ANPOSS/ENPOSS/POSS-RT 2021 Joint Conference
List of Accepted Papers with Abstracts
Dates: March 4-7, 2021
Venue: Hitotsubashi University, Tokyo, Japan (Online)

17:00-18:20, March 4 (Thursday), Japan Standard Time
Session 1 "Algorithms and Reflexivity" (Chair: Eleonora Montuschi)

Henri Galinon. "Silicon Prophets. Computational Social Sciences, Reflexivity and Democracy"

Abstract: In recent years a number of computational social scientists have insisted that social sciences should take a major methodological turn to abide by better scientific standards of hypothesis testing based on prediction tasks – a common currency in natural sciences (e.g. recently Watts 2014, Hoffman & al. 2017, Salganik 2017. Compare to Weber 1968). They have argued that the digital age makes it possible in principle to design complex causal models, calibrate them on available sets of big data and test them on prediction tasks, a methodology which, as a matter of fact, has become standard procedure in the sort of data-mining-based applied social sciences practiced in R&D departments of giant internet firms and security agencies (Lazer 2009).

There are different reasons why old-fashioned social sciences have not given much significance to prediction tasks as validation procedures (e.g. complexity). One of them, as a number of social scientists have long argued, is that a distinctive characteristics of social sciences is their reflexivity in the sense that knowledge of social sciences causally influence the social development of the phenomena they study (e.g. Merton 1948, Guidens 1979, Henshel 1982, Barnes 1983, MacKenzie 2006 etc.). I shall recall in more details the mechanisms involved in such loops, illustrate briefly that they are not marginal phenomena and explain why this reflexivity feature has sometimes been taken to except social science from the methodology of natural sciences especially regarding the status of prediction. Then I shall confront this view with the newcomers' views that social phenomena can and should be predicted, from both an epistemic and a normative standpoint. Indeed I will argue first that reflexivity and predictivity are in no way epistemologically incompatible and recall how it is commonly achieved in a number of fields where high-profile predictors must take into account the causal effects of their prediction announcements (e.g. central banks). At the same time this observation will help to clarify the irreducibly ethical and political dimensions of such public forecasts whose validity (epistemic dimension) is entangled with choices (ethical dimension) in the realization a special kind of equilibrium (the self-fulfilling prophecy problem). In light of this entanglement I shall turn to the distinctive features of IA-based social predictions. My focus will be on analyzing the consequences of two of them: 1. the fact that these predictions are based on "IA"-style models and algorithms which can't be humanly checked nor understood (the opacity problem); and 2. that they (as to now) are dependent on private firms on a market where barriers to entry are very high (the dependency problem). I shall discuss the extent to which the opacity and dependency problems call for specific ethical and political attention to

predictive computational social sciences. I shall briefly ponder the significance of these normative considerations with regards to two other groups of normative considerations: first the emancipatory promises of computer social sciences to deal with major issues faced by modern societies (see e.g. Conte & al. 2002) and, second, more standard general normative considerations about the relationships between science and democracy (e.g. Kitcher 2001).

Miriam Aiello. "R-Awareness vs. Opaque Mind: The Problem of Reflexive Subjectivity between Social and Cognitive Sciences"

Abstract: In the last decades, a "reflexive turn" increasingly runs through the field of social sciences: on the one hand, sociologists attribute reflexive capacity to individuals, institutions and to the entire modern era, on the other hand, social scientists claim for a renewal of their disciplines by the expansion of epistemological reflexivity. A reflexive subjectivity should be able to understand itself as a part of its own scientific object (i.e. affecting and being affected by the social world) and, at the same time, to apply to itself the same objective criteria it uses to enquire this object. From a conceptual point of view, reflexivity conceived both as ontological property and as methodological issue includes three aspects (Ashmore 1989): 1) R-Reference, the property of referring to itself, 2) R-Awareness, as "benign introspection", 3) R-Circularity, which refers to the circular relationship between representations and reality.

In this talk, I will challenge Ashmore's claim (1989, 32) conceiving of R-Awareness as "rarely problematic".

Furthermore, I will suggest that the reflexive turn in social sciences is at odds with a specular and concomitant anti-reflexive turn in psychological and cognitive sciences that 1) points out the opacity and the limitedness of the discursive consciousness at justifying the reasons of the actions and the causes of behaviors and, consequently, 2) undermines that idea of robust subjectivity aprioristically assumed by social scientists to carry out the reflexive tasks.

Finally, I will claim that the reflexive effort in social sciences runs the risk of being ill-grounded, inasmuch as it overlooks crucial data provided by psychological and cognitive sciences which highlight the opacity of introspection and of the entire mental sphere. On the contrary, dealing with the cognitive anti-reflexive turn represents an opportunity for social science to expand and improve interdisciplinary collaboration in the understanding of the psychosocial phenomena.

References:

Archer M. (2007), Making our Way through the World. Human Reflexivity and Social Mobility, Cambridge: Cambridge University Press.

Archer M. (2013), 'Reflexivity', Sociopedia.isa, www.sagepub.net/isa/resources/pdf/Reflexivity2013.pdf.

Ashmore M. (1989), The Reflexive Thesis, Chicago: The University of Chicago Press.

Beck U., Giddens A., Lash S. (1994), Reflexive Modernization, Redwood City: Stanford University Press.

Bourdieu P. (2001), Science de la science et réflexivité, Paris: Seuil.

Carruthers P. (2011), The Opacity of Mind, Oxford: Oxford University Press.
Dennett D. (1991), Consciousness Explained, Boston: Little Brown.
Di Francesco M., Marraffa M., Paternoster A. (2016), The Self and its Defences. From Psychodynamics to Cognitive Science, London: Palgrave-McMillan.
Giddens A. (1991), Modernity and Self-Identity. Self and Society in the Late Modern Age, Cambridge: Polity Press.
Jervis G. (1995), Presenza e identità, Milano: Garzanti.
Merton R.K. (1968), Social Theory and Social Structure, New York: Free Press.
Nisbett R.E., Wilson T.D. (1977), Telling More than We Can Know: Verbal Reports on Mental Processes, «Psychological Review», 84: 231-59.
Wilson T.D. (2002), Strangers to Ourselves: Discovering the Adaptive Unconscious, Cambridge (MA): Harvard University Press.
Woolgar S. (1984), A Kind of Reflexivity, in Id. (1988), Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge, London and Beverley Hills, Calif: Sage.

10:30-12:30, March 5 (Friday), Japan Standard Time

Session 2 "Experiments and Intervention" (Chair: Paul Dumouchel)

Aydin Mohseni. "HARKing: From Misdiagnosis to Misprescription"

Abstract: In a 2019 article in 'Nature', the author, psychologist Dorothy Bishop, describes HARKing as one of "the four horsemen of the reproducibility apocalypse," along with publication bias, low statistical power, and p-hacking (Bishop, 2019, p. 435). The practice of HARKing---hypothesizing after results are known---is commonly maligned as undermining the reliability of scientific findings. There are several accounts in the literature as to why HARKing undermines the reliability of findings. Scholars have argued that HARKing undermines frequentist guarantees of long-run error control, (Rubin, 2017) that it violates a broadly Popperian picture of science, (Mayo, 2019) and misrepresents hypotheses formulated ex post to observing the data as those formulated ex ante (Kerr, 1998). I argue that none of these accounts correctly identify why HARKing can undermine the reliability of findings, and that the correct account is a Bayesian one. Further, I show how misdiagnosis of HARKing can lead to misprescription in the context of the replication crisis in the social and biomedical sciences.

I will show that HARKing can indeed decrease the reliability of scientific findings, but that there are conditions under which HARKing can actually increase the reliability of findings. In both cases, the effect of HARKing on the reliability of findings is determined by the difference of the prior odds of hypotheses characteristically selected ex ante and ex post to observing data. To make this precise, I employ a standard model of null hypothesis significance testing in which I provide necessary and sufficient conditions for HARKing to decrease the reliability of scientific findings.

Understanding HARKing is important on at least two counts. Historically, HARKing is closely tied to questions regarding the relationship between prediction and accommodation. These questions have engaged philosophers at least as early as Mill (1843), were made central in the

philosophy of science by Popper (2005) and continue to be of concern in contemporary discussions in scientific epistemology (cf. Hitchcock & Sober (2004), Douglas & Magnus (2013), and Worrall (2014)). HARKing is also imputed to be among the questionable research practices contributing to the crisis of replication in the social and biomedical sciences. A better understanding of HARKing sheds light on both these issues.

References:

Bishop, D. (2019). Rein in the four horsemen of irreproducibility. Nature. 568: 435.
Hitchcock, C. and E. Sober (2004). Prediction versus accommodation and the risk of overfitting. British Journal for the Philosophy of Science. 55(1): 1–34.
Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. Personality and Social Psychology Review. 2(3): 196–217.
Mayo, D. (2019). Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars. Cambridge University Press.
Mill, J. S. (1843/2011). A System of Logic, Ratiocinative and Inductive. Cambridge University Press.
Douglas, H. and P. D. Magnus (2013). State of the field: Why novel prediction matters. Studies in the History and Philosophy of Science Part A. 44: 580–589.
Worrall, J. (2014). Prediction and accommodation revisited. Studies in History and Philosophy of Science Part A. 45: 54-61.

Alessandro Del Ponte. "Beware of the Proximity Heuristic! Understanding the Potential and Pitfalls of Experimental Research in Political Science"

Abstract: Experiments have gained popularity in political science because they allow researchers to make clean causal inferences. However, common experience suggests that the broad adoption of experiments as a method of inquiry has not been accompanied by proper awareness of their potential and pitfalls. Without sufficient foundations in the philosophy of science behind experiments, researchers may struggle to distinguish a good experiment from a bad one. Even more problematic, they may use misguided heuristics that classify types of experiments as better or worse. A particularly misled rule of thumb is to judge the quality of experiments through the proximity heuristic: an experiment is better the closer it gets to the target population or political behavior that it is intended to study. Accordingly, the following trends emerged in the discipline: field experiments are often rewarded as better research than lab experiments; lab-in-the-field experiments are considered better than lab experiments; survey experiments using nationally representative samples are automatically better than the survey experiments using convenience samples. As appealing as it may sound, the proximity heuristic is based on the incorrect premise that the ideal experiment can lead us to firm conclusions that withstand the test of generalizability across units, treatments, observations, settings (Cronbach, 1982), and time. Instead, experiments are models of the world that should be judged for their usefulness (Clarke and Primo, 2012; Friedman; 1953) to understand politics. Political scientists try to create models of political behavior and uncover the

mechanisms that govern it. In the quest to illuminate the mysteries of political behavior, a good experiment will shine a spotlight on one aspect of politics, until the next experiment refines the picture, confirming or contradicting prior evidence (Popper 1934). This monumental effort is made of many components, each of which is essential to paint a fuller picture of politics. Just like a wind tunnel is no less essential than a testing circuit to design a race car, a lab experiment is no less valuable than a field experiment to study politics. At the same time, experiments are just one method of scientific inquiry: the ability to establish the effect of x on y is valuable but by no means exhaustive to illuminate politics. Science advances with the combined spotlights of the entire toolkit that political scientists have in their arsenal: thorough description, (formal) theorization, (incentivized) experimentation, historical, qualitative, and statistical observation, to name some of the most prominent techniques. A fuller awareness of the place that experiments occupy in scientific inquiry and their important limitations will enable political scientists to harness the potential of experiments to their fullest extent.

Tung-Ying Wu. "Interventionism and Causal Analysis in Social Sciences"

Abstract: Causal analysis plays a key role in quantitative research in social sciences. Several attempts have been made to relate quantitative causal analysis in social sciences within an interventionist framework. The interventionist theory of causation (hereafter, interventionism) aims to non-reductively illuminate the nature of causal relations by interpreting a causal claim as a claim about the outcome of a hypothetical experiment where an intervention on a cause changes its effect. Previous papers have only focused on limited causal analysis techniques such as the potential outcome framework and instrument variables. To the best of my knowledge, however, few literatures have looked specifically at other methodologies such as regression discontinuity designs and difference-in-differences from an interventionist perspective. This paper sets out to provide an interventionist characterization of these methods. Two ideas to make sense of the interventionist accounts of them are presented: first, the notion of intervention can explain the continuity assumption that must hold for the validity of regression discontinuity designs. Second, under a wider interpretation of "holding fixed" or "controlling" variables in the background, the common trend assumption that is vital for the application of the difference-in-differences can satisfy the interventionist requirement of direct causation. It is argued that these methods of causal analysis are readily justifiable on the basis of interventionism and motivate interventionism as a philosophical foundation for quantitative causal analysis in social sciences.

15:40-17:00, March 5 (Friday), Japan Standard Time

Session 3 "Causal Inference and Mechanisms" (Chair: Paul Roth)

Masahiko Igashira. "A Mechanism for Justification of General Claims in History and the Challenges it Might Involve"

Abstract: In recent years, a debate on research methodology in the social sciences has been ongoing. It is regarding the issue of whether general and causal claims can be adequately justified from

studies based on a small number of cases, which is typical of so-called qualitative approaches. The critics who have regression analysis and statistical tests, based on statistical frameworks, in their minds have argued that when the number of cases relied upon is small, errors and biases cannot be sufficiently removed to provide ample justification. The advocates have responded in a variety of ways, such as even though research based on a small number of cases may not allow for theoretical verification, it can be useful in theory construction; if deterministic causality is assumed, the claim can be justified even from a small number of cases; research based on a small number of cases can also contribute to theory testing by relying on prior theory and it can make a partial contribution to theory verification in terms of construction of causal models.

These controversies are not irrelevant to historiography, which has to face the problem of scarcity of historical data. The insufficiency of available data in history does not allow for a convincing statistical analysis, and when any generalized or causal claims are asserted in historiography, they face the same kind of criticism that confronts qualitative approaches in the social sciences.

In this paper, I will outline a case in which a general (limited scope) claim is made and explain the mechanism of its justification. Specifically, I will delineate a particular case of historical practice, in which a common understanding of "matters that people are reluctant to talk about" in a certain region is achieved on the basis of a limited number of testimonies presented in a certain trial and will show that there is a mechanism that can make the conclusion appear convincing despite the small number of testimonies. I will further demonstrate that this mechanism enables us to avoid criticism directed at a research methodology just because it is based on a small number of cases. Moreover, I will evaluate the uncertainty factors that might become problematic when this mechanism is used.

In the course of the philosophy of history, there have been various debates about the forms of historiographical explanations. The mechanism that will be discussed in this paper will show how the parts used in any of these forms of explanation, that is, general statements with limited regions and periods, can be legitimately claimed in historical practice. If the discussion in this paper proceeds successfully, it will provide a clue to bring some harmony and make some progress in this debate on research methodology in the social sciences.

Saúl Pérez-González. "The Roles of Evidence of Mechanisms in Social Sciences"

Abstract: In social sciences, extrapolating a causal claim (e.g., the efficacy of an educational intervention) from a study population to another population of interest is a problematic issue. There are often significative differences between both populations and between their cultural and social contexts. Given that statistical evidence in isolation seems unable to overcome those difficulties, some authors have argued that evidence of mechanisms would be a valuable resource (Grüne-Yanoff, 2016; Marchionni & Reijula, 2019; Steel, 2008). The aim of this paper is to discuss whether and to what extent evidence of mechanisms could contribute to causal extrapolation in social sciences.

In order to analyse the potential contribution of evidence of mechanisms, a distinction

between a supporting and a disproving role is introduced. On the one hand, if the relevant mechanisms at work in the study and the target population are highly similar in the relevant aspects, the extrapolation of the causal claim is supported. On the other hand, if the relevant mechanisms at work in the study and the target population differ in relevant aspects, the extrapolation of the causal claim is disproved. For evaluating the actual relevance of each role, three aspects are considered: (i) required information about the relevant mechanisms; (ii) difficulties faced by mechanism-based approaches, and (iii) real cases that exemplify it.

In the first place, the disproving scenario is evaluated. It is argued that evidence of mechanisms can provide basis for concluding that the extrapolation of a causal claim is not justified. Firstly, comparing the relevant mechanisms in both populations in some aspects in which they are likely to differ may result in the identification of a relevant difference between them (Steel, 2008). Secondly, given the complexity of social mechanisms, masking or irregularity of mechanisms would hardly exactly compensate the identified differences and undermine the negative conclusion. Finally, the main case study in support of the mechanisms approach to extrapolation in social sciences—i.e., the Bangladesh Integrated Nutrition Project (BINP)—exemplifies the disproving scenario.

Regarding the supporting role, it is argued that its actual relevance is uncertain. Firstly, given our usual limited knowledge about target populations, it is hardly possible to specify the degree of similarity between the relevant mechanisms (van Eersel et al., 2019). Secondly, masking and irregularity threat the supporting scenario. Even if relevant mechanisms at work in the study and the target population are highly similar, interfering mechanisms or changes in mechanisms' behaviour could modify the causal relationship held in the target population (Howick et al., 2013). Finally, in the debate about causal extrapolation in social sciences, no real case that exemplifies the supporting scenario has been identified.

References:

Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. Economics & Philosophy, 32(3), 463-483.

Howick, J., Glasziou, P., & Aronson, J. K. (2013). Problems with using mechanisms to solve the problem of extrapolation. Theoretical medicine and bioethics, 34(4), 275-291.

Marchionni, C., & Reijula, S. (2019). What is mechanistic evidence, and why do we need it for evidence-based policy? Studies in History and Philosophy of Science Part A, 73, 54-63.

Steel, D. (2008). Across the boundaries: Extrapolation in biology and social science. Oxford University Press.

van Eersel, G. G., Koppenol-Gonzalez, G. V., & Reiss, J. (2019). Extrapolation of experimental results through analogical reasoning from latent classes. Philosophy of Science, 86(2), 219-235.

18:30-20:30, March 5 (Friday), Japan Standard Time

Session 4 "Theory and Reality" (Chair: Jesús Zamora-Bonilla)

Bele Wollesen. "A Heuristic Approach to Manipulatively - Simulating Simple Strategic Voting"

Abstract: Measures of manipulability in voting theory usually follow the Gibbard-Satterthwaite interpretation of manipulation (e.g Nitzan-Kelly Index). Yet, there is some debate about whether manipulation in this sense is, in fact, a problem in practice and whether the complexity of computing successful strategies provides a barrier to voting strategically (see Conitzer, Walsh, and Xia 2011). Consequently, some authors have wondered whether Social Choice promotes too pessimistic views (see Regenwetter, Grofman, Popova, et al. 2009). Furthermore, it is well established in the political science literature that voters make use of heuristics to deal with complex information and decisions (see Popkin 1995).

Given all of this, it seems like that the formal analyses of the manipulability of voting rules need to be updated in the light of the political science literature to be relevant for real-world application. Therefore, this paper aims to show that taking alternative behaviors of voters seriously leads to important results for formal work in voting theory. Towards this end, the paper describes the result of a multi-agent simulation with voters using simple heuristics to vote strategically and finds that voting rules perform very differently than one would expect based on traditional rankings of manipulatively such as the Nitzan-Kelly Index.

For the simulation voters adapt a Win-Stay Lose-Switch (WSLS) heuristic. Voters using a WSLS will keep doing an action if the previous result was a success and switch the action if it was not. WSLS can better describe how agents behave in many circumstances than more complex decision mechanisms like maximization or reinforcement learning. Therefore, WSLS is one of the simplest, yet plausible strategies of voting that can deviate from sincere voting and by incorporating a strategic element.

The simulation tests different heuristics that follow the Win-Stay Loose-Switch schema. These heuristics all differ with respect to their aspiration level i.e. which rank of the candidate a voter still considers a "win". It focuses on several positional voting rules and includes varying populations of candidates and voters.

The simulations show that among the considered voting rules there is no voting rule that is always, i.e. across all heuristics, most robust or always least robust. Moreover, there is no heuristic that makes every voting rule always robust. However, if the voting rule tracks the aspiration level, i.e. if it gives only as many votes as the rank of the aspiration level, voting rules are always robust against strategic voting.

These results indicate that it is not just important how voters vote or which theoretical possibilities there are for manipulation —the interaction between actual voting behavior and voting rules is equally important. Because of this, measures as the Nitzan-Kelly index do not suffice to make informed decisions and can give misleading recommendations. Moreover, the results give rise to the possibility that robustness comes somewhat cheap if satisfaction and engagement in strategic voting are linked.

References:

Conitzer, Vincent, Toby Walsh, and Lirong Xia (2011). “Dominating manipulations in voting withpartial information”. In: Twenty-Fifth AAAI Conference on Artificial Intelligence.
Popkin, Samuel L (1995). “Information shortcuts and the reasoning voter”. In: Information, partic-ipation and choice: An economic theory of democracy in perspective, pp. 17–35.
Regenwetter, Michel, Bernard Grofman, Anna Popova, et al. (2009). “Behavioural social choice: astatus report”. In: Philosophical Transactions of the Royal Society B: Biological Sciences364.1518, pp. 833–843.

Kevin Leportier. “Opportunity Criterion and Preference for Freedom”

Abstract: Normative economics aims at producing evaluations of economic situations. When comparing different states (for example, before and after the imposition of a tax), the traditional method, grounded on a normative criterion of preference satisfaction, begins by supposing that the agents involved have preferences which remain the same across the different states and then asks what state better satisfies the preferences of the agents. But the results of behavioral economics suggest that agents’ preferences do not usually have this character of stability. This discrepancy between the usual hypothesis of preference satisfaction and the results of behavioral economics raises a “reconciliation problem” (McQuillin and Sugden 2012), which has been addressed in different manners. In a recent book, Robert Sugden (2018) has proposed to replace the traditional preference satisfaction criterion with an opportunity criterion, according to which a situation is better if it gives more opportunity to all agents. This criterion does not require any information about agents’ preferences to be applied and can therefore be seen as overcoming the “reconciliation problem”.

But this criterion may be lacking in other respects. In particular, I argue that it fails to acknowledge that situations where agents are offered the possibility to lose future opportunities—without any benefit—can be bad, even from the point of view of the agents. If agents may change their minds, if their tastes are not fixed, and if, accordingly, they value flexibility, it would be a mistake for them (in the absence of strategic interaction) to commit to some course of action at an earlier stage without receiving any compensation. This freedom to lose one’s freedom—I would call it an “opportunity trap”—is problematic if agents are prone to making mistakes. This argument is general and can as well be directed against the traditional preference satisfaction criterion. How can it be grounded?

I claim that in order to reach the conclusion that opportunity traps are bad, we only need to endorse the principle that evaluations of economic situations should be derived from the evaluations of the agents themselves. This principle, I argue, is at the heart of traditional welfare economics as well as Sugden’s project. I show that if we are to apply this principle consistently, we need to presuppose that agents themselves accept it. And agents cannot accept the principle without valuing in a certain way their freedom, which is to say that they must have a preference for freedom (defined in the precise terms of Puppe 1998). This attribution of a preference for freedom to agents has nothing to do with their psychology, but is a consequence of a (consistent) application of the normative principle defined earlier. Finally, turning to multi-period decision problems, it can be

shown that if agents have a preference for freedom, but may make mistakes, then respecting this preference leads to confer a negative value to opportunity traps, which expose fallible agents to lose their freedom. Far from leading to good old laissez-faire policies, attributing to agents a general preference for freedom can be seen as a motivation for regulation, at least in those situations which present opportunity traps to fallible agents.

Edoardo Peruzzi and Gustavo Cevolani. "Defending (De-)idealization in Economic Modelling: A Case Study"

Abstract: Theoretical models in science, and in economics in particular, typically contain idealizations of various kinds. This is widely acknowledged by both scientists and philosophers. However, there is much less consensus about the nature and epistemic role of those idealizations. One important view of idealizations refers to the notion of "concretization" (or de-idealization). According to this idealization-concretization view (IC for short in the following), important episodes of theoretical change in science can be construed as instance of a process that goes from more idealized to more realistic models of phenomena (Nowak 1980; Cools, Hamminga, and Kuipers 1994; Niiniluoto 2002, 2012, 2018; Hindriks 2013).

In recent discussion, the IC view of theoretical change has been discussed and strongly criticized by scholars interested in the methodology of economics (see, in particular, Alexandrova 2008; Alexandrova and Northcott 2009; Reiss 2012). Despite having different views of idealizations and economic modelling, such critics agree on one point: that de-idealization strategies are actually not used in economic modelling, for the good reason that they are either unfeasible or useless. In short, in their view IC is simply not a viable approach to understanding the nature and role of idealizations, at least as far as theoretical models in economics are concerned.

This paper aims at rebutting this criticism and defend the viability of the IC view in economics. In particular, we put forward two main claims. First, we argue that de-idealization strategies often work and can indeed be applied to the practice of real-world economic modelling. Second, we show that such strategies are actually employed by working economists, pointing to a relevant case-study concerning so-called information economics and the theory of imperfect competition. In our perspective, the IC view then provides both a fruitful theoretical perspective on idealized models in economics, and a way of reconstructing the actual historical development of the discipline.

We proceed as follows. In section 1, we present topic and goals of the paper, and describe its structure. In section 2, we introduce the IC view and summarize the discussion on idealizations and concretizations, with reference to both natural and social sciences. We then focus on the case of economics and discuss recent criticism of the IC view.

To better assess such criticism, in section 3 we present a case-study from the economic literature. This concerns a sequence of theoretical models, originated by the seminal paper of Varian (1980), advanced to rationalize price dispersion in markets for homogeneous goods. We also show how the change from one model in the sequence to the other can be construed as a de-idealization process in the sense defined here. In section 4, we re-evaluate the IC view and its critics in the light

of our case-study. We argue that critics tend to underestimate the importance of de-idealization techniques in economic modelling. Indeed, a striking contrast exists between the claim that de-idealizations are unfeasible and the economists' effort to relax idealized assumptions to get models with better explanatory and predictive power. We conclude that recent criticism of the IC view is ill-founded, and that de-idealization strategies are not only possible but also widely employed in economics. Finally, in section 5 we summarize our discussion and provide a tentative assessment of its implications for the ongoing discussion of realism in economic methodology.

10:30-11:50, March 6 (Saturday), Japan Standard Time

Session 5 "Ethical and Social Perspectives" (Chair: Stephen Turner)

Horacio Ortiz. "Weber Facing Nietzsche: A Tragic Ethics of Social Sciences"

Abstract: Based on a systematic comparison of the complete published work of Max Weber and Friedrich Nietzsche, this paper shows the importance of Nietzsche's analysis of the relation between science, morality and nihilism in Weber's understanding of his own scientific endeavor as "link and motive force" of the disenchantment of the world. Nietzsche explores the will to truth as a movement that destroys all other values, until it uncovers itself and reaches the prospect of nihilism. In a similar reflection, Weber understands Truth to be the value of science, which at once delegitimizes all other values' claims to universality, situating them within the confines of time and space, and delegitimizes its own claims to universality as just another value with the same limitations.

Applying his method of analysis of social action to social science itself, Weber follows Rickert to consider that although values other than truth are the ones that give meaning to the objects of social sciences, the scientist must also put these values aside during scientific analysis. But following Nietzsche's analysis of the death of God, Weber considers that any attempt to realize values in reality is bound to fail, because values and the phenomenal world do not share the same logic, and because asserting one value implies hurting other values within the same person. Yet, Weber considers that it is in this attempt that the human becomes true or authentic ("echt"). What for Nietzsche is then active nihilism, which asserts assertion without having anything else to assert, becomes for Weber the "tragic character" of ethical action and of social sciences themselves.

The paper concludes by showing the importance of this tragic ethics of social sciences for the definition of Weber's analytic categories and research objects. In particular, this understanding of ethical action as inherently tragic structures the notion of meaningful social action, and the relation between instrumental and value rationality. The tragic ethics of social sciences also informs the tension in Weber between universalist and historicist understanding of "logics", as well as his holistic category of "Okzident". Thus, the paper shows that Nietzsche's reflections on science, to which Weber makes reference in several important places of his writing, play a very important role in the way in which Weber understood his own endeavor as social scientist, and what he hoped social sciences could achieve. This analysis is thus important not only to understand Weber, but also to clarify the limits and possibilities of his project for contemporary social sciences and for the

philosophical debates that take inspiration from his work.

Armin Schulz. "Undermining Purposes: Institutional Corruption and Social Functionalism"

Abstract: While corruption has long been recognized as a major social problem, it has only recently become clear that one of the major forms of corruption does not center on public officials abusing their power for private gain, but on the subversion of entire social institutions (Thompson, 1995; Lessig, 2013; Miller, 2017). However, the exact characterization of this institutional form of corruption remains controversial (Thompson, 1995; Lessig, 2013; Miller, 2017). Exactly which actions subvert the relevant institution, and exactly why is it the case that these actions subvert the institution? This paper seeks to make progress in answering these questions.

Specifically, the paper develops and defends a novel theory of institutional corruption (IC), according to which IC is the result of an individual or collective agent acting in ways that prevent a social institution from partially or fully fulfilling its function. In turn, the function of a social institution is spelled out in line with what is argued to be the most compelling account of social functionalism in the literature: Presentist Social Functionalism (PSF).

PSF sees the function of a social institution as those of its features that increase its expected reproductive or survival success in the current socio-cultural environment. This theory is shown to have several important advantages over alternative versions of social functionalism. On the hand, in virtue of its ahistorical nature, PSF sidesteps the "missing mechanisms argument" that befalls the traditional, historical versions of social functionalism (Elster, 1979; Pettit, 1996). On the other hand, PSF improves on the counterfactually-grounded account of Pettit (1996) by being focused on the assessment of actual causal patterns, not on the evaluation of complex counterfactuals.

With PSF in the background, the paper shows that a novel characterization of IC can be developed that has a number of highly desirable features: it is general, fully spelled out, situated in a wider functionalist approach towards the social sciences, and does justice to the complexity of IC. In this way, this novel characterization of IC is in a position to improve on existing accounts of this phenomenon, which—whatever other virtues they have—are either insufficiently general (Thompson, 1995; Warren, 2015), insufficiently well spelled out (Lessig, 2013), or overly moralistic and restricted in outlook (Miller, 2017). The present theory of IC can thus help advance our understanding of, and policy responses to, this critically important social phenomenon.

References:

Elster, J. (1979). Ulysses and the Sirens. Cambridge: Cambridge University Press.

Lessig, L. (2013). "Institutional corruption" Defined. Journal of Law, Medicine & Ethics, 413, 553-555.

Miller, S. (2017). Institutional Corruption: A Study in Applied Philosophy. Cambridge: Cambridge University Press.

Pettit, P. (1996). Functional Explanation and Virtual Selection. The British Journal for the Philosophy of Science, 47(2), 291-302.

Thompson, D. F. (1995). Ethics in Congress: From Individual to Institutional Corruption. Washington, DC: Brookings Institution.
Warren, M. E. (2015). The Meaning of Corruption in Democracies. In P. M. Heywood (Ed.), In Handbook of Political Corruption (pp. 42–55). London: Routledge.

15:00-17:00, March 6 (Saturday), Japan Standard Time

Session 6 "Symposium: From Brain Structures to Cognitive Functions: Philosophical and Neuroscientific Perspectives on Reverse Inference" (Organizers: Fabrizio Calzavarini, Gustavo Cevolani and Davide Coraci; Chair: Alban Bouvier)

Fabrizio Calzavarini, Gustavo Cevolani and Davide Coraci. "Symposium: From Brain Structures to Cognitive Functions: Philosophical and Neuroscientific Perspectives on Reverse Inference"

Symposium Abstract: Since the early days of functional resonance magnetic imaging (fMRI), this technique is being used in two different ways. First, neuroscientists build brain maps by studying which regions are activated by different mental processes (as elicited by different tasks, e.g., face recognition or language processing). This is called forward inference, from mental processes to their putative neural correlates (Henson 2006). Second, researchers routinely employ the inverse reasoning strategy, inferring from specific activation patterns to the engagement of particular mental processes. This reverse inference plays a crucial role in many applications of fMRI, both inside and outside cognitive neuroscience. These include the diagnosis of disorders of consciousness (like vegetative state or coma) in patients with acquired brain injury (Tomaiuolo et al. 2016), the well-known experimental studies of moral reasoning as pioneered by Greene et al. (2001), and most studies in so-called neuroeconomics (Bourgeois-Gironde 2010). In recent years, reverse inference has attracted a great deal of attention, especially after neuroscientist Russell Poldrack (2006) denounced an uncontrolled "epidemic" of this reasoning pattern, cautioned against its (improper) use, and pointed to its crucial weakness. He also sketched a Bayesian analysis and speculated on how to remedy this situation. In further work, Poldrack and collaborators applied machine learning and data mining techniques to automatically explore big fMRI data sets in order to extract relevant correlations between mental processes and activation patterns to be used in making reverse inference more robust and reliable (the NeuroSynth project, see Yarconi et al. 2011). The debate is still open, and the present methodological status of reverse inference is highly controversial (Glymour and Hanson 2016; Hutzler 2014; Machery 2014; Nathan and Del Pinal 2017; Poldrack 2008, 2011; Weiskopf 2020). In addition, a number of philosophical questions about reverse inference remain open. For instance, is reverse inference a form of abductive reasoning? Has it mainly a heuristic role (i.e., that of generating new explanatory hypotheses and assisting discovery), or also a justificatory role (i.e., that of evaluating and possibly accepting selected hypotheses)? In the latter case, can reverse inference provide confirmation to hypotheses concerning cognitive processes? Is reverse inference only correlational, or also causal/explanatory? What is the role of analogical reasoning in inferring from brain activations to cognitive processes? The participants of

the symposium will explore these open issues in a multidisciplinary framework

Abstracts of the Contributed Papers

Fabrizio Calzavarini. "Abductive Reasoning in Cognitive Neuroscience: Weak and Strong Reverse Inference"

Abstract: Abductive inference is reasoning backward from facts to their possible explanations or from effects to their possible causes. It is at play in a wide array of contexts, from everyday life to science. In cognitive neuroscience, researchers often infer from specific activation patterns to the engagement of particular mental processes. This "reverse inference" plays a crucial role in many applications of fMRI, both inside and outside cognitive neuroscience. Poldrack (2006) himself noted that reverse inference is a pattern of abductive reasoning but did not develop further this comparison. In this talk, we argue that the current debate on reverse inference overlooks an important distinction. In the philosophical debate, it has become clear that that there are at least two different ways—respectively, a 'weak' and a 'strong' one—of assessing the proper role of abductive inference. According to the first, weak interpretation, abduction has a primary discovery (or 'strategic' or 'heuristic', see Schurz 2017 p. 203) function, that of suggesting or finding promising or 'testworthy' hypotheses which are then set out to further inquiry or empirical testing. According to the second, strong (or justification) reading, abduction can be formulated as a rule of acceptance, since it gives reasons to tentatively accept its conclusion as the 'best' explanatory hypothesis among the available ones. We suggest that the heuristic and the justificatory role of abduction in cognitive neuroscience can be usefully separated, by distinguishing a weak and a strong form of reverse inference. We claim that, although strong reverse inference may be often fallacious, weak reverse inference plays an essential role as a search strategy that tells us which explanatory conjectures we should set out first for further empirical inquiry; in other words, reverse inference can suggest a short and most promising (though not necessarily successful) path through the exponentially explosive search space of possible explanatory hypothesis.

Davide Coraci. "Reverse Inference and Bayesian Confirmation in Cognitive Neuroscience"

Abstract: Reverse inference is a crucial inferential strategy widely employed in cognitive neuroscience to derive conclusions about the engagement of cognitive processes from patterns of brain activation. A classical example is Greene et al. (2001), who derived implications for philosophical theories of moral reasoning starting from neuroimaging data relative to evaluation tasks in trolley dilemmas. Following the influential critical analysis advanced by Poldrack (2006), a hot debate now targets the methodological statues of reverse inference, which is seen with increasing skepticism in the neuroscientific community. On the one hand, reverse inference is a logically invalid piece of reasoning, reflecting the fallacy of "affirming the consequent"; on the other hand, it can be defended as an instance of abductive reasoning in the sense of C. S. Peirce, or as an inference to the best explanation. In the meantime, neuroscientists are developing tools like the Neurosynth project (Yarkoni et al. 2011), helping researchers to improve the reliability of reverse inference through systematic, computer-aided meta-analyses. In this paper, we offer an assessment of this debate and

advance some positive suggestions about the methodological status of reverse inference from the perspective of Bayesian philosophy of science. First, we precisely characterize reverse inference by analyzing relevant case studies in the neuroscientific literature (e.g., Liebeman and Eisenberger 2015). Second, we survey different proposals in the literature, that conceptualize reverse inference either as Bayesian inference (Poldrack 2006; Hutzler 2014), in purely "likelihoodist" terms (Machery 2014), or as a form of abductive reasoning (Bourgeois-Gironde 2010). Third, we suggest that the notion of confirmation as studied in Bayesian philosophy of science (Crupi 2020, Sprenger and Hartmann 2019) can help to clarify various aspects of reverse inference. In a nutshell, we argue that neuroscientists may actually tackle the problem of reverse inference by choosing the most confirmed among the competing cognitive hypotheses (Poldrack 2006, Hutzler 2014, Cauda et al. 2019). Given the great variety of confirmation measures discussed by philosophers (Festa and Cevolani 2017, Crupi 2020, Sprenger and Hartmann 2019), our analysis also shed new light on promising ways of systematizing and improving current discussion concerning robust inferential strategies in cognitive neuroscience.

Vincenzo Fano and Stefano Calboli. "Reverse inference and Analogical Reasoning"

Abstract: Neuroimaging functional techniques have dramatically increased our knowledge of human cognition making it possible to map neurophysiological states with mental processes under highly controlled experimental environments. The reverse inference, which is the practice of inferring the engagement of cognitive processes from the neurophysiological activation, is an inferential strategy on which neuroscientists persistently rely on. Nevertheless, taking advantage of such inferential strategy is highly controversial, especially due to the lack of selectivity of brain regions, i.e. the fact that each brain region plays a role in many different cognitive processes (Poldrack 2006). In this talk we present the methodological and practical precautions advanced in literature to address the multi-functionality of the brain and guarantee the correct application of inverse inferences (Poldrack 2006, Nathan and De Pinal 2017, Hutzler 2014, Machery 2014). We then discuss the extent to which the precautions advanced succeed in ruling out the careless uses of reverse inferences and argue that none of them allow to assign a causal/explanatory status to the conclusion made through reverse inference. We argue that analogical reasoning could play a pivotal role to raise the epistemic value of the conclusion drawn through reversal inference. More precisely, we suggest to advance along the trajectory indicated independently by both De Pinal and Nathan (2013) and Hutzler (2014), hence take advantage of sequences of specific tasks to find out the cases in which subsist analogies between the causal links among mental processes and the causal links among ne

17:10-19:10, March 6 (Saturday), Japan Standard Time

Session 7 "Measurement and Description" (Chair: Alban Bouvier)

Philipp Schönegger. "Experimental Philosophy and the Incentivisation Challenge: A Proposed Application of the Bayesian Truth Serum"

Abstract: A key challenge in social science research is how to incentivise subjects such that they take the questions seriously and answer honestly. If subject responses can be evaluated against an objective baseline, a standard way of incentivising participants is by rewarding them monetarily as a function of their performance. However, the subject area of experimental philosophy is such that this mode of incentivisation is not applicable as participant responses cannot be scored along a true-false spectrum by the experimenters. We claim that experimental philosophers' neglect of and claims of unimportance about incentivisation mechanisms in their surveys and experiments has plausibly led to poorer data and worse conclusions drawn. As a solution to this, we suggest adopting the Bayesian Truth Serum, an incentive-compatible mechanism used in economics and marketing. It is designed for eliciting subjective data and rewards participant answers that are surprisingly common. We argue that the Bayesian Truth Serum (i) adequately addresses the issue of incentive compatibility in subjective data research designs, and (ii) that it should be applied to the vast majority of research in experimental philosophy.

M.A. Diamond-Hunter. "The Limits of Accuracy for Retrospective Description of Racial Groups"

Abstract: It is taken as non-controversial that how groups of human beings are identified in the past is relevant to both academic and non-academic concerns. With respect to non-academic contexts, how people conceive of their historical connections to groups has an effect upon how they see themselves — whether the group aspect is racial, sexual identity, national identity, linguistic, regional, economic, or social (to name a few). How people are connected to historical groupings is also important in academic contexts, especially for researchers and governmental agencies that are tasked with doing relevant empirical work that connects with group labeling. Non-exhaustive examples of this include demography, epidemiology, economics, sociology, sociolinguistics, and political science. There are plenty of other reasons why people are (implicitly or explicitly) concerned with the ontology of groups: there are claims made about group connections (e.g. free trade agreements between the Brazil and the EU); claims made about the features of groups; the invoking of criteria for membership in groups (e.g. residential neighborhood for the purposes of being part of a post code); and policy decisions are made based upon what are taken to be group membership (e.g. who is allowed freedom of movement throughout the EU based on their country of origin/citizenship).

In this paper, I will provide a solution to a phenomenon that has been part of an under-developed issue for the social and historical sciences broadly construed (including sociology, demography, political science, epidemiology, and history), and which has deep importance for discussions in contemporary philosophy of race and its connection to the social sciences: the understanding of groups in history. My project is centrally concerned with retrospective description — the usage of contemporary racial terms as labels or classifications for historical phenomena. As formulated, my project seeks to provide an answer to the following question: under what circumstances is it correct to apply racial classifications to historical phenomena? My argument sketches out a solution for both of the categories race, advocating that a successful and

comprehensible way for correctly applying racial descriptions retrospectively to empirical phenomena is to take an instrumental approach — an approach that rejects using both biological realist accounts and social constructionist accounts as the bases for ascertaining whether a contemporary racial or has been correctly applied to past phenomena.

The paper will proceed in the following manner: I will begin by motivating the discussion in the form of providing a number of examples where (purported) racial retrospective descriptions are used. Secondly, I will then survey the literature from historical and philosophical disciplines (broadly construed) in order to highlight previous discussions of retrospective descriptions and their applicability to groups in history. Thirdly, I will assess whether contemporary accounts for the reality of race can be used to ascertain whether retrospective uses of current racial categorisations are appropriate. Fourthly, I will sketch and expound upon my solution for a successful and comprehensible way to correctly apply racial group descriptions retrospectively. Fifthly, I will address potential objections to my sketch and provide responses in turn. Finally, I will provide some concluding remarks that discuss the ways in which my account will provide fruitful benefits for the social sciences, including fields like demography, political science, epidemiology, and sociology — all fields that have been traditionally overlooked in the philosophy of science.

Cristian Larroulet Philippi. "When Values in Science are Not an Obstacle but a Solution"

Abstract: What inferences follow from measurement outcomes? When can we use our measurements to infer, say, that some people are happier on average than others? A widespread methodological tradition answers by pointing at different kinds of scales. Depending on the information they provide, scales are classified as ordinal, interval, or ratio; and which measurement inferences are permitted turn on this classification. In particular, making inferences based on averages taken with ordinal scales (versus interval or ratio scales) aren't permitted. Since few scales in the social sciences are considered quantitative (i.e., interval or ratio), and since causal research in these sciences requires taking averages of the measured outcomes (e.g., among subjects in the treatment and control groups), this methodological tradition substantially restraints the research that social sciences can conduct.

Whether some of the attributes measured today with non-quantitative scales will be measured quantitatively in the future remains an open question. Nevertheless, Michell (2009, 2012) argues that social sciences' attributes just are ordinal (vs. quantitative), so that they cannot afford a quantitative measurement. Accordingly, we should not expect to see in the social sciences what happened with temperature: the development of our measurement practices from ordinal (thermoscopes) to quantitative measurements (thermometers). Michel's reason? In typical social science attributes, different levels of the attribute reflect not only differences in degrees, but also in kind. This means that the intervals (i.e., the difference between levels) are not comparable, because they are not "mutually homogeneous." (2012, 262) He illustrates this with a scale of "functional independence" in the elderly. It classifies people in terms of their capacities, which tend to be lost in order (climb stairs=1, transfer to bathtub=2, bath=3, walk=4, dress upper body=5, etc.). The differences between intervals of this scale are not of the same kind—they are heterogeneous. As he

puts it: “There is no homogeneous stuff, independence, adhering in various amounts to each person” (263). Thus, Michel concludes that functional independence itself is ordinal, and not only measured ordinally.

Here I challenge Michel’s argument. To keep things focused, I restrict my discussion to Michel’s example—functional independence. But my argument generalizes.

Note first that functional independence is a thick concept. Claims about it are “mixed claims” (Alexandrova 2017)—they presuppose a normative standard. The aim of a functional independence scale, then, is not to capture ‘mobility’ in some purely physiological sense. Rather, it aims to capture what is valuable about functional independence. Thus, whether the attribute that the scale aims at measuring is quantitative or not is not decided by whether the physiological capacities used to mark levels in the scale are homogeneous in some physiological sense. Accordingly, Michel’s argument against the homogeneity of these capacities is no argument against the homogeneity of independence.

Moreover, I’ll argue, what is needed for making inferences is not as strict as methodologists have argued. We don’t need a fully quantitative scale, but just enough confidence in the approximate equality of the intervals. Importantly, this confidence must come from normative (prudential) theorizing.

10:30-12:30, March 7 (Sunday), Japan Standard Time

Session 8 “Economic Approaches, Poverty, and Gender” (Chair: Francesco Di Iorio)

Ziming Song. “Folk Psychology and a Logic of Evaluation: On the Normative Content of Decision Theory”

Abstract: Decision theory is a cluster of controversial models of behavior. Some theorists see it as a descriptive-explanatory account that makes more precise such folk psychological platitudes as explaining and anticipating a choice by attributing to the agent’s desires and beliefs. Others take decision theory to offer a normative-evaluative account that defines what it means to make rational choices. One chooses instrumentally rationally when she maintains consistency of the way in which she updates her preference for a new course of action based on her existing preference and information in a choice situation. In this paper, I am interested in clarifying this notion of rationality by investigating the normative content of decision theory.

In his book (Bermudez 2009), Bermudez argues that the prescriptive, descriptive, and normative uses of decision theory come apart, and thus implies the failure of interpreting decision theory as a unified account of the folk psychological platitudes about instrumental rationality. He presupposes that the success of such an interpretation requires that decision theory accounts for all dimensions of instrumental rationality.

In addition, the standard model has faced serious empirical refutation from psychology and behavioral economics including the Allais paradox (Allais 1953, Allais 1979), and evidence that shows a systematic deviation of the actual choices that people make from the standard expected utility model. (Kahneman 1979, Starmer 2000, Wu 2004) Acknowledging that the standard decision

theory fails on its descriptive-explanatory front, I will address the question of whether the theory can be normatively valid without being descriptively true. I will make a so-called "defensive methodological move" to preserve the theory. (Hands 2015)

My argument that standard decision theory captures instrumental rationality and folk psychological platitudes about decision making will meet two main challenges. First, Allais and Guala both worry about the consequences of refuting descriptive decision theory. (Guala 2000, Guala 2006) Apparently, decision theory needs to be psychologically true in order to have any explanatory power. My response is that it should be the case that the expected utility model is without empirical content. Normative and descriptive decision theories are completely different. (Thaler 2018) Once interpreted as conditions of consistency on updating evaluative judgments, such a logic of evaluation is not automatically a descriptive theory.

The second challenge is whether such a normative theory is valid. It centers on a principle that the theory is committed to, a principle often known as invariance. (Or else, separability, the independence of irrelevant alternatives, see for example, Joyce 1999, Bhattacharyya 2011, Dietrich 2016.) I illustrate the problem with Sen's cases (Sen 1993) and locate it with the objects of preference. I argue that the objects of human decisions are much finer-grained than what philosophers and economists normally assume them to be. They should be what I will call "holistic outcomes." Therefore, I argue that interpreted as a logic of evaluation, normative decision theory is both indispensable for accounting for, and making sense of, our intentional actions, and that it is defensible as a valid theory given the holism condition on outcomes.

Norihito Sakamoto and Yuko Mori. "Comparative Analysis of Life Satisfaction, Equivalent Income Indices, and Alkire-Foster Multi-dimensional Poverty Index: Empirical Results from India"

Abstract: Since the publication of the Stiglitz-Sen-Fitoussi Commission report, the dashboard approach, which emphasizes that social welfare should be evaluated not only by income per capita but also various factors such as distributions of income and wealth, social capital, quality of environment, health and happiness, has become a more prominent theme in policy-making and policy-evaluation studies. Although many well-being indices have been proposed, their theoretical properties, operations, and implications remain to be further explored.

The purpose of this paper is to compare and analyze the results of five representative well-being indicators: income, subjective well-being as life satisfaction, health equivalent income, generalized equivalent income, and Alkire-Foster multidimensional poverty index. In order to compare these indices consistently, we collected individual data from lower- and middle-income households in and around the slums of Delhi, India. To measure each health equivalent income, we asked respondents about their marginal willingness to pay for becoming perfect health status for one year. Each generalized equivalent income is calculated by the same estimation method proposed by Decancq et al. (2016). To calculate each person's multidimensional poverty index based on Alkire and Foster (2011), we asked respondents about their income, social capital, subjective and objective health, job condition, educational achievement, and their housing conditions, and so on.

We find that a group of poor persons is quite different among the five well-being indices. For example, although a certain respondents' income level is below the Indian poverty line, his/her numerical value in Alkire-Foster multidimensional poverty index is lower than those of persons with their incomes above the poverty line. We also find that life satisfaction has limited correlation to both income and the multidimensional poverty index. In addition, our results show that health equivalent income is associated with income, though many respondents seem to face difficulty with determining their marginal willingness to pay for becoming the hypothetical perfect health for one year. These results imply that there may be an adaptive preference problem in some well-being measures. Estimating generalized equivalent incomes also face some difficulties with their interpretations because of the estimation based on subjective and hypothetical evaluations among the respondents.

References:

Decancq, K., E. Schokkaert and B. Zuluaga (2016) "Implementing the capability approach with respect for individual valuations: an illustration with Colombian data," discussion paper series of KU LEUVEN, DPS16.09.

Alkire, S., and J. Foster (2011) "Counting and multidimensional poverty measurement," Journal of Public Economics, Vol. 95, pp. 476-487.

Daniel Saunders. "Putting the Cart Behind the Horse in the Cultural Evolution of Gender"

Abstract: In The Origins of Unfairness, Cailin O'Connor outlines a novel theory of the origins of gender. She draws on the tools of evolutionary game theory to show gender might have emerged as a device for solving certain classes of coordination problems. Some tasks are best completed through a specialized division of labour. But without social roles, it can be difficult to coordinate on the issue of who should perform which tasks. Who should fish and who should make pottery? Sexual differences are one salient feature in early human societies that could provide a basis for the division of labour. Once endowed with social significance, sexual difference can transform into the autonomous cultural and normative force of modern systems of gender.

Her models are illuminating but have a difficulty. She assumes that agents engage in gendered social learning as the mechanism by which successful strategies spread through a population. But this seems to put the explanatory cart before the horse – how did early humans have a well-developed system of gendered social learning prior to the gendered division of labour? If we want to explain the origins of gender, we should not help ourselves to gender-like behaviour. She suggests that gendered social learning and the division of labour incrementally co-evolved. But no formal model for how this incremental process is currently available.

This paper closes that explanatory gap. I construct an agent-based model that represents an evolutionary environment faced with coordination problems. This model replicates her core results. However, the agent-based model also provides additional structure to explore more complex social learning behaviours. I show that, under a variety of conditions, gendered social learning and the gendered division of labour can co-evolve. Even in populations where they initially have no

gender-specific preference for social learning, mutants who experiment with gendered learning styles can invade and spread through the population. I also explore conditions in which this result does not obtain. This paper contributes to our understanding of the robustness and potential limitations of cultural evolutionary explanations of gender.

15:00-17:00, March 7 (Sunday), Japan Standard Time

Session 9 "Psychological Perspectives on Social Phenomena" (Chair: Julie Zahle)

Lukas Beck and James Grayot. "New Functionalism and the Social and Behavioral Sciences"

Abstract: Functionalism holds that some states and processes should be individuated based on what role they play rather than what they are constituted of. Functionalists endorse that such functional kinds can legitimately figure into explanations in the special sciences. Mechanists, on the other hand, oppose this and argue that the special sciences achieve their explanatory aims, ultimately, by offering mechanistic decompositions.

While functionalism is still dominant in many special sciences, evolving debates in the philosophy of science indicate problems with traditional arguments for functionalism. Early defenders of functionalism, like Fodor (1974), argued that the existence of well-supported special science laws involving functional kinds vindicates functionalism. The main problem with this defense is that it has become doubtful whether there are any such laws.

Consequently, Weiskopf (2011a; 2011b) has posited a reformulation of functionalism on the model-based approach to explanation common in the special sciences. We refer to this reformulation as new functionalism. Roughly, new functionalism holds that functionally individuated states and processes constitute kinds if they figure into a range of successful models instead of well-supported laws.

However, even under new functionalism, much disagreement remains. In a recent iteration of the debate between functionalists and mechanists, Buckner (2015) introduces a dilemma for Weiskopf's account. According to Weiskopf (2011a), functional kinds can be individuated via three different strategies: fictionalization, reification, or abstraction. Each of these strategies indicates why particular functional kinds are not amenable to mechanistic decomposition. However, Buckner argues that functional kinds are either still amenable to mechanistic decomposition (this holds for abstractions) or models involving functional kinds necessarily incur a loss of counterfactual power (this holds for fictions and reifications). He takes this dilemma to pose a serious challenge to new functionalism.

In this paper, we aim at defending new functionalism and recasting it in the light of the concrete explanatory aims of the special sciences. More specifically, we argue that the assessment of the explanatory legitimacy of functional kinds also needs to consider the explanatory purpose of the model in which the functional kinds are employed. In this respect, we hold that Weiskopf's account neglects the diversity of explanatory purposes found in the special sciences. However, we aim to show that once we take these explanatory purposes into account the horns of Buckner's dilemma will frequently reveal themselves to be dull.

We want to demonstrate this by appealing to model-based explanations from the social and behavioral sciences. Specifically, we make the case that preferences and signals as functional kinds are typically not affected by Buckner's dilemma given the explanatory purposes of the game- and decision-theoretic models in which they are employed. We take our argument to not only deflect Buckner's dilemma but to also expand new functionalism to the social and behavioral sciences.
Section 2 introduces new functionalism and Buckner's dilemma. Section 3 closely analyses preferences and signals to highlight the shortcomings of both Weiskopf's account and Buckner's dilemma. Section 4 outlines the implications of our analysis.

References:
Buckner, C. (2015). Functional kinds: A skeptical look. Synthese, 192(12), 3915-3942.
Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). Synthese, 97-115.
Weiskopf, D. A. (2011a). Models and mechanisms in psychological explanation. Synthese, 183(3), 313.
Weiskopf, D. A. (2011b). The functional unity of special science kinds. The British Journal for the Philosophy of Science, 62(2), 233-258.

Raphaël Künstler. "The Theoretical vs. the Methodological Value of Milgram's Experiments"
Abstract: According to Paul Roth (2003, 2004), the controversy between Daniel Jonah Goldhagen (1992, 1996) and Christopher Browning (1992, 1996, 1998) regarding how to account for the behavior of ordinary perpetrators is relevant for ontology because their opposition between replicate the ontological and methodological opposition between interpretivists and naturalists.

The goal of my paper is to examine the role of Milgram experiment plays in this debate. I claim that Paul Roth's theoretical use of Milgram's experiment fails to undercut Goldhagen's position, for he fails to grasp the central methodological rule of Goldhagen's historiography: any explanative investigation on ordinary perpetrators should start with a counterfactual egocentric thought experiment. This rule is implicit in Goldhagen's book, and, to my knowledge, it has not been articulated by its commentators (Browning, 1998; Moses, 1998; Bartov, 2000; Kamber, 2000; Eley, 2002; Lorenz, 2002; Hinton, 1998; Herzstein, 2002; Zangwill, 2003; Roth, 2004; Pleasant, 2019). I think that there is another use, a methodological use, of Milgram's experiment (and of social psychology in general) which respond to Goldhagen implicit methodology.

In the first part of my paper, I will contend that, from a Goldhagenian point of view, the theoretical use of Milgram's experiment is logically circular.

Its second part makes explicit the central methodological point of Goldhagen's account. The main point is that Goldhagen's moralistic and graphic style — which has annoyed many critics, including Roth — should not be taken as an idiosyncratic feature leading to break the academic rules of savoir vivre. This style should instead be taken seriously, as the application of an implicit methodological rule. This implicit rule is the application of Thomas Laqueur's principle according to which, in order to evaluate whether an account of the ordinary perpetrators' problem is satisfying,

one needs to start by looking at the faces of their victims (Laqueur 1997; Friedländer, 2001). The rationale for such a rule is that, if one does not take into account what was the concrete experience of the murders, it becomes very easy to have the impression to explain their behavior. Before comparing the explanans, one should correctly grasp the explanandum. From this rule, it follows that the historian should start his inquiry with a counterfactual egocentric. He should imagine the situation of the perpetrators, and ask to himself what he would have done in the same situation.

In the third part of my paper, I explain how Milgram's experiment can be used to show that counterfactual egocentric thought experiments are not reliable, lead to self-deception and therefore should not be used by an epistemically responsible social scientist. In order to articulate this direct defense of naturalism, it is necessary to take into account aspects of Milgram's experiment that are overlooked by Roth's account. 1. A part of Milgram's experiment involves 'spectators' trying to foresee the outcome of the experiment (Milgram, 1963). Therefore, Milgram conceived his experiment as showing that people could not predict other people's behavior, based on their own introspective knowledge. 2. Following Kurt Lewin, Milgram took into account the emotions of the participants. An emotional interpretation of his findings undermines Goldhagen's emotional methodology.

To conclude, I will distinguish between an indirect and a direct argument for naturalism. The first based on a theoretical use and the second based on a methodological use of Milgram's experiments.

Petr Špecián. "Fake News and the Victim Narrative: Rationality in the Light of the Debate on Disinformation"

Abstract: Since the Brexit referendum and the U.S. Presidential election of 2016, the potential of online disinformation campaigns to disrupt democratic decision-making has become one of the major topics of the public debate. Recently, the coronavirus pandemic seems to have further contributed to our collective disorientation with regard to identifying reliable sources of information and relevant expert knowledge. In many countries, the resulting epistemic crisis has translated into tremendous losses of both life and prosperity.

Against this background, I will address the clash between instrumental and epistemic rationality that characterizes the digital (dis)information exchange. So far, the most common narrative in the scholarly literature portrays the public as passive victims of disinformation (e.g., Gelfert 2018; Britt et al. 2019). According to this view, the source of the victims' vulnerability lies in their bounded rationality: the public's cognitive biases enable fake news stories to shape beliefs regarding the merits of political programs or vaccine safety.

I will present a skeptical take on this 'Victim Narrative,' whose main fault is the implicit neglect of the frequent clash between epistemic rationality and instrumental rationality. In short, truth is not the only goal people pursue, and often not the main one either. Perhaps especially so in the (dis)information-rich environment of the online platforms. Here, truth-seeking can be easily overridden by the concern for managing one's reputation vis-à-vis various competing social groups.

The idea that instrumental rationality may motivate a certain degree of epistemic

irrationality is not new (Caplan 2007). The supporting evidence is growing, however. It comes from research of politically motivated reasoning (Kahan 2015) or evolutionary-psychological theorizing on the adaptive function of reasoning (Sperber and Mercier 2017; Simler and Hanson 2018). Building on these insights, I extend the model of (dis)information exchange to include a signaling function of the manifested beliefs with respect to their intended audience. I argue this model sheds more light on the current epistemic crisis than its contenders do and provides more measured guidance in the realm of policy applications.

References:

Britt, M. A. et al. 2019. "A Reasoned Approach to Dealing With Fake News." Policy Insights from the Behavioral and Brain Sciences 6 (1): 94–101. https://doi.org/10.1177/2372732218814855.

Caplan, Bryan. 2008. The Myth of the Rational Voter: Why Democracies Choose Bad Policies. New edition. Princeton, N.J.; Woodstock: Princeton University Press.

Gelfert, A. 2018. „Fake News: A Definition". Informal Logic 38 (1): 84–117. https://doi.org/10.22329/il.v38i1.5068.

Kahan, D. M. 2015. „The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It". In Emerging Trends in the Social and Behavioral Sciences, eds. R. A. Scott and S. M. Kosslyn, 1–16. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118900772.etrds0417.

Simler, Kevin, a Robin Hanson. 2018. The Elephant in the Brain: Hidden Motives in Everyday Life. New York: Oxford University Press.

Sperber, Dan, a Hugo Mercier. 2017. The Enigma of Reason: A New Theory of Human Understanding. London: Allen Lane.

17:10-19:10, March 7 (Sunday), Japan Standard Time

Session 10 "Symposium: Evidential Pluralism and Causality in the Social Sciences) (Organizer: Yafeng Shan; Chair: Chor-yung Cheung)

Yafeng Shan. "Symposium: Evidential Pluralism and Causality in the Social Sciences"

Symposium Abstract: Causal claims abound in the social sciences. However, there is no consensus about the assessment of causal claims, nor how to understand causality, among social scientists and philosophers of social science. One view, Evidential Pluralism, maintains that in order to establish a causal claim one normally needs to establish the existence of a correlation and the existence of a mechanism, so when assessing a causal claim one ought to scrutinise both association studies and mechanistic studies. This thesis has led to fruitful philosophical work on the role of mechanisms in the biomedical sciences and to suggestions for improvements to evidence-based medicine ('EBM+'). The question arises as to whether it can also be applied to other contexts. The aim of this symposium is to examine the scope of Evidential Pluralism and the concept of causality in the social sciences.

Shan and Williamson will defend the application of Evidential Pluralism in the social

sciences. They argue that there are several benefits to applying Evidential Pluralism to the social sciences. Runhardt will argue for an epistemic account of Evidential Pluralism for case study methods in political science. Maziarz will challenge the application of Evidential Pluralism in economics. He will argue that causal pluralism (i.e. pluralism of concepts of causality instead of types of evidence) can be supported by delivering case studies, where researchers draw causal conclusions from only one type of evidence. Taylor will revisit the debate on mechanistic explanation and causal explanation in cognitive science and argue for an epistemic approach to causal explanation in cognitive science, according to which both mechanistic and non-mechanistic explanations of cognition can be genuine causal explanations.

Abstracts of the Contributed Papers

Yafeng Shan and Jon Williamson. "Applying Evidential Pluralism to the Social Sciences"

Abstract: Since around the year 2000, philosophers of science have produced a great deal of interesting research on the role of mechanisms in science. One strand of this research concerns the role of mechanistic evidence in establishing causal claims. Russo and Williamson (2007) argued that in the biomedical sciences, a causal claim is established by establishing (i) that the putative cause and effect are correlated, and (ii) that there exists a mechanism linking the two which can account for this correlation. This thesis has the following important consequence: while quantitative studies (in particular, randomised controlled studies) provide excellent evidence of correlation and, in the right circumstances, can provide evidence of the existence of a mechanism, it is important to also consider other evidence of mechanisms when assessing a causal claim. This motivates a kind of Evidential Pluralism.

In medicine, this form of Evidential Pluralism has led to a proposed modification to evidence-based medicine, called EBM+. Parkkinen et al. (2018), for instance, developed procedures for evaluating mechanistic studies alongside clinical and epidemiological studies, when assessing the effectiveness of an intervention or when ascertaining the effects of exposure to an agent.

This paper argues that Evidential Pluralism applies equally to the social sciences. In the social sciences, as in the biomedical sciences, establishing causation requires establishing both correlation and mechanism---social mechanisms, in this case. While quantitative association studies can provide some evidence of mechanisms, in addition to good evidence of correlation, other sorts of study also provide good evidence of social mechanisms---notably, certain qualitative studies.

We argue that there is scope to apply Evidential Pluralism to the social sciences. First we show that the lessons from evidence-based medicine can be carried over to evidence-based policy, and that Evidential Pluralism can provide an account of the assessment of evidence in evidence-based policy. We compare this account to that provided by realist evaluation, which also has a central role for mechanisms. Second, we use case studies to argue that Evidential Pluralism additionally applies to more theoretical social sciences research, and can be used to elucidate the confirmation relations in basic social sciences research. Third, we show that Evidential Pluralism can provide new foundations for mixed methods research, because it offers a precise account of the need for mixed methods when establishing causation in the social sciences.

We then respond to two objections to the claim that Evidential Pluralism can be applied to the social sciences: one due to Julian Reiss and a second due to Francois Claveau. We conclude that Evidential Pluralism has much wider scope than originally envisaged, and sheds new light on the use of evidence in the social sciences.

Rosa W. Runhardt. "Evidential Pluralism and Epistemic Reliability"

Abstract: Proponents of Evidential Pluralism argue one ought to combine evidence of different sorts of causal theories (like probabilistic, mechanistic, or regularity causation) in order to "bear on a causal hypothesis and strengthen it" (Reiss 2009, 27). In other words, evidential pluralists suggests that evidence of different sorts of 'things' (Illari 2011), such as correlations, entities and activities, or interventions, can corroborate a causal hypothesis. In political science, evidential pluralists have mainly studied how to fruitfully combine qualitative and quantitative (large-N statistical) methods in multimethod research (Crasnow 2010). How one might adapt Evidential Pluralism for single case studies in political science is rarely explored, even though within case study research there is a plurality of methods and methodologies. In this paper, I expand upon this literature, defending what I will call an epistemic account of Evidential Pluralism for case study methods in political science.

In the first part of the paper, I outline different methods within case study research, using Haggard and Kaufman's *Dictators and Democrats* (2016) research as my primary example. Rather than look at the practical side of combining methods, I focus on what these methods' epistemic assumptions are, and whether they are compatible. For this, I develop a framework in terms of positive and negative reliability of these assumptions. These terms, which stem from theoretical epistemology (Nozick 1981), indicate respectively how likely it is that a social scientist judges a causal claim to be true if indeed the claim is true, given their epistemic assumptions, and how likely it is that a social scientist judges a causal claim to be false if it is false, given their epistemic assumptions.

In the second part of the paper, I show that the reliability of the epistemic assumptions in case study methods crucially depends on the evidential context. This context includes amongst others the heterogeneity of the population being studied and how much information is already collected of other causal relations in the case. I argue that only methods making assumptions with high reliability in the evidential context of the case under study can be fruitfully combined. However, I also show that in certain case study contexts, only one set of assumptions (and thus only one type of method) will have high reliability. Thus, my epistemic account implies that an uncritical combining of methods is not to be recommended. I finish the paper with recommendations for how we may systematically measure both positive and negative reliability of epistemic assumptions.

Mariusz Maziarz. "Evidential Pluralism and Causal Pluralism: Argumentative Strategies and Policy Implications"

Abstract: In the context of causal inference in the sciences, pluralism comes in different flavours. The repertoire of social scientists ranges from qualitative interviews and case studies to laboratory and field experiments and includes both qualitative and quantitative, observational, and

interventional designs. This methodological pluralism raises the philosophical problem of how to make sense of these diverse research practices. In particular, the question arises of whether these diverse research practices deliver different types of evidence for causality, which is a single kind of relation, or, instead, different types of evidence imply different kinds of causal relations (or different concepts of causality). My talk approaches this question.

The two main responses are known as, respectively, Evidential Pluralism (pluralism of types of evidence) and causal pluralism (pluralism of concepts of causality). Evidential Pluralism encapsulates to the view that causal claims need support from both difference-making and mechanistic evidence. Therefore, the philosophers that support Evidential Pluralism as a view on causality descriptively adequate to research practices in a discipline, deliver examples of causal claims established on the basis of the two types of evidence. In contrast, causal pluralism (i.e., pluralism of concepts of causality instead of types of evidence) can be supported by delivering case studies, where researchers draw causal conclusions from only one type of evidence (e.g., mechanistic or difference-making). Drawing on examples from economics, I show that the two argumentative strategies deliver conflicting results: in some cases, economists draw causal conclusions from one type of causal evidence while other causal claims are established in a way that agrees with Evidential Pluralism.

This can turn the interest of philosophers to the question of which of the stances is superior as a normative position. The normative message of Evidential Pluralism that establishing causality requires the two types of evidence trades off the number of potentially spurious causal claims based on either difference-making or mechanistic evidence for their robustness. However, it may set the standard too high for the social sciences since causal mechanisms may not produce observable regularities or some difference-making evidence may lack mechanistic explanation. Evidential pluralism assumes that one type of causal claims (supported by both difference-making and mechanistic evidence) is needed for all types of policy-making. However, it is not yet established that, from the perspective of the user of causal knowledge, all purposes require causal claims in agreement with the epistemic theory of causality.

Instead, causal pluralists point out that different types of policy-making may be based on alternative types of evidence. I elaborate on this view and distinguish among three types of economic policy-making: ‘policy actions’ (i.e., policy-making that does not change the relata of a causal claim), ‘institutional reforms’ (i.e., creating a mechanism in a target), and ‘interventions’ (interventions on a cause to influence its effect). I argue that difference-making evidence from observational studies and mechanistic evidence is sufficient for, respectively, policy actions and institutional reforms, but ‘interventions’ require difference-making evidence from experimental or quasi-experimental research designs.

Samuel D. Taylor. “Causation and Cognition: An Epistemic Approach”

Abstract: According to some philosophers of cognitive science, only mechanistic explanations of cognition count as genuine, causal explanations of cognition, because only evidence of mechanisms reveals the causal structure of cognition (Kaplan and Craver 2011; Piccinini and Craver 2011). On

this view, a causal explanation of cognition will carry explanatory force only to the extent that it identifies, through analysis, the component parts of the mind/brain (e.g. neurons, modules, etc.) and their principles of interaction, before showing how these component parts causally interact to generate some phenomena (Machamer, Darden, and Craver 2000). For example, a causal explanation of categorisation will carry explanatory force only to the extent that it reveals the causally interacting parts and activities that are responsible for our capacity to discriminate between encountered objects.

But this view is not without its critics. In fact, some philosophers now argue that we can have genuine causal explanations of cognition that abstract away from mechanistic detail to characterise the causal structure of cognitive systems in nonmechanistic terms. For example, systems that are "dynamically and interactively bound up with the causal structure of the world on multiple spatial and temporal scales" (Hutto and Myin 2017, 58). On this competing view, we cannot rule out the possibility of developing a genuine, causal explanation of, say, categorisation that makes no reference to the mechanistic parts and activities of the mind/brain. The upshot is that we cannot jump to the conclusion that causal explanations of cognition that fail to cite mechanistic information are *ipso facto* spurious or defective.

There thus exists a tension between two different views of causal-explanatory project in cognitive science. On the one hand, some philosophers take the view that only *causal-mechanistic explanations* count as causal explanations cognitive science. On the other hand, some philosophers take the view that *causal-nonmechanistic explanations* can also count as causal explanations in cognitive science. My aim in this paper is to develop a unified account of causal explanation in cognitive science. To this end, I defend an epistemic conception of causal explanation in cognitive science, which I claim deflates the aforementioned tension by asserting that both causal-mechanistic and causal-nonmechanistic explanations of cognition can be genuinely explanatory.

I first set out the difference between causal-mechanistic and causal-nonmechanistic explanations of cognition. Then, I argue that the debate about causal-nonmechanistic explanations of cognition is grounded in conflicting intuitions about the mechanistic account of causality. I go on to argue that we cannot endorse the mechanistic account of causality without begging the question against causal-nonmechanistic explanations of cognition. Then, I introduce the epistemic account of causality (ETC) and argue that ETC does not prejudge the status of causal-nonmechanistic explanations of cognition. Finally, I argue that by endorsing ETC we can arrive at a unified conception of causal explanation in cognitive science.

References:

Crasnow, Sharon. 2010. "Evidence for Use: Causal Pluralism and the Role of Case Studies in Political Science Research." *Philosophy of the Social Sciences* 41 (1): 26–49.

Haggard, Stephan, and Robert R. Kaufman. 2016. *Dictators and Democrats: Masses, Elites, and Regime Change*. Princeton, NJ: Princeton University Press.

Hutto, Daniel D., and Erik Myin. 2017. *Evolving Enactivism*. Cambridge, MA: The MIT Press.

Illari, Phyllis McKay. 2011. "Mechanistic Evidence: Disambiguating the Russo–Williamson Thesis."

International Studies in the Philosophy of Science 25 (2): 139–57.

Kaplan, David Michael, and Carl F. Craver. 2011. "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective." *Philosophy of Science,* 78 (4): 601–27.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Parkkinen, Veli-Pekka, Christian Wallmann, Michael Wilde, Brendan Clarke, Phyllis McKay Illari, Michael P. Kelly, Charles Norell, Federica Russo, Beth Shaw, and Jon Williamson. 2018. *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*. Cham: Springer.

Piccinini, Gualtiero, and Carl F. Craver. 2011. "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches." *Synthese* 183 (3): 283–311.

Reiss, Julian. 2009. "Causation in the Social Sciences: Evidence, Inference, and Purposes." *Philosophy of the Social Sciences* 39 (1): 20–40.

Russo, Federica, and Jon Williamson. 2007. "Interpreting Causality in the Health Sciences." *International Studies in the Philosophy of Science* 21 (2): 157–70.