Workshop

Reasoning about Evidence: Logical, Historical and Philosophical Perspectives

Book of abstracts

Date: 4-6 November 2019
Venue: Royal Academy of Dutch Language and Literature (KANTL)

Organization:

Centre for Logic and Philosophy of Science

(Ghent University)

Department of Science and Technology Studies
(University College London)

Chairs

Phyllis Illari (London), Erman Sozudogru (London), Erik Weber (Ghent).

Local Organizing Committee

Thijs De Coninck, Stef Frijters, Kristian González Barman, Karina Makhnev, Erik Weber & Dietlinde Wouters.

About the workshop

This workshop is organized within the series "Logic, Reasoning, and Rationality":
`http://www.lrr.ugent.be/`

The series is supported by the Research Foundation Flanders (`http://www.fwo.be/`) through the scientific research network on Logical and Methodological Analysis of Scientific Reasoning Processes (LMASRP):
`http://www.lmasrp.ugent.be/`

Abstracts

Invited speakers

Federica Russo

An informational approach to evidence

Scientific claims are based on evidence. This is hardly contested. What is evidence, instead, is far more contentious. In philosophy of science, one approach to evidence is to analyse the formal and probabilistic relations between evidence (E) and hypotheses (H). However, such an approach remains largely silent about these E and H amounts to. Another approach, evidential pluralism, puts forward an epistemological and methodological thesis according to which evidence of correlation and of mechanisms is needed in order to establish causal claims. This approach is very specific about to the object of evidence needed (of correlation, of mechanisms) and about the purpose (establishing a causal claim). Is it possible to provide a more general and widely applicable account of what evidence is? In this talk, I explore the prospects of an informational approach to evidence and I sketch the consequences this might have on concepts such as model validation.

François Claveau

The Epistemic Risks of Diversifying Evidence: A Bayesian Perspective

Science feeds on evidence like countries feed on natural resources. Science today has its own resource curse: evidence is abundant and varied to the point of impeding the intersubjective convergence on general conclusions. Methodologists and philosophers of science offer remedies to this curse. For proponents of systematic reviews, the aggregation procedure must have strict inclusion criteria for what counts as admissible evidence. For many philosophers of science, the opposite direction is to be privileged: we should include all the relevant evidence or, at least, aim to diversify rather than to restrict evidence.

This paper explores the epistemic risks of this diversification with a Bayesian model of scientific inference. We translate in a formal framework three worries about diversification that are typically expressed informally: (1) disagreement on the relevance and reliability of some types of evidence, (2) discordant evidence, and (3) confirmation bias. We chart the extent to which, in the confines of our model, these worries are warranted. The paper also sheds light on recent propositions to use Bayesian networks to help in the aggregation of widely diverse evidence.

Jutta Schickore

Peculiar blue spots: causes, circumstances, and evidence around 1800

My paper examines how investigators in the late 18th and first half of the 19th century dealt with evidence and causes. I focus on a strange phenomenon that was of interest to researchers in different fields of study, including pathology (veterinary and human), organic and agricultural chemistry, natural history, and public hygiene: the colored spots that sometimes appeared on fresh milk. The episode illustrates continuities and shifts in the investigators' approaches to evidence and causes. Between the 1770s and 1840s, there was a notable change in the characterization of what was in evidence. There was also a significant change in ideas about who can offer credible evidence. Moreover, the emphasis shifted from establishing efficacy of treatment to establishing causes and mechanisms. Nevertheless, despite all these changes, broader ideas about how reliably to establish efficacy of treatment, causes and mechanisms remained relatively stable.

Contributed Talks

Phyllis Illari

Why do we need evidence of mechanisms?

I will present a view of evidence of mechanisms as evidence of the activities, entities, their organization, and the phenomena they explain, using the idea of 'minimal mechanism' (Glennan and Illari, 2018). I will argue that this view allows us to theoretically organize an incredibly diverse array of forms of evidence and empirical practices. I will situate it within a further form of evidential pluralism by showing how this form of evidence can be complementary to evidence of difference-making, following work by (Clarke, Gillies, Illari, Russo and Williamson, 2014). For medical evidence, evidence of difference-making typically takes the form of evidence of correlation or association in a population. The crucial idea of integrating evidence of mechanism and evidence of correlation is that evidence of mechanism helps address the major weakness of evidence of correlation, i.e. the problem of confounding, or the possibility that C and E are in fact common effects of a third variable, D. In reverse, if you are unsure whether the effect of the mechanism you have identified might be 'masked' by the effects of unidentified mechanisms also linking C and E – the major weakness of evidence of mechanism – seeking evidence of a correlation in a population is what you need. This means evidence of both correlation and of linking mechanism is complementary in an important way.

I will then home in on a specific way in which evidence of mechanism is crucial, arguing that it is important even for solid evidence of correlation. Any clinical study, even a well-conducted RCT, which is still one of our best methods of establishing a reliable correlation, needs decent answers to two questions: (i) what are the variables for disease, treatment and outcome? and (ii) how and when are they measured and why?

I will use the case of 'vitamin D deficiency' to show how important these questions are, even when they are not explicitly addressed in published results, because they are regarded as sufficiently standardised to be unimportant. Until recently, vitamin D deficiency was regarded as well-understood, reliably measurable in standardised ways, and linked to diseases such as rickets by well understood mechanisms. However, recent research has linked vitamin D deficiency to other diseases, in ways that expose the fact that different measuring techniques measure slightly different forms of Vitamin D. Those differences are now relevant.

Considering the case shows how deeply integrative our evidential pluralism needs to be, and therefore how complex our practices of reasoning about evidence are. Philosophical accounts need to be responsive to this.

Jaakko Kuorikoski

Mechanistic evidence and a new argument from inductive risk

I present a novel account of mechanistic evidence and show how the contrastive nature of such evidence leads to a thus far unacknowledged way in which non-epistemic values necessarily enter into evaluation of mechanistic models and theories – a new argument from inductive risk.

First, following Carl Craver (2007, 247-255), information about a feature of the investigated phenomenon is evidentially relevant to a theory about that phenomenon, if it constrains the set of possible mechanisms potentially realizing the phenomenon. As evidence constrains the possibility space, it discriminates between alternate possible mechanism hypotheses. Evidence for mechanisms is therefore inherently contrastive. As Lindley Darden (2006) and Craver (2007) have argued, potential mechanisms can be characterized according to their possible component parts and their activities, the spatial and temporal characteristics of the activities of these components. If an established feature of the investigated phenomenon is incompatible with some aspects of these general, then it constrains the space of possible mechanisms. Darden and Craver discuss research heuristics in biology and cognitive neuroscience,

but similar general constraints can be defined and identified also in the social sciences in terms of characteristics of possible social mechanisms (Hedström & Ylikoski 2010).

Second, not all findings about features of phenomena (component parts, activities etc.) constrain the space of possible mechanisms equally efficiently. Findings about some features of the mechanism are therefore stronger evidence than findings about other possible features in that they have a bigger impact on the possibility space. The contrastive nature of mechanistic evidence means that a piece of evidence has an effect not only on a particular mechanism hypothesis H1, but on the whole probability distribution over the alternative hypotheses. A natural way of analyzing the incremental impact of new evidence on a set of alternative hypotheses is in terms of uncertainty or 'entropy' reduction (Crupi & Tentori 2014; Niiniluoto & Tuomela 1973, 66-68; Oaksford & Chater 1994). Entropy is a measure of the 'evenness' of the distribution, and hence of epistemic uncertainty, over the set of the alternative hypotheses. The strength of the acquired evidence can now be defined as a measure of how much more 'uneven' the posterior distribution becomes.

The third step in the argument is to note that there is, arguably, no unique single measure of uncertainty/entropy (Beck 2009; Crupi & Tentori 2014). Consequently, there is no single unique measure of uncertainty reduction. I argue that the 'right' measure of entropy reduction, and hence evidential strength, depends on the pragmatic context, namely what the information (mechanistic hypothesis) is needed for. Hence, the projected end-use of the information is intimately tied with the 'choice' of the measure of information gain and, consequently, of evidential strength. This does not mean that the strength of mechanistic evidence would be arbitrary or subjective, since the different measures of entropy reduction can be theoretically subsumed and analyzed within a common formal framework. A hypothetical example of the mechanisms, and consequent effective treatments, of problem gambling is used as an illustrative case.

Saul Perez Gonzalez

Is evidence of mechanisms indispensable for extrapolation?

In the last two decades, the relevance of mechanisms in science has been underlined by several authors. Initially, this relevance was mainly associated with explaining scientific phenomena and supporting causal claims (see e.g. Glennan 1996; Machamer et al. 2000). Nevertheless, nowadays mechanisms are considered important for many other issues. Particularly, mechanisms are claimed to be relevant for extrapolating causal claims (Steel 2008; Clarke et al. 2013; 2014; Parkkinen et al. 2018; Gillies 2019; Marchionni and Reijula 2019), since it is considered that statistical evidence faces outstanding difficulties for supporting the extrapolation of a causal claim from a study population to another population of interest. Some authors argue that evidence of mechanisms is necessary to overcome those difficulties (Clarke et al. 2013; 2014; Parkkinen et al. 2018) so that reliable "study-to-target" extrapolations should take into account both statistical evidence and evidence of mechanisms (ES and EM, respectively). This idea is often known as the indispensability thesis. The aim of this talk is to examine and evaluate it.

The main problem concerning extrapolation from study population to target population, which are indeed different populations, is how to justify a similarity claim between the former and the latter, that they are similar in all the relevant respects (i.e.:they are not dissimilar in any relevant respect). ES has been claimed to face relevant difficulties for dealing with that issue. Two well-known arguments usually invoked to stress the intrinsic limitations for standard statistical methodology concerning external validation are: (i) extrapolating from statistic frequencies in the study population assumes a "biological universal response" (Victora et al. 2004; Clarke et al. 2014), and (ii) differences between experimental ("artificial") and "real life" conditions may be crucial. However, even if there are significant limitations for standard statistical methods based on ES, we do not think that these difficulties can be avoided just by resorting to EM, nor that EM is indispensable for overcoming them.

We will argue that (i) does not point at an intrinsic limitation for ES. Moreover, examples like those about antihypertensive treatments (see Clarke et al. 2014) are not sufficient to defend the indispensability of EM as testing evidence. Rather, it could play at most a subsidiary role to define the sample space for testing subsequent statistical hypotheses. Concerning (ii), it is not clear that extrapolations based on EM are in a better position than those based on ES. In the end, it must be assumed not only that the same mechanism is operating in both populations, but also, that it operates in the same way even though conditions may be rather different. Contextual variations may be so harmful for extrapolations based on EM as for those based on ES (for examples see Howick et al. 2013; on the unpredictability of mechanisms see DesAutels 2011).

However, even though EM is not indispensable for reliable study-to-target extrapolation, EM may be a valuable resource. In order to qualify this claim, we distinguish between a positive and a negative role. On the positive side, if the relevant mechanisms at work (and factors that influence them) in the study and the target populations are highly similar in the relevant aspects, the extrapolation of the causal claim is more justified. On the negative side, if the relevant mechanisms at work (or factors that influence them) in the study and the target populations differ in relevant aspects, the extrapolation of the causal claim is not justified. According to the aforementioned considerations, we will defend that the negative role of EM is actually relevant, while the relevance of the positive role is dubious. No proper method for acquiring the knowledge required for it seems to be available (Reiss 2010), and, unlike the negative side, it is severely affected by the limitations of the mechanisms approach to extrapolation (Howick 2011a; Howick et al. 2013; van Eersel et al. 2019).

Luis Mireles-Flores

What 'policy' in evidence-based policy

The recent empirical turn in economics (see Angrist and Pischke 2010) has become an essential backbone of a general approach to using scientific evidence for policy purposes, the so-called evidence-based policy (EBP) movement. The main idea motivating EBP is that empirical sciences should devote more effort to improving and systematising their evidence-evaluating methods and standards to secure producing scientific research that is more reliable to guide policy. Not surprisingly, EBP has rapidly grown in popularity, for it makes economists feel their research is highly policy relevant, and at the same time it makes policy makers feel their decisions are more scientifically grounded.

There have been, however, a number of criticisms raised against the EBP movement. For example, questions about the epistemic priority of randomised controlled trials (Worrall 2007; Cartwright 2010), and criticisms about how the EBP methods are not good at providing information about the mechanisms underlying the causal relations (Weber 2007; Marchionni 2017; Steel 2013; Grüne-Yanoff 2016). Furthermore, Nancy Cartwright (e.g., Cartwright 2009; Cartwright and Stegenga 2011; Cartwright and Hardie 2012) has put forward a general account to conceptually characterise the evidential requirements for successful and effective evidence-based policy.

In this article, I claim that existing philosophical accounts have mainly focused on the "science" side, so to speak, of the science-policy interaction, and mostly ignored the details of the "policy" side, which are essential for understanding how scientific evidence can be relevant to policy. I argue that philosophers should not only be analysing and questioning the methodological problems related to EBP's inferential techniques and dubious hierarchies of evidence, but they should also open the "policy" black-box in "evidence-based policy", and start studying policy-making not as a simple outcome variable in a formal causal framework, but as a complex process with distinct stages, and different dimensions, aims, and causal factors interacting dynamically at each stage.

Finally, I put forward an alternative approach to the policy relevance of economics and EBP. Instead of exclusively relying on the well-known philosophical accounts of evidence, causal inference, extrapolation, and the like, I

will develop a framework for philosophical analysis structured in line with current accounts of the policy-making process in public and social policy research.

Two of the main advantages that I expect to obtain from this interdisciplinary move are the following: First, in policy studies policy making is understood as a complex dynamic process, and studied in terms of a number of clearly characterised stages (Birkland 2016; Dunn 2016; Hill and Varone 2016; Kraft and Furlong 2017). Second, there is a plethora of discussions and substantial characterisations of many of the contextual potential disturbing factors that can affect the outcomes of science-based policy making (Head 2013; Jasanoff 2013; Cairney 2016).

A promising consequence of my approach is that by taking seriously and understanding policy-making as a complex process, one can analyse and assess separately and more precisely how different types of scientific knowledge and evidential techniques play different roles at each distinct stage of the policy process. Thus, instead of general philosophical debates about whether economic science is policy relevant or not, or about which particular type of evidence or inferential method is better or worse for "policy", the discussion could now be refocused towards more informative questions, for instance, about how different pieces of scientific knowledge can be more or less relevant to the specific aims and needs at the different stages of the policy process.

Maria Jimenez Buedo

Background knowledge and experimental evidence in the social sciences

The experimental revolution in the Social sciences is one of the most significant methodological shifts undergone by the field since the turn of the century, having effects in the possibility of cross-collaboration of formerly separate research areas and above all, changing the way social scientists view and deal with problems of causal identification. One of the often valued features of social science experimentation is, precisely, the fact that there are clear methodological rules regarding hypothesis testing that allow for the adjudication among contentious causal claims. The paper tries to spell out the conditions under which this kind of role is possible and underlines the often crucial though often ignored role of background knowledge in mediating between experimental results and the inferences that can be drawn from them.

For reasons that are partly contingent, as I discuss in the paper, a fundamental component of the conceptual set of tools that we use to describe social scientific experimental results and practices in includes the distinction between internal and external validity, as first conceived by Campbell (1957) and Cook and Campbell (1979). I develop the idea that one undesirable consequences of the extensive use of the distinction between internal and external validity is that the terms, and the conceptual and methodological approach in which they are embedded (the Campbellian project) tends to assume that there is a correspondence between experiments and the inferences that can be made from them.

This alleged correspondence between experiments and their inferences has, in turn, as a consequence, the underplaying of the role of background knowledge in inferring causal statements from experimental data, flattening the role of experiments to that of objective or impartial trials that can conclusively adjudicate among contentious causal claims. While this characterization may suit certain aspects of concrete experimental research programs, it does not represent the role of experimentation in much of the laboratory practices in growing subfields such as theoretical experimental sociology or behavioral economics.

Bert Leuridan & Sydney Green

How much is enough? On the role of the Principle of Total Evidence in evidence amalgamation in practice

In many scientific disciplines, an overwhelming amount of – often discordant – evidence for causal claims exists. As a remedy, several ways of amalgamating evidence exist. Recently, several authors have invoked Carnap's Principle of Total Evidence to assess current practices of evidence amalgamation. Based on the work of Carnap, Ayer, Good, and Graves, we argue that the PTE as such is not rational and that the relevance, cost, reliability, and availability of evidence are crucial in deciding what our evidential base should be. We also show that these issues provide a useful, multi-dimensional tool for understanding the nuanced, local, and contextualized versions of the PTE which are used in current practice. The guidelines of IARC, the Cochrane Collaboration, and Borenstein et al. (2009) are used as case studies.

Elena Rocca & Frederik Andersen

Scientific evidence evaluation and experts' disagreement: how to make it work

Scientific safety assessment of technologies and human interventions is often underdetermined by evidence. In these cases, it would be advantageous to build an approach to evidence evaluation that takes advantage, instead of suffering, from diverging experts' interpretations of the same evidence. One example is Douglas' explanatory approach (Douglas 2012), which works as an inference to the best explanation. In this approach, a plurality of different possible explanations of the same evidence are first collected, and then evaluated according to the criteria of internal consistency, empirical competency, and predictive potential. These criteria, however, are not always stringent; it is often the case that two or more explanations apparently meet all of them, and are still equally scientifically defendable. Here, we propose one more constraint to what counts as the 'best explanation', and we call it 'the criterion of unity of background ontological assumptions'. The criterion is based on the premise that a number of different explanations of the same evidence may be motivated by diverging background assumptions about the nature of things. Accordingly, the criterion demands consistency between the background ontological assumptions of a specific explanation, and the background ontological assumptions of current general scientific knowledge in the field. We motivate and explain both the premise and the criterion, by recalling Galilean-Keplerian arguments relating to evidential underdetermination (Galilei 1615, Galilei 1632, Kepler 1600, Kepler 1609). We are going to use the case of underdetermination of theories of planetary motion for two reasons. The first is that it is a classic example of empirical equivalents: alternative theories that are equally well supported by any possible evidence, which makes our point easier to show. The second is that underdetermination was solved through selection of the most defensible ontological basic assumption, which then played the role of tie breaker. In other words, we want to offer a clear example where ontology picked up where epistemology dropped off, and show how and why such a strategy can succeed.

Our reference to the Galilean-Keplerian method should not be seen as anachronistic and out of place, though. Indeed, we bring it back to the context of evidence-based policy by showing how it would apply to the case of expert disagreement about the evidence of safety of stacked genetically modified plants.

Mauricio Suarez

The contextual character of causal evidence

I argue that causal evidence is contextual, i.e. it is relative to a context. More specifically, I defend the view that the same causal claim may be warranted by the same piece of empirical evidence in one context but not another. This view must be distinguished from the superficially similar view that causal claims themselves are contextual, or relative to context. My claim is about the context-relativity of evidence, not about the relativity of truth. I do not argue that what is a true causal claim in one context may be a false claim in another. On the contrary, I do not believe that the concept of truth in general is relative (i.e. I do not believe that the truth conditions of declarative statements are generally relative to context – although I do of course accept that there are truths about particular contexts). I also find that causal claims are amongst the least relative, uncertain, or indeterminate truths that we possess. I provide some examples and argue that there is no interesting sense in which these are socially constructed truths, or truths only for particular communities, practices or cultures.

I will consequently assume here that our knowledge of causal truths is as absolute and context-independent as any other knowledge we possess, because we can only possess knowledge of what is in fact true in the non-relative or absolute sense described above. Yet, although my views about truth in general — and causal truth in particular – are conservative, my views about evidence — in particular causal evidence — are not. I do believe that evidence for causal truths is contextual in an interestingly radical sense. Maybe this is also a sense in which any evidence for any claim may be said to be contextual; if so, it is still the case that the contextual nature of causal evidence has not been appreciated sufficiently so far.

My most general claim is that causal evidence is contextual however causal evidence is understood. In other words, I defend that the same causal claim may be warranted by the same piece of empirical evidence in one context but not another – and that this is so independently of the theory of causation, and causal evidence, that is adopted. There will always be important contextual presuppositions determining both what I call default entitlements and relevant confounding factors. In this talk I only have enough space to argue for the claim in connection with the manipulability theory of causation developed by Jim Woodward (2003). However, at the end of the talk I go on to outline an extension of this argument to another theory, namely the counterfactual theory. This suggests that the context determines the objective standards required for evidence for a particular causal claim regardless of how causal evidence is understood.

Brice Bantegnie

"Constraint", constraints, and evidence in the philosophy of cognitive science

This talk has both a therapeutic and a prophylactic goal. I argue against the use of the concept of constraint in the context of discussions of intertheoretical relationships in philosophy of cognitive science. In order to do so, I show that the word "constraint" is multiply ambiguous and I argue that this ambiguity muddles debates on the relationship between psychological theories and neuroscientific theories. The talk has three parts.

In the first part, I analyze the meaning of the word "constraint" as it is used in philosophy of cognitive science. I distinguish four distinct concepts of constraint which go by pairs. First, there are two evidential concepts of constraint. According to the first one, psychologists believe that they ought to appeal to mechanistic explanations to support their functional analyses. According to the second one, psychologists ought to appeal to mechanistic explanations to support their functional analyses. Second, there are two instrumental concepts of constraint. According to the first one, mechanistic explanations are used to narrow down the space of possible empirically equivalent functional analyses to one. According to the second one, mechanistic explanations can be used to narrow down the space of possible empirically equivalent functional analyses to one.

In the second part, I show that the contemporary debate on the relationship between psychological theories and neuroscientific theories has been muddle by the confusion between instrumental concepts and evidential concepts. In a nutshell, those who have argued against the autonomy of psychology from neuroscience have used an instrumental concept of constraint which couldn't deliver the normativity they needed in support of their argument.

In the third and last part of my talk, I draw the genealogy of the "constraint" talk in the philosophy of cognitive science. I contend that it comes from a focus on the metaphysics of the mind/brain relationship. I argue that a metaphysical viewpoint is not the right viewpoint to adopt in order to understand the relationship between psychological theories and neuroscientific theories. First, the metaphysical issue of the relationship between the mind and the brain is an issue of folk metaphysics. Second, evidential questions should be dealt with first before answering the metaphysical question.

Yin Chung Au

Evidence in biological basic research: what they are and how they become representations

This study uses concrete examples of the practice of searching molecular and cell biological mechanisms to argue that experimental data serve as either evidence or representations according to the contexts they are used for finding explanations. Initially, I define evidence in biological basic research as something that supports this process: researchers conduct experiments to approach the certainty that the mechanism of the phenomenon of interest contains only specific known components and that the components are only connected by specific known causal relations.

I discuss the evidential status of biological data by adopting and extending Russo and Williamson's (2007) framework, which I consider as useful for understanding both health sciences (i.e. their case study) and biological basic research. Based on their classification of mechanistic and probabilistic evidence, I raise some points regarding 'probabilistic evidence'. Upon having obtained new data, biological researchers determine its evidential validity. Here, the data is used to prove that an experimental intervention is effective for producing a hypothesised outcome. In this context, both quantity and diversity of evidence are required. The former is necessary for statistically testing difference-making between the effect and the postulated cause. The latter is to offer probabilistic independence between evidence kinds so that the researchers can build robust causal conclusions. In short, experimental data in biological basic research not only statistically prove the probabilistic dependence between cause and effect but also serve as probabilistically independent evidence because of their diversity in rationales and materials.

Meanwhile, in technical expressions such as papers and lab meetings, practitioners sometimes refer to some data as 'evidence' in another context. The relevant data in this context tend to appear in groups, and with obvious diversity, to stand for specific components of the developing mechanistic explanations. I argue that while the practitioners might not necessarily be aware, they use such diverse experimental data collectively as representations, which normally contain at least one causal relation, for their surrogative reasoning regarding the mechanistic explanations of interest. In this context, difference-making between cause and effect has been confirmed. Researchers no longer treat data as mere evidence of effective interventions but indicators of specific components of mechanisms. This is consistent with the practitioners' commonplace phrase 'evidence that indicates a mechanism'. To clarify, it is not that I am concerned about the rhetorical triviality but that the connection between words and practices, in my view, can be philosophically rich. Researchers conceptually organise and manipulate these indicators. Thereby they reason about the mechanisms responsible for particular phenomena and design future experiments for revealing more possible indicators/representations.

Lastly, I do not consider the evidential and representational statuses to be mutually exclusive. It is because the researchers know what the evidence supports that they know how to turn a group of evidence into representations for surrogative reasoning.

Femke Truyens

Validity of data as precondition for evidence. A practical analysis of what is taken to count as evidence in psychotherapy research

The Evidence-Based paradigm in mental health care emphasizes the use of the best available methods to provide a sound evidence-base for clinical practice. However, there is strikingly little consensus on what evidence is. Nonetheless, (mental) health researchers conduct a vast amount of research in which the outcome is taken as evidence. To derive outcome, quantitative data are collected from samples of patient-participants. In the generation of 'evidence' on therapeutic efficacy, 'the data' thus play a vital role.

In 'gold standard' psychotherapy research, 'the data' are commonly collected by validated self-report questionnaires. This way, quantitative data are collected that are aggregated to become a data set, which is subsequently used as input for the inference of evidence on treatment efficacy. Commonly, researchers take the validation of measures as a warrant for the soundness of collected data. In this talk, I discuss concrete research data from three patient-participants who participated in the Ghent Psychotherapy Study (Meganck et al., 2017) to illustrate how the practical process of data collection results in 'the data'. Based on our empirical analyses of the data collection process, however, I show how the 'the data' can yield validity issues despite (or because of) them being collected by validated measures.

When these data are straightforwardly aggregated to form the data set that is used as input for deriving evidence, the validity issues that I have shown are easily overlooked, and as the statistical principles of aggregation and distribution are not sufficient to cover these validity issues, they can become inherent to the data set. Consequently, the validity of evidence can be threatened by validity of raw data despite (or because of) the use of validated measures. Therefore, I argue that validity of data is a precondition for sound evidence.

Steven Tresker

The Russo–Williamson Thesis and medical treatment

The Russo–Williamson Thesis (RWT) states that "In order to establish a causal claim in medicine one normally needs to establish two things: first, that the putative cause and effect are appropriately correlated; second, that there is some mechanism which explains instances of the putative effect in terms of the putative cause and which can account for this correlation" (Russo and Williamson, 2007). One reading of this thesis takes the claim to be about what makes for good evidence in medicine. Medical treatment is arguably one of the most important aspects of medicine, yet little has been written on how the RWT applies to medical treatment; i.e. is it useful to practicing clinicians or research scientists? For example, can it suggest research programs, establish effectiveness from studies already performed, and be conducive to making successful treatment recommendations? My purpose is to evaluate this. I use two examples: a non-pharmacologic medical treatment (psychotherapy) and a pharmacologic treatment (empagliflozin, an SGTL2 inhibitor). Drugs are the most extensively tested medical treatments, yet non-pharmacologic treatments (e.g. the expressive arts therapies, such as dance, music, and art therapy; physical therapy; exercise; medical devices; psychotherapy; nutrition therapy; surgery) are also an important part of medicine. Whether and how evaluation of evidence for these treatments' effectiveness should differ is thus important to assess and is the avenue I take for exploring the applicability of the RWT to medical treatment.

My approach applies the RWT to these two types of medical treatments by evaluating them through the lens of what can be seen as drivers of RWT's epistemological framework: Austin Bradford Hill's guidelines (also called indicators of causality), a focus on mechanisms, and a deference to community standards for establishing effectiveness claims. Much of the evidence regarding my case studies' putative effectiveness is based on a randomized controlled trial (RCT)/evidence-based-medicine epistemic framework. Given the limitations with this approach (Cartwright

and Stegenga, 2011; Gupta, 2007; Reiss, 2015), alternative methods of evidence generation/evaluation (e.g. investigating underlying mechanisms) may be better suited to medical treatments, particularly non-pharmacologic ones. However, idiosyncratic features of the medical treatments I explore (e.g. multiple "communities", non-amenability to RCTs, and inapplicability of some of Hill's indicators, such as dose-response relationships and specificity) reveal important limitations of RWT's drivers. I conclude that what constitutes sufficient mechanistic and difference-making evidence for establishing the effectiveness of a medical treatment is domain-specific and depends on the goals, stakes, and features of the treatment in question. To improve the RWT's usefulness for evaluating the effectiveness of medical treatments, its drivers should be operationalized in a way sensitive to the realities of medical practice and specific treatments.

Maarten Kleinhans

What on earth are we doing... 500 million years of evidence for causes of river meandering and still the puzzle is incomplete??

Past failures of causal explanation accounts leads some authors to a resignation to pluralist accounts. I will attempt to show how a theory for a ubiquitous earth-scientific phenomenon was based on integrated reasoning about various kinds of evidence, used in difference-making and in production.

Rivers, such as the prehistoric Scheldt river close to Ghent, do not simply follow the steepest descend from mountains to the sea; rather, they may have braided, multiple channels, or a winding, meandering channel flanked by an erosion-resistant floodplain with clay and vegetation. The dynamic pattern of meandering is associated to vegetation. The relevant literature of several earth-scientific disciplines reports on a wealth of kinds of evidence and kinds of inference, on physically-based mechanisms, numerical simulation and analogue scale experiments with intervention in the main variables with the aim of an integrated theory that causally explains and predicts the existence, properties, occurrence through geologic time. Indeed, earth science seeks both causal explanations and plausible 'historic' descriptions of the Earth's development.

Various disciplines have studied present-day active meandering systems through observation and interference in real rivers, in numerical models and in analogue experiments. Some emphasized general causal factors (difference-making), which led to hypotheses of contrasting physical mechanisms. Others emphasized inference and testing of specific physical and physico-biological mechanisms (production).

On the other hand, geologic reconstruction of past systems quantified spatiotemporal correlations between signs for meandering and braiding and signs, or absence thereof, for vegetation, to infer a relation with timing and evolution of land plants. The evidence and the inference are insensitive to the precise invoked physical mechanisms and are 'only' implicitly assumed ("causation, whichever mechanism it invokes, implies correlation"). Even though the geologic record is very poorly preserved, the positive association of meandering with vegetation appears overdetermined by the evidence, while the choice between alternative mechanisms for meandering is underdetermined by the available geologic evidence. However, the precise mechanisms in these theories and simulations differ considerably, and indeed there exist meandering rivers without vegetation, including a fossil case on planet Mars that shows all the hallmarks of dynamic meandering, where vegetation likely did not evolve but another factor had the same effect on river bank erodibility.

The theory for meandering appears to consist of mechanisms, processes with physical, sedimentological and biological variables, and specifications of INUS conditions. How these are precisely integrated into a theory is not yet clear to me, but collectively these explain occurrence and variations in characteristics between cases on present-day Earth and, in fossil form, on planet Mars and on Earth since land plants evolved. There are also failed and successful predictions about other kinds of meandering channels, such as meanders on glaciers, in estuaries and on the ocean floor.

Maarten Van Dyck

The temporal constitution of evidence

This talk will use Galileo's experimental confirmation of his law of fall as an exemplar through which to investigate the relation between experimental data and the laws that they confirm. It will be shown that Galileo's grounds for trusting his law of fall were not just constituted by the relation between the mathematical ratio's encoded in the law and the numerical values established by his experimental trials. The experiment was designed to answer a precise question that arose in the context of a wider research program investigating the relation between motion on inclined planes and pendular motion. It is this position in an independently but empirically motivated research program that convinced Galileo of the possibility to establish an exact mathematical law. Its proposed form yielded the further opportunity to reinterpret an earlier experimentally established phenomenon concerning the trajectory of projectiles, thus integrating the latter in the framework of the same research program. It will thus transpire that confirmation is not to be characterized as a two-place relation between data and law but as a three-place relation between data, law and the progress of further research predicated on the acceptance of the law. Evidence is temporally constituted in ongoing research rather than by purely formal relations.

Louis-Etienne Villeneuve

How to put the right thing in the right box? Evidences about the past

For whom studies the past as a causal chain of transmission of information, following E. Sober (1988) and A. Tucker (2004) conceptual frameworks, the evidences are easy to identify: every remain left by the past is, in principle, an evidence. Still, the usefulness of theses frameworks are controversial in the eyes of many contemporary philosophers of history such as J.-M. Kuukanen (2016) or P. Roth (2019). Recently, P. Roth asked me after a talk on Tucker's framework what is the consistency of information and how an historian can distinguish, in a causal chain of information, what is an event and what is an evidence resulting from it, or in other words, "how to put the right thing in the right box".

This question is crucial if you want to preserve A. Tucker's framework. But I think the question also shows a misunderstanding of the research goals and procedures attributed to history by Tucker. I will propose in this presentation that the critic of differentiation between event and evidence makes sense only if you subscribe implicitly to the thesis that history is a discipline trying to justify sentences or narratives about events by the use of evidences. On the contrary, in Tucker's framework, history is a discipline trying to explain the very existence of the remains. Description of events, for Tucker, are not the things to be proven, they are the hypotheses proposed to explain the traces of what truly happened. As Tucker puts it, when it comes to the evidences, description of the past and explanation of the evidences are the same. The difference between events and evidences is then very easy to make: evidences are what is left by the past, materials, to be found everywhere but mostly in the archives. Events are hypotheses to explain them, using information theories.

Christian Hennig

Can and should statistical model assumptions be tested?

Methods of statistical inference such as tests and Bayesian procedures are regularly used to quantify evidence. These methods rely on statistical model assumptions. I focus on frequentist inference here, but Bayesian inference has similar issues. Statistical folklore says that in order to make sure that model-based inference is valid, the model assumptions need to be fulfilled, which can be tested ("misspecification testing"). There are a number of problems with this. Firstly, if model assumptions are tested, inference is applied conditionally on passing the misspecification test, whereas the theory on which the inference is based is unconditional. In other words, model assumptions are actively violated by testing them ("goodness-of-fit" or "misspecification paradox", Hennig 2007). Secondly, "combined procedures" can be defined in which a model-based method of inference is applied if a misspecification test passes the model, and another method otherwise. Literature investigating the performance of such combined procedures is surprisingly critical of this approach. Thirdly, the problem of misspecification testing may be seen as ill-posed, given that no model can ever be precisely true, and that passing of the model assumption by a misspecification test does not imply that the model assumption holds. I will give an overview of existing arguments and results. I will argue that the problem of misspecification testing is usually misinterpreted, and that its aim cannot be to make sure that the model assumption really holds, but rather that model violations are ruled out that would lead to misleading inference. Most existing model misspecification tests are not tailored to this aim. There are situations and setups in which preliminary misspecification testing is helpful for subsequent inference, and other situations in which this is not the case. This depends on the specific characteristics of the model-based method, the alternative method, and the misspecification test. I will give conditions under which misspecification testing is advantageous, but I will also highlight limits of this approach. This is based on joint work with Iqbal Shamsudheen (Shamsudheen and Hennig 2019).