



Description of the workshop

Evidence plays a crucial role in science. With the development of evidence-based decision-making, nowadays also in public decisions and policy-making, the trend is towards using evidence-based methods to assess policy-making in sectors like health, education, or the economy. Choices on school programs may be made on the basis of evidence from randomized control trials, as well as choices on how to invest in development programs in third-world countries. Researchers use various methods for gathering evidence, from formal ones like models, controlled experiments, or simulations, to less formal ones like historical analysis, or case studies. In most cases expert judgment is one of the most significant components of the process of gathering and amalgamating evidence.

Experts thus enter the evidence process at all levels: from the choice of methods, to the gathering of the evidence itself, to the amalgamation of evidence coming from different sources. Often, experts are themselves sources of evidence, through their experience and personal knowledge; for instance in the case of economists as members of monetary policy committees, or in the teams of scientists working on the IPCC reports on climate change.

Philosophers of science, methodologists, and epistemologists have long discussed the concept of evidence; this workshop will focus on whether the current concepts and accounts of evidence are adequate for capturing the subjective, reason-based and argumentative component of evidence-based science and policy-making.



Finnish Cultural
Foundation



Trinity Project

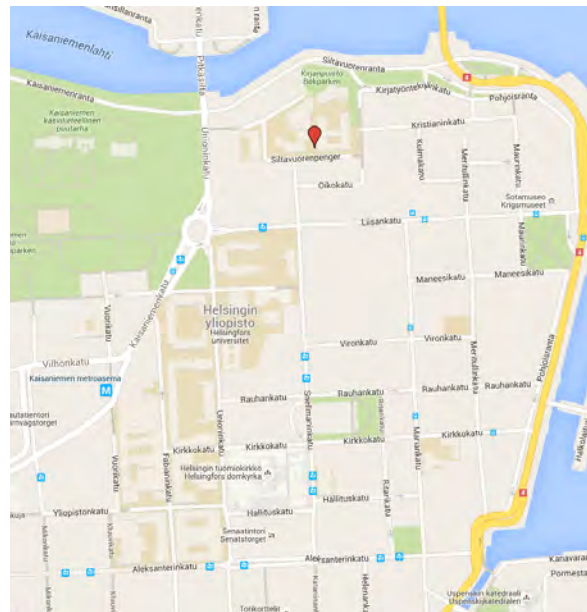
The Trinity of Policy-Making: Evidence, Causation and Argumentation

The workshop is the first in a series of meetings within the project The Trinity of Policy-Making: Evidence, Causation and Argumentation. More information of the project can be found at the following websites:

martinicarlo.net/research/

evidencecausationargumentation.wordpress.com

The workshop is kindly supported by the Finnish Cultural Foundation with a two-year Workshop Grant.



Evidence and Expertise

October 29-30, 2015. Athena Building, Room 360. Siltavuorenpenger 3A, 00170 Helsinki



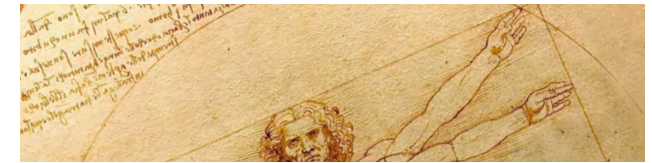
Program

Friday, October 30, 2015. Athena Building, room 360. Siltavuorenpenger 3A, 00170 Helsinki

Thursday October 29, 2015. Athena Building, room 360. Siltavuorenpenger 3A, 00170 Helsinki

- 9.45-10:00 Opening
- 10:00-11:00 Evidential Pluralism in the Causal Sciences – **Rani Lill Anjum** (Norwegian University of Life Sciences) and **Stephen Mumford** (University of Nottingham)
- 11:15-12:15 Experts: Who Needs Them – **Eleonora Montuschi** (Ca' Foscari University of Venice and London School of Economics)
- 12:15-13:45 LUNCH BREAK
- 13:45-14:45 Confirmation Meets Social Epistemology: A Theory of Inferential Judgement – **Julian Reiss** (Centre for Humanities Engaging Science and Society, Durham University)
- 15:00-16:00 Here, There, Everywhere: External Validity of Randomized Experiments in Development Economics. **Judith Favereau** (TINT, University of Helsinki) and **Carlo Martini** (TINT, University of Helsinki)
- 16:15-17:15 Experts in Fabula – **Barbara Osimani** (Munich Center for Mathematical Philosophy, LMU)

- 9.00-10:00 Using Expert Judgment to Measure Forecast Uncertainties for Policy-Makers – **Roel Visser** (Erasmus Institute for Philosophy and Economics, Erasmus University Rotterdam)
- 10:15-11:15 The Target of Policy-Relevance in Think Tank Research – **Anita Välikangas** (TINT, University of Helsinki)
- 11:30-12:30 “The Economists Enter the Courtroom”: Evidence and Expertise in the Recent History of Forensic Economics on Discriminations – **Cléo Chassonnery-Zaïgouche** (Centre d'études interdisciplinaires Walras Pareto, University of Lausanne)
- 12:30-14:00 LUNCH BREAK
- 14:00-15:00 Expertise in EBM: A Tale of Three Models – **Sarah Wieten** (Centre for Humanities Engaging Science and Society, University of Durham)
- 15:15-16:15 Expert Judgement in Climate Services – **Casey Helgeson** (London School of Economics), **Erica Thompson** (London School of Economics), **Roman Frigg** (London School of Economics)



Trinity Project

Running from 2015 to 2017, TRINITY brings together scientists, philosophers of science, argumentation scholars, and policy-makers to explore the notion of evidence as the three-place relation: 'evidence is evidence of something for someone'. This is done by considering evidence in a wider philosophical context, including causal reasoning and methodology. The organizers are Carlo Martini (Helsinki), Rani Lill Anjum (NMBU, Aas), and Frank Zenker (Lund/Konstanz).



Finnish Cultural Foundation

Support

Generously supported by the Finnish Cultural Foundation, TRINITY cooperates with the CauseHealth project (Rani Lil Anjum), the project group on Causal Argumentation (Frank Zenker), and the ArgLab at the New University of Lisbon (Fabrizio Macagno). The project's official home is the [Academy of Finland Center of Excellence in the Philosophy of the Social Sciences \(TINT\)](#) at the University of Helsinki.



Description

Evidence plays a crucial role in science. With the development of evidence-based decision-making, nowadays also in public decisions and policy-making, the trend is towards using evidence-based methods to assess policy-making in sectors like health, education, or the economy. Choices on school programs may be made on the basis of evidence from randomized control trials, as well as choices on how to invest in development programs in third-world countries. Researchers use various methods for gathering evidence, from formal ones like models, controlled experiments, or simulations, to less formal ones like historical analysis, or case studies. In most cases expert judgment is one of the most significant components of the process of gathering and amalgamating evidence.

Experts thus enter the evidence process at all levels: from the choice of methods, to the gathering of the evidence itself, to the amalgamation of evidence coming from different sources. Often, experts are themselves sources of evidence, through their experience and personal knowledge; for instance in the case of economists as members of monetary policy committees, or in the teams of scientists working on the IPCC reports on climate change.

Philosophers of science, methodologists, and epistemologists have long discussed the concept of evidence; this workshop will focus on whether the current concepts and accounts of evidence are adequate for capturing the subjective, reason-based and argumentative component of evidence-based science and policy-making.

Abstracts



Anita Välikangas

The Target of Policy-Relevance in Think Tank Research. The last three decades have witnessed a significant increase both in the amount of think tanks and the quantity of research conducted by them. These think tanks attempt to influence politics and their surrounding society. From the viewpoint of research it means that they are usually oriented towards generating policy-relevant research, much more than traditional research communities.

This paper examines how the connection to policy-making is present in think tank research. In previous research, the work conducted by think tanks has sometimes been regarded as knowledge-transfer processes. This stance highlights that think tanks are institutions that translate ideas generated in scientific research to policy domain. Recently, these views have been present for instance in Research-Integration-Utilization (RIU) model (Böcher & Krott 2012). In this paper, I attempt to demonstrate why this three-fold distinction, where think tanks serve as integrators between research and knowledge utilization, is not able to accurately describe the way how think tank research is constructed.

In order gain an understanding of the typical types and characteristics of think tank research, this paper conducts a small survey of think tank research. This empirical part examines research published by eleven think tanks during a two-month period (June–July 2015). Total number of examined reports is 67. By identifying common types of this research, the paper demonstrates that policy practices and policy-orientation are and can be manifest in many different ways in think tank research.

This consideration of policy-relevance of think tank research contains some implications also from the viewpoint of understanding scientific research and evidence in the context of policy-making. Most of think tank research – perhaps apart from medical research – is quite local or otherwise specific in its nature. It can thus be asked how much this task of generating policy-relevant research comes at the cost of placing less weight to traditional scientific norms, such as universality, disinterestedness and originality.



Barbara Osimani

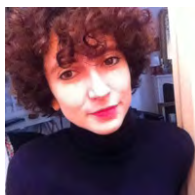
In the debate over Evidence Based Medicine (EBM), expert judgment, which is ranked at the lowest level in the hierarchies, is considered by critics of such paradigm as not dispensable, not only for the essential role played by clinical reasoning in the medical setting, but also for its importance in assessing the overall meaning of a body of evidence for the target population (see e.g. Cartwright 2007). Expert judgment should, among other, provide subject-specific knowledge which supports inferences from the limited scope of study results to their applicability to target populations. However, what if this judgment itself is “biased” by the very *forma mentis* that underpins the EBM methodology? This means that the flaws attributed to the EBM paradigm globally, just replicate in expert judgment locally, only to result in increased opacity. Indeed this phenomenon can be observed at different levels: in cases of dissent, where experts disagreeing on substantive matters struggle in articulating intuitions which lie outside the methodological standards developed within the EBM framework; more systematically and more obviously, it is also present in the recommendations and guidelines for evidence selection in systematic reviews.

The present talk addresses the problems related to this practice with a special focus on causal assessment for harm, and illustrates a proposal for evidence amalgamation based on Bayesian inference networks (Osimani, Pöllinger, Landes; forthcoming). While this approach does not exclude either the incorporation of expert judgment, or the use of “higher level evidence”, it has epistemological, methodological, as well as heuristic virtues. From an epistemological point of view, the graph hierarchically structures the inferential problem and distinguishes between diverse inferential levels. This allows us to gain greater insight into methodological problems by specifying the distinctive epistemic roles of their components. Furthermore, it also helps to discern various issues on the philosophical dissent around EBM, by showing that criticisms may address different nodes in the causal inference, although they regard the same study type. From a heuristic point of view, such approach may help scientists articulate their intuitions about the informative role of various pieces of heterogeneous evidence and to model the structure of the problem in a unitary framework.



Casey Helgeson

Expert Judgement in Climate Services. Climate change adaptation is largely a local matter, and adaptation planning can benefit from local climate change projections. Providing these projections (and/or scenarios) as well as accompanying decision support is the work of climate services, currently provided by a combination of state agencies, private consultancies, and universities. (Major efforts are underway to expand and develop climate services, both in Europe — through the European Joint Programming Initiative JPI Climate and the environment theme of the European Commission’s Research and Innovation framework, and globally — through the Climate Services Partnership and the World Meteorological Organization’s Global Framework for Climate Services). Local climate change projections available from climate services providers are typically generated by accepting climate model outputs in a relatively uncritical way. We argue, based on the treatment of model outputs by the Intergovernmental Panel on Climate Change (IPCC), that this approach is unwarranted and that subjective expert judgment should play a central role in the provision of local climate change projections intended to support decision-making. — Work with **Erica Thompson and Roman Frigg.**



Cléo Chassonnery-Zaïgouche

“The Economists Enter the Courtroom”: Evidence and Expertise in the Recent History of Forensic Economics on Discriminations. Discrimination has been an ongoing concern in economics, as an object of theoretical enquiries as well as expertise measurements. In the early 1970s, empirical methods — micro-econometric estimations and, later, lab and field experiments — expanded the scope of the economic theories of discrimination, focusing on gender and racial issues, from labor to other market (credit, housing) and non-market areas (education, health). Results were largely used to assess the evolution of discrimination in the long run, in the context of policy

evaluation, such as the Civil Rights Act of 1964. But since the early 1980s, the economics of discrimination made its way to the Courtroom. In a famous paper, Ashenfelter and Oaxaca (1987) stated the superiority of the economic definition of discrimination over the legal one, especially in terms of implementation. The objective of the paper is to recast this point of view and to analyze the influence of the economics of discrimination in the courtroom in practical terms. From the intelligibility question of the more technical scientific procedures to the definition of science in the rules of evidence, the paper explores the underlying epistemological issues concerning the credibility of evidences: what are the conditions of the transposition of evidences from academic to legal arena? Two case studies are developed: the use of wage regression analysis as proof of indirect wage discrimination and the use of field experiment (audit in-person) as a proof of direct discrimination.



Eleonora Montuschi

Experts: Who Needs Them? Experts are often listed among the ‘sources of evidence’ — as for example in the well known evidence ranking schemes — so expert opinion gets compared to methods such as RCTs, or cohort studies, etc. However, at least if we stick to how the quality of evidence is assessed by these schemes, experts score badly in this capacity: they always come last in the list.

In this talk I will argue that the reason (one of the reasons) why experts often appear as lesser or disqualified sources of evidence is that it is somehow wrong to treat them as such. Experts in the domain of use and practice are not primarily providers of evidence, they are rather processors of evidence. Expertise has to do not only with the amount and quality of knowledge produced in a certain field but also, and more specifically, with the skill(s) of making that knowledge matter, or count towards a question, prediction, or course of action. I will explore what this skill amounts to by analyzing the process of making so-called ‘judgement calls’. I will approach the topic from a particular context of use (a policy case) where expertise crucially bears on evidence.



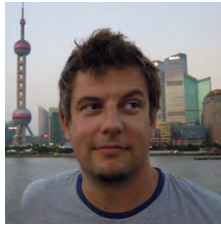
Judith Favereau and Carlo Martini

Here, There, Everywhere: Policy Validity of Randomized Experiments in Development Economics. In the last few decades the use of randomized field experiments (RFEs) has increased dramatically and is still increasing. The internal validity of such experiments is undoubtedly their most attractive feature; however, their external validity is still problematic. RFEs are considered to have a strong internal validity, allowing them to produce what is

labelled as evidence, but strong internal validity may be coupled with very weak external validity (see Cartwright 2007, 2009, 2010, and Cartwright & Hardie, 2012).

There are a number of issues related to external validity: one is the problem of “global external validity” — i.e., how do we generalize from the test-population to the whole population. Another issue is “local external validity” — i.e., how to infer, from the success of an experiment in a test-population, to some conclusions about a different but to equivalent (for the purposes of the test) population. Both issues are discussed extensively in the literature, and RFEs are problematic under both aspects. In this paper, however, we claim that the subdivision does not capture all the possible ways in RFEs can go wrong. We postulate and explain the concept of policy validity.

Policy validity is “validity for the purposes of policy making”. Policy validity may still fail when both global and local external validity are carefully checked for. We present a case study about the use of fertilizers: The study assesses the impact of a development program (SAFI) on the use of fertilizers by Kenyan farmers. We conclude that policy validity cannot be managed through the standard techniques that are sometimes employed for other types of validity, but requires subjective judgment.



Julian Reiss

Confirmation Meets Social Epistemology: A Theory of Inferential Judgement.

There is no universal logic of induction (cf. Norton 2003). Inductive arguments can be more or less compelling or ‘cogent’ but the extent to which they are so depends on contextual factors that go well beyond what could be described as an inductive logic. William Rehg (in his book *Cogent Science in Context*) focuses in particular on three extra-logical dimensions of scientific argumentation: the dialectical, the rhetorical, and the socio-political. Dialectical elements have to do with the way in which arguments are being produced — for instance with whether or not an argument has had a sufficient chance of being challenged by a sufficiently inclusive choice of experts. Rhetorical elements are features of the presenter of the argument (e.g., ‘Is she an honest scientist?’) and its audience (e.g., ‘Does it mainly speak to other scientists or the public at large?’). Socio-political elements are the institutional realities within which an argument is made (scarce resources may mandate a smaller degree of inclusivity than would ideally be desirable, for instance).

This paper looks at one logic of induction in particular — eliminativism — and, starting from accepting that it is indeed the case that logic and ‘the facts’ alone indeed underdetermine the conclusion of an inductive argument, develops a richer ‘theory of inferential judgement’ that includes dialectical, rhetorical, and socio-political elements in Rehg’s sense. The theory is then applied to a case from contemporary social science that illustrates the importance of the extra-logical elements in making a good social-scientific argument.



Rani Lill Anjum and Stephen Mumford

Evidential Pluralism in the Causal Sciences. What type of

evidence should we consider? And what counts as evidence? Evidence is often used as an objective notion, something that is generated by scientific activity. But different types of scientific methods generate different types of evidence: correlation data, difference-makers, probability-raising,

mechanisms, etc. In the causal sciences, evidence should be evidence of causation. We adopt a position of methodological pluralism in which we attempt to identify causation scientifically by looking for evidence of its symptoms. These symptoms will be plural and our methods are based on finding evidence of them. Not all of the symptoms appear in every case of causation and no one method will suffice for every case of causation. Ideally, we would find a convergence, where plural methods make the same judgement on whether A causes B. We look at what would justify the adoption of various methods including statistical data, trials and expert opinions. The latter form of evidence may be of a special nature or higher order: for instance, where it is the taking of an overview of all the other evidence.



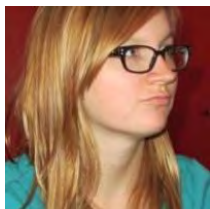
Roel Visser

Using expert judgment to measure forecast uncertainties for policy-makers. To

measure and reduce forecast uncertainty, both the econometric and measurement literature advocate the use of expert judgment in addition to statistical measures of uncertainty (Franses, 2014; Boumans, 2015). However, the authority of forecasting institutes critically depends on their reputation as provider of objective policy-advice (Morgan & den Butter, 2000). How can forecasting institutes use expert judgment to improve their policy-advice, without jeopardizing their relation with policy-makers? To answer this question I investigate how the role of the forecasts of the CPB Netherlands Bureau for Economic Policy Analysis in the budgetary policy-making of the Dutch government is affected by the use of expert judgment to measure forecast uncertainty. This research is based on an internship at the CPB and 12 interviews with economists and policy-makers.

Policy-makers assign little credibility to measures of uncertainty that are directly based on expert judgment. The CPB has therefore considered different strategies to make its expert judgment 'more objective' (cf. Martini & Boumans, 2014) by incorporating it only indirectly in its uncertainty measurement, through the use of probabilistic models (cf. Sims, 2008) and formal elicitation methods (cf. Cooke, 1991). Both of these attempts have been unsuccessful, so far. The problem with the currently available probabilistic models is that they are insufficiently realistic for detailed policy-analyses. The problem with formal elicitation methods is that the CPB knows little about the sources of its expert knowledge, uses a forecasting process that leaves little room for variety among expert judgments, and is unsure what data to use to calibrate the expert judgments that are used to measure uncertainty.

Due to these constraints, the CPB currently restricts the use of expert judgment to qualitative scenario analyses, while probabilistic measures of uncertainty are based on historical forecast errors. I suggest two solutions to improve on this uneasy compromise. Firstly, to develop a non-ideal account of expert judgment, that helps forecasters and policy-makers to evaluate the use of probabilistic models and formal elicitation methods in research contexts that are not fully compatible with the requirements of such models and methods. This may also help to evaluate alternatives such as the ad-hoc combination of scenarios and historical uncertainty measures. Secondly, my interviews reveal that the expertise of forecasters often originates with policy-makers, whereas the literature typically locates expert judgment only in between of science and policy. Acknowledging this may improve our understanding of expert judgments and reduce the suspicions of policy-makers.



Sarah Wieten

Expertise in EBM: a tale of three models. In this talk I will argue that there have historically been three main models in EBM (Evidence-Based Medicine) for the integration of expertise. Differences between these models and a lack of clarity regarding if the older models should be abandoned and replaced by the new, or used in concert, mean that the EBM conception of expertise remains impoverished. There is, however, a wealth of theoretical resources available to enrich EBM's account of expertise. These theoretical resources come from many disciplines including philosophy, sociology, artificial intelligence, computer science and decision theory. These resources currently cannot be deployed because of the conflicting signals within EBM about the role of expertise. Other roles for expert judgement are important and they deserve more discussion within the EBM literature. Until this discussion happens within EBM, these resources cannot be deployed. In the first portion of the talk I will discuss EBM's standard "GOBSAT" joke and what it means about the conversation regarding expertise in EBM. Next, I will explain the three historical models of expertise integration in EBM and discuss the possible difficulties in putting them into practice. I will explain the ways in which all three accounts leave something to be desired, and lastly suggest that, with some additional clarification from EBM proper, theoretical resources from other disciplines might augment the current EBM account of expertise.