interest, or was he deliberately misguiding me for petty reasons? As we deliberate on these questions, the strength and objects of my beliefs may shift, without arriving at definitive evidence to align my subjective judgments and preferential attitudes.

I posit that this presents a unique type of choice problem-one that requires committing to a particular belief-set among alternatives, making the choice of a viewpoint a complex and consequential decision, similar to transformative choices (Paul, 2014). Paul employs decision theory to argue that fundamentally new discoveries bring about significant preference changes by allowing individuals to entertain new knowledge. I argue that entertaining a proposition alone, such as "what if I will make a bad parent," can have similar effects on one's viewpoint by committing them to a specific set of beliefs. While any of the alternative views on this question might resonate as I deliberate on parenthood, it may not be possible to harmoniously integrate my beliefs over all the propositions I entertain on the topic. That is, since there are likely incompatible ways of viewing this

*Speaker

question, which cannot be resolved through research and evidence collection, there are multiple viable ways for me to form my beliefs that will then influence how I perceive a choice about parenthood. Hence, my beliefs regarding the first question will shape the nature of the parenthood choice: if I opt to believe that I would be a good parent, that belief determines the salient features of the decisions I make regarding parenthood.

How should a rational choice framework, designed to represent rational individuals, be ap- plied to deliberation and belief formation in such cases? I contend that this phenomenon presents a challenge for decision theory, as it introduces alternative ways of framing one's choice setting. Within decision theory, models are subject to multiple interpretations and applications, each with distinct variations, as model-based studies inherently omit certain characteristics of the target phenomena (Titelbaum, 2013). A central challenge for modellers, thus, has been determining the appropriateness of models for their intended domain (Mäki, 2018). Ideally, a decision model should represent how an individual perceives and un- derstands their decision setting to ensure its suitability. However, my point here is that the possibility of different belief-sets shaping how decision-makers engage with a choice problem reveals the possibility of incommensurable alternative framings, presenting another choice problem, akin to those discussed by Kuhn (1983) and Okasha (2011) on scientists' theory choice. As adopting a modelling framework resembles how phenomena can be viewed and handled, changing this framework can be likened to an epistemic change by altering how a scientist–and a decision-maker in our case–perceives their subject.

This shortcoming of modelling decisions-and how the salient features of decisions can take different shapes under different model outlooks-has been discussed as the framing effect by Bermudez (2020; 2022), who argues that in complex decision situations, the way one frames their decision setting reflects their reasons for a specific outcome or action. Thus, the framing of a decision problem influences behaviour by altering how one perceives the relevant aspects of an outcome (Weisel & Zultan, 2022). The chief problem for decision theory, then, is that once a specific framing is adopted, it shapes the decision-maker's perception of the available options, as determined by their beliefs as well as their wants. However, a model may reveal different orderings of preferences, each yielding distinct optimal course of actions with markedly different consequences for different belief-sets.

Works Cited

Bermúdez, J. L. (2020). *Frame it again: New tools for rational decision-making*. Cam- bridge University Press.

Bermúdez, J. L. (2022). Rational framing effects: A multidisciplinary case. *Behavioral and Brain Sciences*, *45*, e220. doi:10.1017/S0140525X2200005X

Collins, J. (2015). Neophobia. *Res Philosophica*, *92*(2), 283-300. https://doi.org/10.11612/resphil.2015.92.2.6

Kahan, D. 2017. Misconceptions, misinformation, and the logic of identity-protective cognition. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2973067.Google Scholar

Kuhn, T. (1996). *The structure of scientific revolutions* (3rd ed., Issr collection). Chicago, IL: University of Chicago Press.

Mäki, U. (2018). Rights and wrongs of economic modelling: refining Rodrik, *Journal of Economic Methodology*, 25:3, 218-236, DOI: 10.1080/1350178X.2018.1488475

Mercier, H. 2020. Not Born Yesterday. Princeton, NJ: Princeton University Press.Google Scholar

Paul, L. A. (2014). Transformative experience. Oxford University Press.

Savage, L. (1954). The foundations of statistics. Dover Publications.

Thoma, J. (2019). Decision Theory. In Pettigrew, R. & Weisberg, J. (eds.), The Open Handbook of Formal Epistemology. PhilPapers Foundation. Pp. 57–106.

Titelbaum, M. (2013). *Quitting certainties: A bayesian framework modeling degrees of belief* (1st ed.) (1st ed.). Oxford: Oxford University Press.

Weisel, O., & Zultan, R. (2022). Frames, trade-offs, and perspectives. *Behavioral and Brain Sciences*, *45*, e245. doi:10.1017/S0140525X22001005

Williams, D. (2020). Socially adaptive belief. *Mind & Language*. https://doi.org/10.1111/mila.12294. Williams, D. (2023). The marketplace of rationalizations. *Economics and Philosophy*. 39(1):99-123. doi:10.1017/S0266267121000389

Keywords: Beliefs, Rational choice, Decision models, Framing effects, Theory choice

Social Preferences and Social Welfare in Behavioral Welfare Economics

Jack Vromen*[1]

[1]Erasmus Institute for Philosophy and Economics – Netherlands

Abstract

Debates in Behavior Welfare Economics (BWE, understood here broadly as attempts to solve McQuillin and Sugden's (2012) reconciliation problem, not narrowly as just Bernheim's solution) have mostly focused on the issue of the extent to which (revealed) preferences can still be treated as reliable measures of individual well-beings. Research in behavioral eco- nomics (BE) that prompted discussion of this issue clearly is research done in the "heuristics and biases program": the detection of biases and heuristics that can explain observed choice anomalies. Apparently, research in BE that is engaged with identifying social preferences and developing social preference models is supposed not to pose serious challenges for standard welfare analysis. *Prima facie* this is surprising, since welfare analysis is ultimately always about social welfare, inevitably involving distributional issues, and social preferences in social preference models are either about desired distributions or about desired behaviors of others (having distributional consequences). Welfare economics arguably consists of two parts: first, how should individual well-beings be evaluated, and second, how should individual well-beings be socially aggregated. Ongoing debates in BWE proceed as if social preferences of stakeholders do not challenge standard practices in either part (with only a few exceptions such as Manski 2025).

I will discuss what individual well-being and social welfare in BWE, and especially the issue of what welfare criteria are appropriate, could mean from the perspective of social preferences. I first note that in the context of welfare analysis "social preferences" is an ambiguous concept (as has been observed by several commentators; e.g., Fleurbaey 2019). It can refer both the outcome of a welfare analysis (following from applying particular welfare criteria) – about what social state ought to be preferred – and to concerns stakeholders have about what others have and do. How the two might differ and how the two might be related are rarely discussed. As an outcome of a welfare analysis, social preferences are meant to be ethical preferences (often cashed out in terms of what enhance social welfare most from an impartial benevolent point of view). Despite several remarks to the contrary, suggesting that the social preferences that figure in social preference models always stem from such moral concerns (such as social efficiency and fairness), I argue that such social preferences are best seen as a behavioral concept, rationalizing observed behavior that allegedly cannot be explained in terms of a "self-interest only" utility-maximizing model (in a sense, I argue, social preference models in BE work from the presupposition that there is an "inner self-interested agent" in each of us). As such, moral motivation might but need not underlie the latter sort of social preferences. To keep the two more clearly apart, I propose to label the first sort of social preferences *ethical social preferences* and the latter sort *behavioral social preferences*. Concerning their possible relationship, I discuss how behavioral social preferences might figure in several places in welfare analysis. They might be included or excluded in the input

*Speaker

of welfare analysis, for example, in evaluating individual well-being. Alternatively, to the extent that the latter sort of social preferences stem from moral concerns, they might also inform the choice of the functional form of the Social Welfare Function (SWF), determining the distributional weights assigned to individual well-beings. I argue that social preferences should not be ”doubly counted” at any rate: they should not be included both in individual well-being and in the SWF’s functional form.

It is also possible, however, that behavioral social preferences of stakeholders, and indeed all of their preferences, are disregarded altogether in arriving at ethical social preferences. In- dividual well-beings of stakeholders can be ”paternalistically” understood and distributional weights can be assigned to them by an ”ethical deliberator” in ways that are independent of how stakeholders themselves look at these. Welfare economists might hold that the resolution of these critical issues should be the outcome of taking a principled substantive ethical stand (and this might underlie their ”paternalism”), not of a popularity poll among the populace. I argue that bypassing all preferences of stakeholders is an open option only for welfare economists who think that a (what I will call) a *welfarist* justifier grounds all of welfare analysis. A welfarist justifier involves the acceptance not just of the belief that social welfare is all that matters in the evaluation of social states, but also of the belief that social welfare is a function solely of the individual well-beings of stakeholders. I shall also argue that the welfarist justifier is not the only possible justifier for welfare analysis, however. A *democratic* justifier (emphasizing individual sovereignty in self-governing one’s own life and one’s own society) is another justifier that is actually sometimes endorsed. I shall argue that from the perspective of this democratic justifier, bypassing preferences of stakeholders is not an open option. To the extent that preferences of stakeholders are taken to accurately reflect their judgments of their own well-beings and of the social welfare of their societies, they ought to be deferred to in welfare analysis, regardless of whether these preferences are believed to be reliable indicators of individual well-being and social welfare.

Keywords: Social preferences, social welfare, welfare criteria, behavioral welfare economics

Senian conceptions of agency and commitment for individual choice and welfare assessment in the environmental domain

Julia Wdowin*[1]

[1]University of Cambridge – United Kingdom

Abstract

Capitalism, economic growth and the welfare state all hang in the balance as an environ- mental crisis unfolds. Whilst de-growth circles criticise economic growth for its environmental costs, some circles point to 'green growth' as a directed economic growth with diminished environmental costs, whilst providing a context for continued flourishing in other areas of human welfare. The directed growth aims to align private and social value more closely. In modelling green growth, individual preferences are key, yet, a strong economic modelling error in conventional approaches that advocates of green growth point towards is the strong assumption of exogenous preferences. Other assumptions in conventional economic individual choice analysis are single preference rankings and making observations based on revealed preferences. A focus on these assumptions neglects the possibility that individuals adjust or change their behaviour on the basis of changes to policy or incentives; have a narrow set of values motivating behaviour and choice; and that observed individual choices reflect individuals' true preferences, respectively. This paper explores an application of relevant concepts from Sen's capability approach to the question of modelling individual choice in the environmental domain, whereby these critiques of exogenous and revealed preferences, as well as single preference rankings, are addressed.

The Senian capability approach models a two-fold criterion to broader individual welfare assessment. One dimension refers to individuals' opportunities to achieve welfare, and the other to individual agency in the process (Sen, 1980a; 2009). We will focus on the latter and its particular definition in Sen's approach to welfare assessment, with implications for the process of formation of individual preferences and their role in individual choice. When preferences are believed to be influenced by changes in policy or incentives, a much more realistic assumption in economic modelling of individual choice and behaviour is to account for endogenous preferences. These, in turn, are contingent on acknowledging a broader set of values inherent to individual choice. Nested in the argument is the point that the types of preferences modelled have implications for the outcomes of economic welfare assessments.

A second issue is that of modelling preferences according to a single preference ranking in standard economic approaches. Yet, the environmental domain, in particular, is associated with a broad range of values that guide individual behaviour and choice, for which unidimensional preference rankings are a poor fit. Hence, the paper will look to the mechanism of commitment proposed by Sen (1977) in forming multiple rankings of preferences as tools for better modelling of individual choice in the context of the environmental domain.

*Speaker

In exploring relevant notions of the capability approach, namely agency and commitment, as a basis to the formation of (endogenous) preferences, the paper thus takes a broader- than-usual approach to modelling individual choice and behaviour with regards to the environmental domain. The Senian application contributes to conceptually and analytically moving beyond a unidimensionality of description (Sen, 1980b) and of economic modelling for understanding human welfare linked to the environmental domain by addressing issues linked to exogenous and revealed preferences, and single preference rankings as potentially misleading methods for analysing individual choice. In this light, the paper also makes a contribution to a conceptually more inclusive approach to welfare assessment.

Keywords: Sen, agency, commitment, natural environment, welfare, assessment, preferences

Regret aversion

Hayden Wilkinson*[1] and Konstantin Genin

[1] Germany University of Oxford – United Kingdoom

Abstract

Suppose I offer you a choice between two options Probably Worse and Probably Better, each of which would result in some number of those strangers being saved from painful deaths. And I tell you this: with probability ¾, choosing Probably Worse would turn out to save one life fewer than would Probably Better; and, with probability ¼, choosing Probably Worse would turn out to save three more lives than would Probably Better. From this description, you might think to yourself "Well, Probably Better will probably turn out better. Choose Probably Worse and I will most likely come to regret it." You might thus intuit that Probably Better is better. You might still intuit this if I give you a fuller description of the two options. A fair coin will be flipped twice, with states s_1 to s_4 corresponding to the four possible sequences of flips, each having probability 1/4. For each such state, the number of lives that would be saved under each option are given below. (All other evaluatively relevant features are held fixed.)

	S_1	S_2	S_3	S_4
Probably Worse	1	2	3	4
Probably Better	2	3	4	1

Even with this fuller description, you might intuit that Probably Better is a better option than Probably Worse. And, at first glance, it doesn't seem irrational to have at least one of these strict preferences. Indeed, a sizeable minority of real-world agents-between 31% and 42% of experimental subjects, depending on the study—reveal preferences of this sort (Loomes and Sugden, 1987; Loomes, 1988b,a; Starmer and Sugden, 1989; Loomes et al., 1992).

But these verdicts are almost universally rejected by normative decision theory, as it currently stands. Take, for instance, the orthodox view of which risky options are better than others: expected utility theory. An option's expected utility is entirely determined by the option's possible outcomes and its probabilities of resulting in those outcomes. And Probably Better and Probably Worse have exactly the same probabilities of the same outcomes. They will have the same expected utility, and so the theory must say that they are equally good. The theory cannot accommodate the verdict that Probably Better is strictly better.

Nor can various alternative theories accommodate any of the above verdicts. Various alternatives to expected utility theory have been proposed, often to allow for sensitivity to risk-to give additional weight to possible outcomes in which an option turns out particularly badly or particularly well, beyond the weight given by expected utility theory. These include risk-weighted expected utility theory (from Buchak, 2013; cf. Quiggin, 1982), chance-sensitive utility theory (Stefánsson and Bradley, 2019), and weighted-linear utility theory (Bottomley and Williamson, 2024; see also Chew, 1983). But, in the above case, these theories agree with expected utility theory. Again, Probably Better and Probably Worse must be equally good.

Does this mean that the above preference is irrational? I will argue that, plausibly, it does not. Instead, we have an objection to all normative decision theories so far proposed. Every such theory fails to accommodate a

pattern of preferences, exemplified by those above, that I will call regret aversion: a preference against an option being more likely to actually turn out worse (or perhaps actually turning out a great deal worse), even holding fixed its outcomes and their probabilities.

*Speaker

In regret aversion, we have a plausible new consideration for normative decision theory. And, as I argue in this paper, it is a well-motivated one there are compelling reasons to think it rational to be regret-averse. One such reason is that it is necessary to reflect a particular concern for what is objectively best. Another such reason is that, in cases of extremely low probabilities and extremely high stakes, regret aversion allows us to correctly diagnose and avoid the so-called 'fanatical' verdicts that many decision theories espouse. And, helpfully, regret aversion can be captured in a comprehensive theory of decision-making—we can adapt an existing theory of descriptive decision-making from economics, known as regret theory (originating from Loomes and Sugden, 1982), to accommodate regret aversion in the normative setting. And, although both regret theory and the more general pattern of regret-averse preferences are subject to some objections, I will suggest that these objections are not fatal. It remains plausible, I will argue, that it is rational to be regret-averse.

References

BOTTOMLEY, C. AND WILLIAMSON, T. L., 2024. Rational risk-aversion: Good things come to those who weight. Philosophy and Phenomenological Research, 108, 3 (2024), 697-725. (cited on page 2)

BUCHAK, L., 2013. Risk and Rationality. Oxford University Press. (cited on page 2)

CHEW, S. H., 1983. A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox. Econometrica, 51 (1983), 1065-1092. (cited on page 2)

LOOMES, G., 1988a. Further evidence of the impact of regret and disappointment in choice under uncertainty. Economica, (1988), 47-62. (cited on page 1)

LOOMES, G., 1988b. When actions speak louder than prospects. The American Economic Review, (1988), 463-470. (cited on page 1)

LOOMES, G.; STARMER, C.; AND SUGDEN, R., 1992. Are preferences monotonic? testing some predictions of regret theory. Economica, (1992), 17-33. (cited on page 1)

LOOMES, G. AND SUGDEN, R., 1982. Regret theory: An alternative theory of rational choice under uncertainty. The Economic Journal, 92, 368 (1982), 805-824. (cited on page 2)

LOOMES, G. AND SUGDEN, R., 1987. Testing for regret and disappointment in choice under uncer-tainty. The Economic Journal, 97, Supplement (1987), 118-129. (cited on page 1)

QUIGGIN, J., 1982. A theory of anticipated utility. Journal of economic behavior & organization, 3, 4 (1982), 323-343. (cited on page 2)

STARMER, C. AND SUGDEN, R., 1989. Probability and juxtaposition effects: An experimental investigation of the common ratio effect. Journal of Risk and Uncertainty, 2 (1989), 159-178. (cited on page 1)

STEFÁNSSON, H. O. AND BRADLEY, R., 2019. What is risk aversion? The British Journal for the Philosophy of Science, (2019). (cited on page 2)

Manipulation in Collective Decision Mechanism: Solutions from Economics and Computer Science

Bele Wollesesn*[1]
[1]London School of Economics and Political Science – United Kingdom

Abstract

In designing institutions for collective decision-making, a primary concern is ensuring that the mechanisms employed are not vulnerable to manipulation by strategic agents. Economists commonly invoke rational-agent models, employing insights from mechanism design and social choice theory to devise incentive-compatible mechanisms for collective decision-making. The gold standard of incentive compatibility in this literature is strategy- proofness. In a strategy-proof mechanism, no participant can benefit by misrepresenting their preferences, meaning that truthful revelation is a weakly dominant strategy. Meanwhile, computer scientists are building on these economic models and developing complementary approaches (Zuckerman & Procaccia, 2009; Von Krogh et al., 2021), emphasizing algorithmic methods that make strategic manipulation computationally intractable or opaque (e.g., Bartholdi et al., 1989; Elkind & Lipmaa, 2005). Despite these shared goals, the two fields operate with different descriptive assumptions: discussions of which approach to use to reduce the threat of manipulation often center on the descriptive accuracy of the idealized assumptions made about potential strategic agents.

This paper argues that the normative significance of these approaches does not depend solely on the accuracy of their underlying behavioral assumptions but also on the broader normative concerns motivating the design of a manipulation-resistant system. Assessing the realism of descriptive assumptions alone cannot resolve this issue if the underlying normative concerns differ. ”Manipulation” itself is a thick concept: it describes a misrepresentation while also carrying normative implications about harm. The harm of manipulation in mechanism design is typically associated with welfare loss (e.g., if we are not satisfying the ‘true’ preferences, what is the normative appeal of other desiderata such as Pareto efficiency?). By contrast, work in fair machine learning examines how to prevent strategic behavior that undermines fairness or predictive accuracy. In this context, ”manipulation” frequently refers to altering input attributes to achieve a preferred result-e.g., an individual modifying their stated demographic information to secure a more favorable loan or admission outcome. As these approaches integrate and are used in the same applications (e.g., predicting student success and using the prediction to match students with universities), it becomes less clear which normative goal is achieved by implementing different safeguards against manipulation. Often, models are simply transferred without realizing that solution concepts also incorporate such normative assumptions.

A further complication arises from the failure of algorithmic approaches to distinguish be- tween inducing substantive changes in underlying attributes (e.g., acquiring additional skills, reflecting genuine adaptation) and incentivizing strategic misrepresentation of attributes (e.g., falsifying qualifications, undermining decision integrity). While the former may be normatively desirable as an adaptive response to incentives, the latter is generally problematic due to its distortionary effects on decision-making criteria. Nonetheless, both behaviors can negatively impact predictive accuracy, necessitating a nuanced approach to their regulation.

*Speaker

This paper examines two principal strategies for counteracting strategic manipulation: (1) constraining the possibilities for strategic interactions, as in strategy-proof mechanisms, and (2) reducing transparency to limit the feasibility of manipulation. While reduced transparency may effectively mitigate the epistemic harm of manipulation, it does not necessarily align with fairness considerations. Conversely, mechanisms designed to preclude strategic interactions may better serve fairness objectives but do not directly address concerns related to epistemic accuracy. In practice, decision-makers must navigate competing priorities, such as optimizing welfare, adhering to fairness constraints, and safeguarding epistemic integrity. A critical examination of how different approaches to manipulability embed distinct assumptions about harm is essential for clarifying the trade-offs involved. By disentangling the specific harms that different methodologies seek to mitigate, this paper contributes to a more rigorous and normatively informed approach to designing collective decision-making mechanisms.

References

Bartholdi, J. J., Tovey, C. A., & Trick, M. A. (1989). The computational difficulty of manipulating an election. Social choice and welfare, 6, 227-241.

Elkind, E., & Lipmaa, H. (2005, December). Hybrid voting protocols and hardness of manipulation. In International Symposium on Algorithms and Computation (pp. 206-215). Berlin, Heidelberg: Springer Berlin Heidelberg.

Procaccia, A. D., & Rosenschein, J. S. (2007). Junta distributions and the average-case complexity of manipulating elections. Journal of Artificial Intelligence Research, 28, 157-181.

Von Krogh, G., Ben-Menahem, S. M., & Shrestha, Y. R. (2021). Artificial intelligence in strategizing: Prospects and challenges. Future of strategic management, 625-646.

Zuckerman, M., Procaccia, A. D., & Rosenschein, J. S. (2009). Algorithms for the coalitional manipulation problem. Artificial Intelligence, 173(2), 392-412.

Keywords: Manipulability, Rational Choice, Algorithmic Decision Making

Negative Externalities in Attention Economies

Adrian K. Yee*[1]

[1]Lingnan University (Dept. of Philosophy, Hong Kong Catastrophic Risk Centre) – Hong Kong SAR China

Abstract

A core aspect of human conscious experience is that in any given instance, whenever we pay attention to some things, we necessarily neglect attending to other things. While this may seem trivially obvious, there are considerable opportunity costs that arise as a direct corollary of our psychological attention being structured this way. Given that attention is a finite resource, and necessary for any agent to obtain information about goods and services, attention is a prime candidate as a subject of economic analysis (Greenberger 1974; Bueno 2017). In this sense, attention is a bounded resource that comes with intrinsic costs, both as internalities and externalities, rendering attention a potential object of economic analysis. An attention economy is an economic system in which consumers' psychological attention, of focusing on one item rather than another, is the primary source and measure of economic value, with products and services given more attention tending to be conferred higher value. Whereas classical political economists Smith, Ricardo, and Marx argued that economic value is constituted by labor, and 19th-century Marginalist School economists such as Edgeworth and Jevons that it is constituted by changes in psychological utility, contemporary social media companies active on the internet and corporate advertising has now created a relatively new economic reality in which our mind's psychological attention is a key constituent and cause of economic value. As social scientist Herbert Simon wrote (Greenberger 1971, 40-41):

(I)n an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.

Existing literature on attention economies has typically occurred in three fields: philosophy, psychology, and economics. In philosophy, most have focused on ethical issues,such as how social media usage can violate users' psychological autonomy (Aylsworth & Castro 2021; 2022), can generate toxic ecosystems which exploit our frequent failure to retain con- text sensitivity to claims made while using internet platforms (Miller & Record 2022), and lower overall social well-being (Williams 2018; Castro & Pham 2020; Bahrgava & Velasquez 2020). This philosophical work is critical for articulating different senses in which attention economies can produce negative side effects. In psychology, thinkers at least as early as Wil- helm Wundt, Hermann von Helmholtz, and William James were concerned with the limited nature of attention as a bounded resource, and attempted to develop empirical methods to test various hypotheses about objects of perception that compete for our attention (Festre & Garrouste 2015). Recent psychology work has made rapid developments in precise measures of visual attention processing with applications in marketing (Wedel & Peters 2007).

*Speaker

In economics, there have been Marxist analyses arguing that advertising companies exploit 'surplus value' via attention economies (Bueno 2017), arguments suggesting the revision of microeconomic models of consumer preferences in light of empirical work on psychological attention (Gabaix 2019), and arguments that the social welfare effects of social media are largely negative, including smartphone addiction, increased depression, and self-esteem issues (Alcott 2020; 2022). However, no work has been done as of yet to synthesize this literature and articulate a theory in which there are negative externalities (which we provisionally de- fine as 'harmful unpriced spillover effects' in attention economies that require measurement and government response towards to mitigate. This is especially so in that research on psychological models of attention in the context of social media activities remains under-studied despite increasing efforts (Lazer 2020).

This paper provides a theory of negative externalities in attention economies. Current models face difficult challenges articulating an account of information deprivation, which is a phenomenon especially afflicting information ecosystems within attention economies. Section 1 begins by articulating the general sociological and economic structure of contemporary attention economies. Section 2 articulates an account of informational preferences relevant to our discussion of negative externalities. Section 3 then surveys the extant microeconomic literature on externalities. Section 4 advances a theory of negative externalities in attention economies by way of the previous account of informational preferences. Section 5 provides a sketch of potential government policies by exploring the extent at which nudges and Pigovian taxes might mitigate the creation of negative externalities in attention economies.

Keywords: attention economics, social media, negative externalities, measurement, policy interventions

Predictions, Potential Outcomes, and Performativity

Sebastian Zezulka*[1] and Konstantin Genin

[1]Universität Tübingen – Germany

Abstract

Predictions can change the world. They do not only play an epistemic role in describing how the world will be in the future, but they often also have a causal effect on the world themselves: predictions can be performative. Performative effects are especially relevant in the case of automated decision-making. Pre- dictions, the output of a statistical model trained on historical data, here inform or even fully automate consequential decision-making. Such systems exist and are used worldwide for the allocation of social goods. Take as an example a public employment service in a social welfare state. Once a person registers as unemployed, a statistical model predicts their likelihood of becoming long-term unemployed. This risk score then influences the decision of which resources are made available to that person. The predictions shape decisions and the decisions are causally relevant to the likelihood of that person ultimately becoming long-term unemployed.

Two types of performative effects can be distinguished: self-defeating and self-fulfilling prophecies. Note that, throughout this paper, we adopt a causal understanding of performativity and leave open the relationship to a constitutive use. *Self-defeating prophecies* are predictions that are incorrect because making the prediction caused the predicted out- come not to occur. *Self-fulfilling prophecies* turn out to be correct because we have made the prediction. In this case, they can lead to statistical fatalism, a term we borrow from Hacking (1990, p. 115). That is, we take the consequences of our actions resulting from the predictions as evidence that things could not have been different.

In this paper, we illustrate how both types of performative effects arise in automated decision- making and can harm people. The prediction of the mortality risk of pneumonia patients to decide about hospital admission serves as case study for self-defeating prophecies. A statistical model predicted low mortality risk for pneumonia patients with asthma due to the aggressive treatment they typically receive, which results in good survival rates (Caruana et al., 2015). Under the proposed policy, this would have led to fewer hospital admissions for these patients, resulting in self-defeating prophecies. The prediction of long-term unemployment (LTU) introduced above is an example for a self-fulfilling prophecy. Under an Austrian policy proposal, people predicted to be at high risk of LTU would only be given restricted access to labor market programs (Allhutter et al., 2020), potentially trapping them in a cycle of seemingly accurate self-fulfilling predictions, ignoring the possibility of alternative resource allocations that could reduce LTU rates in this group.

Performative effects drastically complicate how we evaluate predictions, as pragmatic and epistemic considerations are entangled. Thus, the notion of accuracy becomes ambiguous.

*Speaker

In the case of self-defeating prophecies, the problem is that test set accuracy can be high, failing to indicate drastic failures once the model is deployed. In the case of self-fulfilling prophecies, accuracy values remain high even after deployment but for the wrong reason. The possibility that predictions steer people into a response that is evaluated as empirically adequate (self-fulfilling) or cause the prediction to fail by making the very prediction (self- defeating) requires value-laden decisions. Moreover, the notion of inductive risk must be revisited: values do not only come into play because the consequences of different kinds of error must be weighed but also because we worry that predictions might be right for the wrong reasons.

Many proposals have been made on how to deal with performativity, including ”endogenizing” performative effects, or making predictions that steer outcomes in socially desirable directions (Perdomo et al., 2020, Khosrowi, 2023). We argue that these approaches are misguided. Many undesirable performative effects arise because the predictions that in- form policy decisions are ambiguous between *predictions* that predict outcomes by implicitly guessing which policy will be adopted, and *counterfactual predictions* that predict outcomes for different policy options. Once we make the right distinction between prediction of actual and counterfactual, pragmatic and epistemic issues can be distinguished.

Methodologically, counterfactual predictions can be realized as potential outcomes. Potential outcomes predict the relevant *counterfactual* outcome under the policy options; they predict what *would* happen under the respective allocation decision. Therefore, they allow us to make accurate predictions of outcomes for the policy options without triggering outcome performativity. Similar arguments have recently been made by Tal (2023) in the case of hospital admission and by Khosrowi and van Basshuysen (2024) for recidivism predictions. We build on both but argue against Tal (2023) that the relevant counterfactual is not specified by ideal treatment conditions but by the set of available policy options. This allows us to propose a concrete methodological approach, as Khosrowi and van Basshuysen (2024) have called for.

To summarize, we develop a methodological proposal for dealing with performative effects in automated decision-making. Our contribution is threefold. First, we utilize the potential outcomes framework to conceptualize outcome performativity in policy contexts and show that estimating this target at a sufficient level of granularity is feasible with methods such as doubly robust machine learning. We discuss the necessary assumptions and limitations.

Second, we show that accurate predictions of potential outcomes are *decision-supportive*: they are (a) not self-defeating and (b) ”performative omnipredictors” for a rich set of policies (Kim and Perdomo, 2022). That is, they enable an optimal policy implementation of different policies without presupposing an allocation decision like maintaining the status quo.

Third, we argue that accurate predictions of potential outcomes are *discourse-supportive*: furnishing the relevant outcomes for various policy options to allow discussants with a variety of goals and values to engage in public discourse about which of a set of policy options represents a reflective compromise. It is for this reason that potential outcomes are the value-adequate estimation target that allows disentangling relevant pragmatic and epistemic issues in automated decision-making.

References

Doris Allhutter, Florian Cech, Fabian Fischer, Gabriel Grill, and Astrid Mager. Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. Frontiers in Big Data, 3, 2020.

Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Read- mission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’15. ACM, 2015.

Ian Hacking. The taming of chance. Cambridge University Press, 1990. ISBN 0521380146.

Donal Khosrowi and Philippe van Basshuysen. Making a Murderer. American Philosophical Quarterly, 61(4):309–325, October 2024. ISSN 2152-1123.

Michael P. Kim and Juan C. Perdomo. Making Decisions under Outcome Performativity, 2022.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative Prediction. In Hal Daumé III. and Aarti Singh, editors, Proceedings of the 37th International Conference on Machine Learning, volume 119 of Proceedings of Machine Learning Research, pages 7599–7609. PMLR, 2020.

Eran Tal. Target specification bias, counterfactual prediction, and algorithmic fairness in healthcare. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, AIES '23. ACM, 2023.

Keywords: Automated Decision, Making, Performativity, Causal Inference, Econometrics, Values in Science

Microfoundations in measurement and theory

Junyi Zhu*[1]

[1]Deutsche Bundesbank – Germany

Abstract

This paper clarifies that, to address the aggregation problem, macroeconomic models must return to the empirical aggregates of national accounts (NA), thereby reconciling theoretical constructs with observed macro outcomes. However, some fundamental micro–macro gaps emerge between the bottom-up, microfounded macro and macro measurement. The primary gap stems from the divergence between closed aggregation-rooted in micro reality and theory-and the open aggregation approach employed by NA, which implicitly im- poses macro restrictions that lack full microfoundations (Schlicht, 1985). In practice, two barriers-information friction and policy-oriented objectives-prevent macro measurement from adopting a closed aggregation framework. Using the aggregate production function and the contrast between Real GDP and Physical Value Added as illustrative examples, we demonstrate these barriers and rationalize NA's reliance on open aggregation. A microfounded explanation is achievable only when the corresponding micro restrictions are satisfied, ensuring consistency between bottom-up aggregation and macro measurement; otherwise, NA aggregates diverge from those derived from micro data. Recognizing these gaps, we argue that many current efforts to redistribute the aggregate data exhibit a paradoxical regression to representative agent thinking. Tentative solutions include the development of satellite ac- counts and/or the imposition of top-down micro–macro restrictions-drawing on index number theory and practice-on the bottom-up aggregation process.

Keywords: micro and macro gap, aggregation problem, index number

*Speaker