ZiF Zentrum für interdisziplinäre Forschung
Center for Interdisciplinary Research
Universität Bielefeld

ZiF WORKSHOP

MARCH 14 - 16, 2013

DIMENSIONS OF MEASUREMENT

CONVENORS

Alfred Nordmann | TU Darmstadt | GER

Oliver Schlaudt | U Heidelberg | GER

PLENARY SPEAKERS

Marcel Boumans | U Amsterdam | NED

Hasok Chang | U Cambridge | GBR

Nadine de Courtenay | U Paris 7 | FRA

Michael Heidelberger | U Tübingen | GER

Martin Kusch | U Wien | AUT

Luca Mari | U Cattaneo | ITA

Joel Michell | U Sydney | AUS

Mary Morgan | LSE London | GBR

Simon Schaffer | U Cambridge | GBR

Eran Tal | U Bielefeld | GER

Laura Dassow Walls | U Notre Dame | USA

For further questions please contact **Marina Hoffmann** | Tel +49 521 106-2768 | marina.hoffmann@uni-bielefeld.de

For further information see www.bicoda.info

The conference is supported by the Deutsche Forschungsgemeinschaft.

Plakatgestaltung: Stefan Adamek

Program Committee

Emily Brock, *History, University of South Carolina*
Andreas Gelhard, *Philosophy, Technische Universität Darmstadt*
Ann Johnson, *History and Philosophy, University of South Carolina*
Andreas Kaminski, *Philosophy, Technische Universität Darmstadt*
Johannes Lenhard, *Philosophy, Universität Bielefeld*
Leah McClimans, *Philosophy, University of South Carolina*
Alfred Nordmann, *Philosophy, Technische Universität Darmstadt*
Carsten Reinhardt, *History, Universität Bielefeld*
Oliver Schlaudt, *Philosophy, Universität Heidelberg*
Astrid Schwarz, *Philosophy, TU Darmstadt and Universität Basel*
Michael Stöltzner, *Philosophy, University of South Carolina*

> Measurement has long been considered a hallmark of science properly practiced, and once a new discipline has developed a mathematical discourse, it has almost immediately laid claim, at least in the language of its most enthusiastic disciples, to the significant status – science!

With these words, Harry Woolf opened a special issue of *Isis* that was dedicated to the "History of Quantification in the Sciences." Based on a conference that took place on November 20-21, 1959 at the Social Science Research Council, the collection of papers included Thomas Kuhn's "The Function of Measurement in Modern Physical Science" alongside other disciplinary perspectives – Kuhn's was the physics paper that was followed by contributions about measurement in chemistry, medical science, psychology, economics, sociology, and last, but not least, biology.

Dimensions of Measurement promises an equally comprehensive, yet somewhat more eclectic consideration of the history and philosophy of measurement in the sciences. It respects not only the particulars of all the various special sciences, but also the variety of philosophical approaches to questions of measurement, as well as the entire scope of "science in practice" which ranges from research practice to the ways in which science answers to demands of practice. During the next few days, we will be watching each other to discover commonalities and differences in the questions we ask – do our agreements and disagreements mark a particular moment in the history and philosophy of the sciences as did the 1959 conference on quantification?

With curious anticipation, therefore, we welcome you to Bielefeld and a conference that arose from interactions with the BiCoDa Alliance (www.bicoda.info). This alliance of historians and philosophers from Bielefeld, Columbia (University of South Carolina), and Darmstadt considers science in its technological setting: How do science-technology interactions shape research-practice today and in the past?

As with any conference of this size, many thanks are due for material and moral support, administrative and financial assistance – to the BiCoDa program committee, to Martin Carrier, to Stefan Adamick, Jutta Braun, Daniela Brinkmann, Marina Hoffmann, and Daniel Schindler, to the ZiF and the Deutsche Forschungsgemeinschaft.

Bielefeld – March 14, 2013 Alfred Nordmann and Oliver Schlaudt

Thursday, 14 March • 10:00-10:45

Alfred Nordmann & Oliver Schlaudt: *Welcome and Outlook*

Thursday, 14 March • 11:00-13:00

The Epistemic Value of Visual Data *(p. 7)*

Chair: Aud Sissel Hoel Room: **Plenary Hall**

Patrick Maynard: "*Photo Mensura"*
Laura Perini: *Data images, representation, and the creation of scale*
Nicola Mößner: *Visual Data – Reasons to Rely on?*
Tobias Schöttler: *Pictorial Evidence: On the Rightness of Pictures*

Sociology of Metrological Knowledge *(p.10)*

Chair: Hans Radder Room: **Round Table**

Hector Vera: *A Single Language of Measurement for Humanity: The Global Dissemination of the Decimal Metric System*
François Hochereau and Adel Selmi: *Measuring Animal Performance: Sociological Analysis of a Double Transaction between Genericity and Singularity*
Nadine de Courtenay, Oliver Schlaudt: *Round Table with* **Alexandre Mallard, Simon Schaffer et al.,** *On the Social Construction of Units of Measurement*

Disciplinary Formations *(p. 13)*

Chair: Donna Drucker Room: **Long Table**

Godfrey Guillaumin: *Scientific Measurement as a Cognitive Integration. The Case of Kepler's Astronomy*
Andrew Maul, David Torres Irribarra and Mark Wilson: *On the Conceptual Boundaries between Physical and Psychological Measurement*
Sébastien Plutniak: *Measuring the Foreign Past: A Sociohistorical Inquiry of the Use of Multivariate Analysis in French Prehistorians' Community*
Joshua McGrane: *Good Logic Does not Make Good Measurement: A Very Shaky Discussion of the Foundations of Measurement by a Social Scientist*

Thursday, 14 March • 14:00-16:00

Plenary Lectures *(p. 16)*

Chair: Astrid Schwarz

Joel Michell: *Militant Pantometry, the Logical Limits of Measurement, and the Fabrication of Psychometrics*
Michael Heidelberger: *The Function of Instrument Measurement*

Thursday, 14 March • 16:30-18:30

Images as Measurements *(p. 17)*

Chair: Nicola Mößner **Plenary Hall**

Sophia Efstathiou: *The Shiny and the Real: Fluorescent Dying Techniques as Means of Scientific Measurement*
Aud Sissel Hoel: *Images as Measurement*
Liv Hausken: *The Photographic Prototype in Cognitive Neuroimaging*
Tom Vogt: *Imaging at the Edge*

Studies on Calibration *(p.21)*

Chair: Johannes Lenhard **Round Table**

Léna Soler, Catherine Allamel-Raffin, Frederic Wieber:
What is Calibration? An Analysis of Calibration Focused on Normal Practices of the Investigation of Nature by Means of Already Standardized Instrumental Devices ***(three papers)***
Genco Guralp: *Calibrating the Universe: The Beginning and End of the Hubble Wars*

Measurement Practices *(p. 24)*

Chair: Leah McClimans **Long Table**

Elizabeth Neswald: *Creating Commensurability in Early Nutrition Science and Metabolism Studies*
Ramona A. Braun: *The Pump, the Woman, the Graph: Measuring Tubal Pressure in French and German Gynaecology, 1940s-50s*
Laura Cupples: *Epistemic Iteration: Enrichment and Self-Correction of Patient-Reported Outcome Measures*
Emily Brock: *Measuring Darwin's Entangled Bank: The Tools of Ecological Fieldwork*

Thursday, 14 March • 18:30-20:00

Dinner at the ZiF

Thursday, 14 March • 20:00-21:00

Plenary Lecture *(p. 27)*

Chair: Carsten Reinhardt

Simon Schaffer: *The Silent Trade: On the Boundaries of Measurement*

Friday, 15 March • 9:00-11:00

Plenary Lectures *(p.28)*

Chair: Michael Stöltzner

Nadine de Courtenay: *Connecting the Actual and the Virtual: The Double Interpretation of the Equations of Physics*
Luca Mari: *Outline of the Current Status of Measurement Science: From the Point of View of the International Vocabulary of Metrology*

Friday, 15 March • 11:30-13:00

Measuring the Unmeasurable *(p. 29)*

Chair: Andreas Kaminski **Plenary Hall**

Leah McClimans: *Can We Measure Quality of Life? You Decide*
Ann Johnson: *Measuring Complex Phenomena: The Problem with Proxies*
Johannes Lenhard: *When a Pea is Equal to the Sun: The Mathematics of the Unmeasurable*

Conceptual Issues *(p. 31)*

Chair: Insa Röpke **Round Table**

Flavia Padovani: *A Logical Space for Measurement*
Fabien Gregis: *Can we Dispense with the Notion of "True Value" in Metrology?*
Christian Hennig: *Measurement as Constructive Act - a Statistician's View*

Measurement and Engineering *(p.33)*

Chair: Marcel Boumans **Long Table**

Mieke Boon: *The Measurement of Properties in the Engineering Sciences*
Sjoerd D. Zwart: *The Foundations of Indirect Measurement and Model Laws*
Pablo Schyfter: *Measuring Promoters to Make the Field: Metrology as a Tool for Building Knowledge, Building Artifacts, and Building Fields*

Friday, 15 March • 14:00-16:00

The Politics of Measurement *(p. 36)*

Chair: Emily Brock **Plenary Hall**

Anne Harrington: *Nuclear Values*
Carlo Martini:*"Subjective" Measurement in Economics, but not Arbitrary*
Sharon Crasnow: *The Measure of Democracy: Coding in Political Science*
Cheryce von Xylander: *The Metrics of Tact*

Friday, 15 March • 14:00-16:00

Standards in Context *(p. 39)*

Chair: Ann Johnson **Round Table**

Shaul Katzir: *Frequency and Time Standards from Acoustics to Radio - the Road to the Quartz Clock*
Sharon Ku and Frederick Klaessig: *A Matter of Size Does Not Matter: The Social Construction and Application of Gold Nanoparticle Standard*
Lara Huber: *Measuring by which Standard? Plurality Challenges Epistemic Singularity*
Wolfgang Pietsch: *A Revolution Without Tooth and Claw: Redefining the Physical Base Units*

Human Dimensions *(p.42)*

Chair: Andreas Gelhard **Long Table**

Donna J. Drucker: *The Klein Sexual Orientation Grid and the Fluid Measurement of Sexual Identity*
Erik Angner: *The Problem with Happiness Measurement*
Zed Adams: *Man a Measure?*
Hans Radder: *Benjamin Libet's Measurement of Freely Willed Decisions: A Critical Analysis*

Friday, 15 March • 16:30-18:30

Plenary Lectures *(p.45)*

Chair: Alfred Nordmann

Laura Dassow Walls: *With Compass, Chains, and Sounding Line: Taking Thoreau's Measure*
Martin Kusch: *"A Branch of Human Natural History": Wittgenstein and Metrology*

Friday, 15 March • 19:30

Dinner at Brauhaus Joh. Albrecht
(Hagenbruchstraße 8, downtown Bielefeld)

Saturday, 16 March • 9:00-11:00

Plenary Lectures *(p.47)*

Chair: Martin Carrier

Mary Morgan: *Accounting Reasons - Reasoning Accounts*
Marcel Boumans: *Clinical Measurement*

Saturday, 16 March • 11:30-13:00

Measurement of Mind *(p.48)*

Chair: Laura Cupples **Plenary Hall**

Andreas Kaminski: *Measuring Intelligence and Temperature: Is it the Same?*
Andreas Gelhard: *Measuring Competence Rather than Intelligence? McClelland's Claim*
Malte Bachem: *Matching Personality to Vocation. A Historical Perspective on an Epistemological Model around 1920*

Simulations, Models and Measurements *(p. 49)*

Chair: Eran Tal **Round Table**

Annamaria Carusi: *Modelling Measurements and Constitutive Realism*
Wendy Parker: *Data Assimilation, Measurement and the Construction of Global Climate Datasets*
Teru Miyake: *Uncertainty and Modeling in Seismology*

History of Measurement and Measurement Theory *(p. 53)*

Chair: Cheryce von Xylander **Long Table**

Francesca Biagioli: *Empirical and Formal Conditions in Helmholtz's Theory of Measurement*
Stéphanie Dupouy: *How Meaningful is it to Measure Sensation? Discussions of Fechner in France in the 1870's and 1880's*
Jochen F Mayer: *On the 'Mathematisation' of Measurement in (West) German Academic and Official Statistics c.1920-1950*

Saturday, 16 March • 14:00-16:00

Plenary Lectures *(p.55)*

Chair: Oliver Schlaudt

Hasok Chang: *Operationalism: Old Lessons and New Challenges*
Eran Tal: *A Model-Based Epistemology of Measurement*

Abstracts

Thursday, 14 March • 11:00-13:00

The Epistemic Value of Visual Data

Visualisations – such as diagrams, computer graphics, graphs, photographs, maps etc. – are an essential part of current scientific practices. Scientists include them in their presentations and publications as a part of their research results. They use them as means to separate the essential from the mere contingent data – via curve fitting and the consideration of error bars –, to discuss them with their colleagues, and to find out more about the object, process etc. under investigation. Scientific visualisations, therefore, are both part of the context of discovery and of the context of justification.

In this sense, they are likewise objects of research and records of research activity. We can illustrate this with a little example: To measure the frequency of sunspots, on the one hand, and, thereby, to test whether the theory predicting upcoming events is correct or not also means analysing incoming data, i.e. to compare a corresponding diagram with other graphs in this context. Diagrams of these kinds then allow for further analysis, to search, for example, for maxima or minima and to predict the average development. Accordingly, you can find out more about your research object – the sunspots – and about the correctness of your theory at hand – displayed in older diagrams – with the aid of such visualisations. In the same sense, to proof the existence of the Higgs boson means that you have to discover the relevant peak in your diagram in a significant amount of measurement results. On the other hand, recording the success of your research means distributing, showing, and discussing the relevant diagrams amongst peers and the public.

In the context of measurement processes, we are thus often confronted with visual representations as the output of such data collecting activities. One reason for this consists in ubiquitous computing, i.e. the application of information technology in scientific experimentation. Software takes over the task of data processing, supplies researchers with elaborated plots and merely leaves the qualitative task of interpreting and assessing the resulting visual representations to the trained eye of the observer. Mediating software, however, is only one aspect why visual representations have gained such a significant part in measurement processes.

Another important point is that more and more measurement devices directly yield visual data. A variety of imaging techniques are applied in medical research and diagnosis, for example. Moreover, technologies such as electron or video

microscopes in microbiology or digital photography in astrophysics are used as detecting devices.
Taking these developments in science as background, the focus of this section is on the epistemic value of these visualisations. A variety of questions can be asked in this context such as: Do we need these visual representations in science? Which functional role do they play? Are they just a redundant way of information representation, i.e. fulfilling merely an illustrative function? Or is there more involved than mere heuristics? Can they be regarded as evidence? What do they tell us about the existential status of our research object? Can they serve as visual proofs? Which kinds of epistemic problems are maybe correlated with the task of interpretation in this context? After all, the digital nature of those images allows us to apply different techniques for visual enhancement here. Then what does it, for example, mean to the epistemic status of such visual representations if we use techniques like false colours depiction etc.? And, finally, what about the scientists' responsibilities with respect to their peers and to the public when using such figures in their work?
Questions of this kind will be considered in the four talks of this section, namely:

Patrick Maynard: "*Photo Mensura"*

Protagoras' 'homo mensura'—"humans the measure of all things"—has distinct applications to our topic: for the collection of ordered data and for its presentation; for what shows up when we measure nature, and for how we show that, making our measurements accessible to ourselves. Issues of objectivity arise for each. Perhaps we can begin to grasp some of photography's complex contribution to scientific measurement by approaching it as photography—a constantly developing family of technologies, powers for doing things—rather than in terms of certain objects, 'photographs'. From its 1839 presentation, photography was already understood in terms of several distinct powers—notably those of detective registration, reproduction and depiction—taking places beside already established procedures for each, and it has never ceased to be so. Common confusions about photography have been caused by failures, first to distinguish these roles, then to understand how, like most multi-functional entities, its different functions frequently interact, and in a variety of ways. After making a case for photography's contribution to objectivity in measuring, as contrasted with extreme relativity rather than (as on current understandings) with subjectivity, we briefly consider two concrete cases of detective/depictive photographic combinations for measuring and displaying measurement.

Laura Perini: *Data images, representation, and the creation of scale*

Basic research often yields visual data, but that is only a first step in the production of an image that can serve as evidence in scientific reasoning. Historical, social science, and philosophical studies have shown that often data images are subjected to intensive practices of interpretation before they can be used as representations. Furthermore, data images are often altered in various ways and combined with other images when they are published in research articles. In this talk I will show how the evidential function of data images depends on these practices of interpretation and alteration.

Nicola Mößner: *Visual Data – Reasons to Rely on?*

In today's science, the output of measurement processes are often visual representations of the data detected. Moreover, we find such visual data as parts of scientific reasoning in different contexts. In this talk, we will take a look at two of them. On the one hand, visual representations are used as a kind of surrogate for the real object to ask question about it – we will call this the exploratory use of visual data. On the other hand, visualisations are often used to communicate research results to scientific peers or to laymen. Instances of the latter case we will subsequently call the explanatory use of visual data. However, in both contexts the question arises what might justify these epistemic practices. Believing without reasons is epistemically irresponsible, it leads to gullibility. Thus, we will try to answer the question whether there are any reasons to support our epistemic practices concerning scientific visualisations in the above-mentioned senses and whether these are good reasons to rely on.

Tobias Schöttler: *Pictorial Evidence: On the Rightness of Pictures*

The epistemic value of pictures results from their use in scientific reasoning. This kind of use presupposes their truth-aptness or more generally their aptitude to be right. We can distinguish two strategies to justify or to explain the rightness of pictures: a naturalistic and a pragmatic strategy. According to the naturalistic strategy, the objectivity of pictures is based on their similarity or causal relation to the depicted objects. The naturalistic account fails. The similarity relation is neither necessary nor sufficient for the picture's reference. In most cases, we use pictures as surrogates for the depicted objects since we do not perceive the objects themselves, thus we cannot compare the picture and its depicted object. Therefore, we cannot know if the picture resembles the depicted object. The causal relation alone does not explain the picture's content, too. Therefore, I am going to argue for the pragmatic strategy. This strategy justifies a picture's objectivity by its

›rightness of fit‹ (Goodman). Since the attempt to establish an external objectivity by the picture's intrinsic relation to depicted objects fails, the pragmatic account is contend with the internal objectivity and investigates its criteria.

Sociology of Metrological Knowledge

Hector Vera: *A Single Language of Measurement for Humanity: the Global Dissemination of the Decimal Metric System*

Ninety-five percent of the world population live today in countries where the metric system is the only legal system of measurement (or at least it is committed to achieve full metrication); the other five percent lives in the seven countries (Liberia, Myanmar, United States, Independent State of Samoa, Federated States of Micronesia, Palau, and Marshall Islands) where the metric system is only optional, i.e. it can be used legally, but its utilization is not mandatory. This makes the metric system the only measurement system with which a commercial or civil contract can have validity in every single nation of the planet. Surprisingly, there is a considerable lack of thorough research on the international propagation of the metric system. Most accounts pay a great deal of attention to the creation of the system and to its technical novelties. Little is known, however, about the details of how the meter spread throughout the world, and why. Since the 1960s almost no academic work has traced the global metric spread, which means that the considerable metrication activity that took place during the last decades of the twentieth century has been completely overlooked. Besides, what has happened outside Europe has been ignored.

This paper analyzes how geographical and socio-political conditions have shaped the international diffusion of the decimal metric system of weights and measures, from its invention during the French revolution at the end of the eighteenth century to the latest national adoptions of the system at the end of the twentieth century. The results are based on a dataset on international metrication specifically assembled for this research. The dataset covers all 196 countries according to the grid of national states in 2010. How and when countries adopted the metric system as their official system of measurement is explained considering geographical factors (for example vicinity) and historical circumstances (like revolutions, national unifications, foreign military imposition, colonialism, decolonization, and periods of rapid and massive social transformation). Moreover, the paper underlines the significant influence of colonization in spreading the metric system outside Europe (particularly, but not exclusively, through French colonialism in Africa), stresses the pioneering role of Latin American countries in making the metric system a legitimate international convention during the nineteenth century, highlights the American imprint in the seven remaining countries that have not adopted the

metric system, and analyzes why the United States have failed to adopt the metric system as its exclusive and mandatory system of measurement despite several metrication campaigns design by the federal government.

François Hochereau and Adel Selmi: *Measuring Animal Performance: Sociological Analysis of a Double Transaction between Genericity and Singularity*

Today, measurement is diffuse in all aspects of social life It is mobilized. in order to assess human activities, their mutual relations, their results and their resources. Its influence is thus twofold: it acts as a powerful means of governability (Foucault 2004), yet it also supports knowledge acquisition, mutual understanding and agreement between people on the things that they deal with on a daily basis (Dagognet, 1993). Measurement proceeds as a dual process of aggregation that makes things governable and specification that allows it to operate at different levels as a device of proof and of agreement. .The purpose of the metric is thus to reduce the spectrum of measurement so to provide useful information for decision-making, while building a common language between people to reach agreement.

In order to investigate this double transaction between genericity and singularity, we have studied how standard measurements of animal performance were constructed and appropriated in French animal selection. These standards reflect the collective ambition to modernize French breeding by guiding animal selection toward high performance breeds, which is supported by an intercomparison tool for each farmer that can help them to better choose their animals. Moreover, two temporalities can be distinguished in the measurement process: animal performance is first captured through quantifiable characteristics such as weight and growth rate; and second, qualitative criteria such as carcass quality or animal behavior are assessed.

In this paper, we focus on the collective process of establishing metrics through the practical organization of people and resources, constitutive of the animal selection system. We analyze the social impact of measurement, i.e. the way in which measurement is institutionalized by explaining the chains of cognitive translation and mediation through the construction of shared categories. We identify the "sense of measurement", that is, what we measure, how we measure, why we measure, what is hidden through measurement, and the system of values that support the act of measurement. This leads us to reflect upon the complex relationship between measurement, quantity and quality already described by Dewey (1967).

Nadine de Courtenay, Oliver Schlaudt: *Round Table with* **Alexandre Mallard, Simon Schaffer et al.,** *On the Social Construction of Units of Measurement*

At present, there is an important reform of the International Systems of Units (SI) going on, aiming at a fundamental redefinition of the base units. It consists essentially in fixing the values of some fundamental constants and deriving from them the units of measurement. This idea appeared for the first time in 1983 when the meter was linked to velocity of light, and according to the International Committee for Weights and Measures the same method should be applied to all base units. As regards the definition of the kilogramme - the last unit still defined by an artifact -, there are two competing methods, the Avogadro project and the Watt balance, none of them however currently providing the precision demanded by the International Committee for Weights and Measures.

Though of fundamental importance, this reform seems not to attract much attention from philosophers, historians, or sociologists of science. Some metrologists indeed speak of metrology in general as an "invisible infrastructure" of our society (cf. Quinn and Kovalevsky 2004). There is some sociological work on metrology in general and its various aspects (standardization, quantification, commensuration, precision, history of the metric system). But the current reform of the SI still lacks attention.

It is our aim in this round table to discuss the interest of sociological perspectives on the current reform of the SI, to single out the points of interest for sociological investigation, and perhaps even to frame such a study. Two points will be of special interest to us:

(1) In how far does the strategy of linking measurement units to to the most fundamental and thus most invariant features of the universe pose a challenge to the sociology of scientific knowledge? Does it challenge sociologist's of science account of "absoluteness"? Or, on the contrary, does the "(non-vicious) circularity" of this strategy offer any special perspectives to sociology?

(2) How far---with respect to epistemological questions---can the sociological analysis of the reform be pushed? Does it confine itself to the social causes for the choice between scientifically equivalent options in the realisation of the reform? Or can it even account for what gives a value to a method and what makes different methods to be of the same value ("equivalent")? In particular, which perspectives for sociological investigation offers the concurrence of the Avogadro project with the Watt balance?

Participants will be Simon Schaffer (Cambridge), Alexandre Mallard (Center for the Sociology of Innovation, Mines ParisTech), as well as the other speakers of this panel, Héctor Vera, François Hocherau and Adel Selmi.

Disciplinary Formations

Godfrey Guillaumin: *Scientific Measurement as a Cognitive Integration. The Case of Kepler's Astronomy*

Scientific measurement was a central factor behind the rise of modern science. Nevertheless, despite the epistemic and methodological importance of measurement for the development and growth of scientific knowledge during the modern epoch, few historical-philosophical studies seek to understand measurement's epistemic, methodological, and cognitive aspects. This paper aims to propose a historical- philosophical approach for scientific measurement in terms of an integration of mathematical, conceptual, and instrumental aspects. This integration was achieved after long periods of time, as result of "feedback" processes between these three aspects, where each other were mutually adjusted. I illustrate this dynamic and integrative notion of scientific measurement analyzing some of Tycho Brahe's and Johannes Kepler's determinations of celestial parameters.

Andrew Maul, David Torres Irribarra and Mark Wilson: *On the Conceptual Boundaries between Physical and Psychological Measurement*

The concept of measurement has been a cornerstone element of epistemology in the physical sciences since their inception. More recently, the social sciences have developed a variety of techniques that purport to be instances of measurement as well. However, it is not clear how the way in which the concept of measurement is understood in the social sciences accords with the way it is understood in science more generally.

In recent years, a number of scholars have subjected the conceptual and philosophical foundations of psychological measurement to vigorous investigation and critique; on one reading, this body of work seems to suggest that the way in which 'measurement' is understood by social scientists is entirely dissimilar to the way in which it is understood by physical scientists and philosophers of science.

In our experience, professionals of measurement in the social sciences have reacted to this body of work in one of two distinct ways. The first is to accept the premise that there is but one meaning of 'measurement' throughout all of science, with which social scientists are out of touch, and therefore, there may have not yet been any successful instances of genuine measurement in the social sciences. The second is to deny that the concept of 'measurement' has or needs to have a consistent meaning across scientific disciplines, and so the fact that social scientists have a different understanding of 'measurement' than do physical scientists is not,

in itself, problematic. In the first case, the word 'measurement', when used by social scientists, is a metaphor at best and a conceptual error at worst; in the second case, the word is merely a homonym.
We argue that neither of these positions is necessary: in fact, it is possible to formulate a philosophically coherent account of how successful psychological measurement works, in a manner consistent with mainstream positions in philosophy of science. In order to formulate such a position, we first review three major intellectual strands that characterize a great deal of thinking about psychological measurement: first, empiricism, including positivist, operationalist, and, later, representationalist modes of thinking; second, scientific realism; and third, pragmatism. We argue that there are misconceptions and shortcomings associated with each of these strands of thought, but that once these issues are addressed, the traditions are not as incompatible as they might initially appear. In particular, more mature forms of realism, such as Putnam's natural or pragmatic realism, acknowledge that while a natural world does exist independently of observation, there is not necessarily only one true and complete description of that world; rather, it is the interaction of the world and the rich fabric of our linguistic, conceptual, and mathematical schemes and models that jointly determines what we see. Although psychological and physical measurement activities differ in many respects, including the particulars of the schemes and models that are brought to bear on their respective situations, they do share common foundational philosophical commitments, and need not be thought of as fundamentally different.

Sébastien Plutniak: *Measuring the Foreign Past: A Sociohistorical Inquiry of the use of Multivariate Analysis in French Prehistorians' Community*

At the end of the 1960's, sociologists and historians of science developed a particular interest in the notion of discipline and the dynamics of science activities (Kuhn 1962, Ben-David 1966, Mullins 1972, Mulkay 1975). Thirty years later, this topic sparked attention, more particularly through the propositions published by Callon (1995) and Abbott (2001). The gap between these two periods was occupied by the development of science studies and by the application of ethnographic methods to science practices: sociologists aimed at following science "in action", the inscriptions, instruments, and actors at the heart of laboratories. Considering these two perspectives, my communication aims to investigate the relationship between an instrument and the dynamics of a discipline. The use of multivariate analysis in French prehistoric archaeology will be the case I shall propose to investigate.

Archaeologists deal mostly with tangible facts; they start by accumulating them, and then try to discriminate between them before finally interpreting them in terms of anthropological factors.
Basically, archaeology can be understood as the science of evaluating and interpreting the differences between the material forms — stratigraphic as well as artefactual ones — which are persistent through time. During the 1960's, the simultaneous occurrence of the increase of excavations, ever more accurate methods of artifact description, and the rise of rescue archaeology resulted in giving the archaeologist an increasing mass of data to face. Contemporary innovations in statistical research, and especially in multivariate analysis, provided a way to analyze these data. Principal component analysis, or J.-P. Benzécri's correspondence analysis, aroused an enthusiastic interest in various research fields, including archaeology. These statistical instruments helped the archaeologists to produce distance measurements between collections of archaeological materials. The quantitative results of these statistical distances – or, simply, their visualizations – were directly involved in establishing assertions about foreign prehistoric phenomenon, in cultural or sociological terms.
Based on researchers' life narratives and on a bibliographical survey, my communication will consider two aims. Firstly, in a sociohistorical perspective, I shall establish a description of the conditions of the possibility of the introduction of multivariate analysis in the archaeological field including the material context, the actors, the relationships, the resources, the places, and the epistemic roles lent to these analyses; all of them being bound together. Thus, I shall underline the effects of the introduction of this new kind of measurement on both the scientific community and on the definition of their object of knowledge. Secondly, based on this study, in more abstract terms, I shall suggest a local model of discipline dynamics, dealing with the notion of self-similarity that has been proposed by A. Abbott (2001). My proposal will focus on the permeability of disciplines and the recursivity of their interrelationships with regard to methods and instruments.

Joshua McGrane: *Good Logic does not make Good Measurement: A Very Shaky Discussion of the Foundations of Measurement by a Social Scientist*

"The most distinguished physicists, when they attempt logical analysis, are apt to gibber; and probably more nonsense is talked about measurement than about any other part of physics." (Campbell, 1938, p. 121)
Throughout the 19th and 20th centuries, numerous physicists, mathematicians and philosophers attempted to explicate a logical foundation for measurement, which culminated in the representational theory of measurement. This approach was seen as shifting theoretical discussions of measurement from, "incomprehensible

verbiage about units and dimensions" (Campbell, 1938), to a discourse grounded in the logical principles of mathematics and the axiomatisation of key concepts, including number, quantity and continuity. Such discourse dominates contemporary philosophical and mathematical discussions of measurement and is at the core of the understanding of measurement in the social sciences, e.g., economics and psychology.

Over the same period, less logically inclined physicists, or those 'apt to gibber' in the above quotation, developed the systematic approach to measurement. Under this approach, measurement is made possible by physical theory and laws that determine the definition and realisation of a system of units. The coherence between physical laws and the unit-system is ensured by principles of dimensional analysis, including dimensional homogeneity. Systematic measurement, despite its minimal consideration in more philosophically inclined discussions, forms the basis for measurement throughout the physical sciences (and trade, commerce, manufacturing, etc.) by way of the International System of Units (SI).

Whilst some, such as Campbell, have suggested that the empirical success of the systematic approach is contingent upon a representational foundation, particularly in explaining measurement de novo, I will put forward a number of arguments that the two are incompatible paradigms. Firstly, the representational, set-theoretic theory of number will be compared with the systematic, realist conception of number, whereby the former is argued to entail a category error. Secondly, the representational and systematic definitions of measurement will be contrasted, with only the latter argued to be consistent with physical theory and law. Thirdly, the representational conception of continuous quantity will be argued to violate the systematic principle of dimensional homogeneity, as first elucidated by Zeno's metrological paradox. Fourthly, the unit concept under the representational and systematic viewpoints will be contrasted. I will then, being a psychologist by training, attempt to elucidate some non-representational, systematic principles of measurement for social scientists

Thursday, 14 March • 14:00-16:00

Plenary Lectures

Joel Michell: *Militant Pantometry, the Logical Limits of Measurement, and the Fabrication of Psychometrics*

The virtues of measurement are jewels in science's crown, impatiently coveted by proponents of aspiring sciences, convinced that measurement lacks logical limits and, thus, that all attributes are in principle measurable. This is pantometry. Taken as an unquestioned axiom of investigation, it means that the possibility of

alternatives to measurement is denied. This is militant pantometry. It ignores at least one of the following: that measurement is the estimation of the ratio between a magnitude and a unit of the same attribute; that ratios require that relevant attributes are quantitative; and that attributes logically incompatible with quantitative structure exist. That is, it ignores that measurement has logical limits. Despite this, militant pantometry has forged modern disciplines, all in the name of measurement and science. One pivotal case is the prefabrication of psychometrics through the work of Galton and Binet. Since becoming an independent, applied discipline a century ago, its proponents have always held, as an absolute presupposition, the proposition that psychological attributes are quantitative despite the fact that the attributes tests assess present themselves to us as logically incompatible with quantitative structure. It provides a salutary litmus test: for any practice shaped under the auspices of militant pantometry, it is never safe to infer that those called "measurement" are measurement. Evidence of quantity is always necessary.

Michael Heidelberger: *The Function of Instrument Measurement*

Thursday, 14 March • 16:30-18:30

Images as Measurements

Sophia Efstathiou: *The Shiny and the Real: Fluorescent Dying Techniques as Means of Scientific Measurement*

Highly sophisticated methods and equipment are used for making things on the molecular level visible, for telling them apart, and for tracking their evolutions in cellular or molecular spaces.
This paper focuses on two techniques using fluorescent dying as a means of bioscientific measurement. Both cases are developed through collaborative humanist, physiology and systems biology work, on gastrointestinal and heart research. The first case examines the use of Fluorescence Recovery After Photobleaching (FRAP) to study protein mobility and activity in living cells: specifically as applied to the study of cancer cell responses to stimulation with the hormone gastrin, specifically the role of nuclear receptor NR4A2 when gastric adenocarcinoma cells are stimulated by gastrin. The second case involves the study of calcium signaling during the contraction of heart muscle, also studied using fluorescent dyes and live recording and monitoring dynamics, while the heart muscle (freshly harvested from animal models) is alive (see for instance http://www.youtube.com/watch?v=7LyZkNeyw9s).In both cases the ability to

establish biological activity at a normally invisible scale, to measure and to quantify it relies on making it visible; and making visible relies in part on the contrasting colours (and colour-uptake) of the thing stained or dyed and its background: creating a visible, often shiny effect. In both cases the biological phenomena studied are dynamic, and the distribution and change of a colour pattern is meant to communicate and help measure a change in the thing itself, under measurement. The shininess of the vision is seen as an access to its reality.
This specific mobilization of our everyday ability to tell apart colours and their intensity is investigated in conjunction with the computational and mathematical expressions that frame and accompany such measurements (Lynch 1988). Sociotechnically orchestrated and experimentally challenging, technologically and computationally mediated, these techniques STILL build on everyday human abilities for colour-sorting, and the allure and ease that sorting things by visible intensity and colour has for most of us (cf. Osbeck and Nersessian for affect and reason in systems biology). Playing on the work of holist neurobiologist Kurt Goldstein on colour-sorting tests as measures of ability for "abstract thinking" (Goldstein 1995[1935], Harrington 1996), the paper discusses the specific abstractions that these colour-sorting bioscience techniques make possible in the context of this bioscience research.
In this way, the paper argues, these highly sociotechnically mediated, seemingly everyday practices of looking for shiny things, can be understood as installations of everyday, nonscientific practices into scientific contexts: a case of what I call "found science", by analogy to "found art" (Efstathiou in press). Found science proposes that some everyday notions and practices can become embedded, "founded" into scientific contexts, and thereby transfigured into ideas and practices that we recognize as scientific ideas or methods, while still retaining symbolic and actual relations to everyday contexts, often through use of the same "names" for concepts or tools. Fluorescent colour-sorting techniques occupy that space between the ordinary and the extraordinary and scientific.

Aud Sissel Hoel: *Images as Measurement*

Drawing on Ernst Cassirer's functional approach to concepts (Substanzbegriff und Funktionsbegriff, 1910), and subsequently, to symbolic forms and technology (Philosophie der symbolischen Formen, 3 vols , 1923-1929; "Form und Technik," 1930), I develop a notion of images that conceives of images as measurement technology. This approach is spurred by scientific imaging and visualization practices in the current computerized situation, but I go on to make the broader argument that all images, even in the case of simple line drawings, stands in a measured relation to their target phenomena. My aim is to develop a notion of images that is no longer based on representationalist beliefs and, hence, on the

substantivist and dualist scheme criticized by Cassirer. I approach images, not as representations or documents of pregiven realities, but as imaging tools that stand in a generative and what I have coined as a "differential" relation to their objects. What I hope to show is that imaging tools operate by enacting a specific and targeted division between figure and ground.

The differential mode of operating is the link between quantitative methods of measurement and imaging methods, which are commonly classified as observational. However, in order to establish this inner connection between methods of measuring and methods of observation, the terms "measuring" and "measurement" need to be broadened. In my interpretation, such a broadening is precisely what Cassirer sets out to do, first in the context of mathematical and scientific thought, and later also in the wider context of the cultural sciences. The clue to this broadening is the emphasis Cassirer puts on mediating symbolic forms, or as I prefer to put it, mediating apparatuses. In "Form and Technology," for example, he states that technology "grounds and steadies … a type of mediacy that belongs to the essence of thought." Interestingly, Cassirer conceives mediating forms as tools, each of which incorporates a dynamic principle that serves at once as a standard or measure according to which the object is delineated, and as a guiding point of view according to which the object is seen and accessed. Hence, in Cassirer's view, the logical and the visual are not radically distinct. Thus understood, the mediating apparatus takes on a dual role as generative principle and guiding sight, which is to say that the intervening tool actively participates in the formation of the object's identity. Even if there is a sense to which the intervening apparatus produces the object of knowledge (it introduces a new "object" specific for that particular measurement situation, or what Lorraine Daston and Peter Galison have referred to as "working objects"), the mediating instruments are understood to reveal aspects of reality. The intervening apparatus does not cover up reality but provokes it to answer back along the lines specified by its parametric set-up. Mediating apparatuses, in other words, are understood to enter into a differential exchange with reality.

In this paper I argue that images, conceived as imaging tools, operate by a differential logic of the kind outlined above. I substantiate my case by examples from medical imaging.

Liv Hausken: *The Photographic Prototype in Cognitive Neuroimaging*

This paper will argue that the photograph figures explicitly or implicitly as a prototype in scientific imaging. It is not just treated as an image modality comparable to others, but as a model for other modalities, and the first and most typical examples of images as such, sometimes even the standard of all imaging. This prototype or standard image is based on a very specific conception of

photography often referred to as transparency, optical empiricism and mechanical objectivity while neglecting other conceptions of photo which are equally important in the history of photography in science, namely the camera as a measuring instrument. In the 19th century the photographic camera emerges as a truth-apparatus despite of the crisis of faith in optical empiricism, due to the idea of its metrical accuracy. As Allen Sekula convincingly has illustrated, “Here was a medium from which exact mathematical data could be extracted, or as the physicist François Arago put it in 1839, a medium 'in which objects preserve mathematically their forms.' For the nineteenth-century positivists, photography doubly fulfilled the Enlightenment dream of a universal language: the universal mimetic language of the camera yielded up a higher, more cerebral truth, a truth that could be uttered in the universal abstract language of mathematics" (Sekula 1986 p.17). From the history of ethnology and physical anthropology, the development of social statistics, criminology, and also of motion studies in the 19th century through the 20th century and until the current culture of biometrics and medical and social surveillance, photography and camera based imaging has been an important instrument for measurement. In this paper I will demonstrate how this alternative conception of photography would make a difference in the discussions of brain imaging. Brain imaging technologies has been enthusiastically embraced both by scientific and medical community as well as by the public through popular media reports of neuroscientific methods and findings (Weisberg 2008). Brain images are seen as epistemologically compelling (Roskies 2007) and are believed to have a particularly persuasive influence on the public perception of research on cognition (McCabe & Castel, 2008). In academic discussion of this rather popular topic it is more often than not made clear that Neuroimages Are Not Like Photographs of the Brain (to rephrase the title of A.L. Roskies article in Philosophy of Science, 74/2007). I will obviously not argue that an MRI scan indicating internal structures of the brain equals the photographic registration of visible surfaces. I will however suggest that these debates about the persuasive power of brain imaging in contemporary society demonstrates a problematic conception of photography which among other features neglect the importance of measurement in photography. This is important not just for our conception of photographic images but also for how we deal with the role of neuroscience in psychology, education and law and, more generally, for the conception of scientific imaging as such.

Tom Vogt: *Imaging at the Edge*

Recent advances in imaging use novel concepts based on the fact that images are sparse and therefore might allow us to circumvent the constraints of the Shannon-

Nyquist sampling theorem. This talk will give an introduction to compressed sampling and outline a paradigm shift in imaging.

Studies on Calibration

Léna Soler, Catherine Allamel-Raffin, Frederic Wieber:
What is calibration? From calibration in practices of investigation of already well understood phenomena by means of already standardized instrumental devices to more problematic cases ***(three papers)***

This symposium deals with calibration, starting from calibration in a certain kind of scientific practices, namely practices which investigate relatively well-understood natural phenomena with already standardized instrumental devices – for short, say 'normal practices of nature investigation'.
Calibration is a crucial but neglected topic. It is a crucial topic from an epistemological point of view, since it essentially conditions the reliability of instrumental procedures which play a major role in science. The outputs of measurement apparatuses are no more than marks deprived of a determined physical significance, unless the apparatuses in question have been correctly calibrated. Yet although important, calibration is a relatively neglected topic. This is probably because in many configurations at least, calibration is treated as a preliminary unproblematic procedure that precedes the 'main show' – namely the characterization of objects of interest by means of what has been previously calibrated in this purpose. Under closer examination, however, calibration is rarely a straightforward matter. The study of calibration practices shows that even in the at-first-sight less problematic configuration, that is, in relatively routine scientific practices, calibration procedures are often delicate, complex, and not obvious activities. This a fortiori holds for calibrations involved in more creative, not to say "revolutionary" scientific practices.
Given that, we think that more attention should be devoted to the topic of calibration. The present symposium attempts to take a step in this direction. The aims are: (i) to characterize calibration in normal practices of nature investigation; (ii) beyond this particular case, to provide conceptual and taxonomic tools of broader scope that helps to get a better understanding of what is (or can be) calibration; (iii) To consider more briefly calibration in other types of scientific practices than normal practices of nature investigation; (iv) to propose a general definition of calibration.
To analyze a case of calibration, we investigate four questions. 1) The target T of calibration: What kind of thing can be the object O of a calibration? 2) The presuppositions Ps of calibration: what is taken for granted about O, which delimitates what is not granted and has to be checked and controlled? 3) The aim of calibration applied to the object O under presuppositions Ps. 4) The procedure of

calibration: the nature of its structural elements and the kind of logical stages through which the aim of calibration is achieved.

When attempting to introduce some order into the network of activities that can be identified with calibration, it appears useful to distinguish different types of scientific practices with respect to which calibration takes different forms and might be more or less problematic. We shall provide a rough typology, and will then focus, as a starting point, on one of the most simple and less problematic case: calibration in normal practices of nature investigation. This case corresponds to calibration from the standpoint of scientific practitioners who are users, rather than designers, of instrumental devices: to situations in which practitioners use already well-mastered and entrenched instruments as means to achieve their ultimate aim which is the exploration of the natural world. Moreover, within such practices, we start from configurations in which the natural phenomena under scrutiny do not pertain to a completely new domain but are, as far as their main global features are concerned, already understood.

But even focusing on normal practices of nature investigation, multiple, not always homogeneous activities might be called calibration. To cope with such a situation, we start with the construction of a simple "exemplar" of calibration. To exemplar corresponds to determined answers to the four questions of the target, presuppositions, aim and procedure of calibration. The simple exemplar is part of the analytical framework we propose in order to understand calibration, and it has two functions: it provides a first core prototypical characterization of calibration (for normal practices of nature investigation); and it works as a tool for the scrutiny of more complex, more problematic and less prototypical candidates to calibration – the latter being analyzed in reference and by contrast to the simple prototype.

Next we turn to more complex and problematic cases of calibration. First, we examine one more complex case of calibration in normal practices of nature investigation: calibration in X-ray diffraction experiments. Compared with the simple prototype, this calibration involves a treated-as-composed instrumental device, which draws the attention to the relations between calibration of an instrument as a whole and calibration of its parts. In relation to this complexity, the case is specific and instructive in respects that are analyzed. Second, we broaden the scope, and give some insights on calibration in scientific practices different from normal practices of nature investigation: calibration in the investigation of new phenomena with already standardized instruments; and calibration in the invention of new instruments. Finally, we propose a general definition of calibration.

*This paper has been inspired by a collective work developed by some members of the PratiScienS research group directed by Léna Soler in Nancy (France). The six participants of this group were: Léna Soler, Catherine Allamel-Raffin, Catherine Dufour, Jean-Luc Gangloff, Emiliano Trizio and Frédéric Wieber.

Genco Guralp: *Calibrating the Universe: The Beginning and End of the Hubble Wars*

For many cosmologists, we are in the golden age of cosmology today. The latest example of this is the 2011 Nobel Prize in physics that was awarded to two teams which, working independently, confirmed the striking fact that the expansion of the universe is accelerating. This prize marks another major point in the chain of successful results cosmology obtained in its relatively short history of being an "experimental science." In fact, modern cosmology prides itself for becoming a precision science, breaking sharply with its "speculative" past. In this study, I examine one episode that played a key role within this epistemological dynamics of modern experimental cosmology, namely, the intriguing history of the measurement of the Hubble Constant.

Historians of science generally agree that Edwin Hubble's discovery of the expansion of the universe, on the basis of his observations of "extra-galactic nebulae," opened a new period in the history of observational cosmology. Hubble observed a linear relationship between the distance and the velocity of the receding galaxies, yet the value he obtained for the constant of proportionality between them was stricken by gross systematic errors. His former student, Alan Sandage, who was a major influence on the 20th century astronomy through his observational program, set the stage for the post-war attempts to determine the Hubble constant. His calibration scheme of "principle of precision indicators" aimed at selecting the best standard candle at each level of the cosmic distance scale and using only this method for calibration throughout the measurement process. This approach was opposed by the French astronomer, Gérard de Vaucouleurs, who meticulously devised his alternative calibration scheme of "spreading the risks," which advocates the methodology of using as many techniques as possible and then averaging over them. For more than three decades, these towering figures produced two incompatible values for the Hubble constant that they measured on the basis of their respective methods; Sandage promoting a value of 50 km/s/Mpc, opposing de Vaucouleurs's 100 km/s/Mpc. This conundrum, which then came to be known among astronomers as the Hubble Wars, was only resolved in the early 2000's, mainly due to the efforts of a collaboration known as The Hubble Space Telescope Key Project, which was specifically formed with the aim of determining the Hubble constant to an accuracy of ±10%.

In my paper, I analyze how this conundrum of measurement was resolved. I argue that the best way to understand this resolution is through situating it within the general transformation that cosmology underwent in the early 90's, which is known as the "precision era." Drawing on Galison's work, I claim that, within the precision era, a new material culture of calibration came into play which opened up an experimental trading zone in which various working-groups using different methods sought error-reduction in measurements as their primary goal, as

opposed to a "philosophical" commitment to a single methodology that we see in the case of Sandage and de Vaucouleurs. Following Chang, I further observe that the realization of this epistemic goal of error-reduction exemplifies the process of epistemic iteration, in the sense that each step in the calibration process is iteratively built on the previous one with the singular aim of minimizing error and thus resolving the discrepancy. In other words, the material culture of the precision era does not privilege particular methodologies but prioritizes an episteme of measurement which focuses on the production of a unique value of the constant that the scientific community consents to, albeit within error limits.

Measurement Practices

Elizabeth Neswald: *Creating Commensurability in Early Nutrition Science and Metabolism Studies*

This paper will look at questions of measurement in defining the objects, approaches and methods of nutrition physiology in the later nineteenth and early twentieth centuries. A young science, nutrition physiology emerged in a period in which quantification and measurement were becoming established as central methods in the life sciences. Encompassing areas as diverse as metabolism studies, dietary surveys and respiration calorimetry, it struggled from the outset with questions of what should be measured and how. The combination of laboratory experiment with a direct application orientation placed it at the intersection of the physiological and the social sciences with their differing methods of measurement and definitions of precision. As such, it reveals perhaps as no other science, the difficulties and compromises that were necessary to define the objects, goals and meanings of measurement. The aim of this paper is to show on the case study of nutrition physiology on the one hand the role of measurement in the definition and formation of a research field and its contested areas, and on the other the ways in which different dimensions of measurement intersect, strengthen and undermine one another. The paper will focus on two dimensions of measurement in nutrition science and their "apparatus of commensurability": dietary standards and norms and metabolism measurement.

Ramona A. Braun: *The pump, the woman, the graph: Measuring Tubal Pressure in French and German Gynaecology, 1940s-50s*

The Fallopian tube is a vague object. It needs to be measured and tamed. This was the notion established in 1920s gynaecology and that was still valid in the 1940s and 50s, the focus period of this paper.

Measuring pressure and measuring infertility were intricately linked in medicine up until the 1970s. Carbondioxide pumped into the tubes seemed to translate the organic cause of infertility directly to the graphical method.
Discussions about how, why and when the tubes move were at the centre of two World Congresses of Fertility and Sterility in 1953 and 1956. Tracing the most prominent experiments performed around the dogma of "infertility made visible", the paper analyzes French, German, Swedish and American proposals to get behind the secrets of reproductive physiology. Used as a therapeutic and even 'operative' method, tubal pressure made the process of measurement a procedure questioning the very nature of a stable object in the life sciences.
Arguing that mechanical objectivity was very much alive in Twentieth century gynaecology, the paper describes the importance of lab culture for the stabilization of the gynaecological discipline.

Laura Cupples: *Epistemic Iteration: Enrichment and Self-Correction of Patient-Reported Outcome Measures*

In this paper, I describe the problem of epistemic circularity in measure development in the physical and social sciences. I discuss Hasok Chang's coherentist framework for negotiating this circularity in the development of temperature measures. Using the process of epistemic iteration, metrologists successively refine and correct their measures, learning about the behavior of those measures' target constructs by putting them to work in concrete situations.
I argue that Leah McClimans's strategy for learning about patient-reported outcome measures (PROMs) and their target constructs, quality of life and subjective health status, parallels Chang's strategy of epistemic iteration. Like Chang, McClimans advocates learning about these measures by putting them to work in real populations. However, McClimans advocates a second commitment as well. According to McClimans, the questions that make up PROMs should be treated as genuine, in the Gadameran sense, meaning that they should be treated as open to interpretation and that even unexpected responses should be taken seriously.
I contend that while these commitments on McClimans's part allow us to advance our understanding of PROMs and their target constructs, they carry important and perhaps unforeseen consequences. Part of taking patients' responses to PROMs' questions seriously is admitting to variability in the meaning of PROMs' target constructs, quality of life and subjective health status. This variability of meaning prevents PROMs from converging around a well-behaved, ideal standard. We cannot expect PROMs' target constructs to behave in law-like fashion. Whether this limitation is a serious problem depends on our goals for PROMs.

Emily Brock: *Measuring Darwin's Entangled Bank: The Tools of Ecological Fieldwork*

Charles Darwin's evolutionary theory balanced on the cumulative effects of natural selection through ecological competition. Individuals within each species competed against one another, and also struggled against pressures from the ecological milieu. As important as these struggles were for Darwin's theory, they were rarely studied in a rigorous, direct way during his time. Nineteenth century natural history focused on describing and measuring static states of nature—the way a species looked and where it was found. Those forms of observation left the action of evolutionary struggle to be inferred. Evolutionary theorists constructed their work using the evidence generated by natural history, but natural historians offered little response to evolutionary theory. Nineteenth century natural history and evolutionary theory, while topically adjacent, could never be fully interacting, commensurable disciplines.

Growing scientific enthusiasm for the study of evolution prompted new goals and directions in the study of living things in their natural environments. The discipline of ecology formed as scientists began to examine the complex dynamics of interacting biological entities in real-world settings. But such examination, these ecologists soon found, could not be done with the instruments of natural history. While the discipline of physiology focused on the direct measurement of biologicalprocesses, it was grounded in the controlled conditions of laboratory practice. To rigorously and directly measure the phenomena of interest to evolutionary theory, ecologists they would have to fabricate their own instruments. In the early twentieth century some ecologists began looking to the discipline of plant physiology for methods to study a plant's interactions with the larger world surrounding it. Bringing plant physiology out into the field required not just a shift in research focus but also fabrication of dedicated instruments for so-called field physiology. This paper examines a case study of a new field instrument designed and fabricated to provide answers to specifically ecological questions about how plants responded to environmental conditions. At the Carnegie Institution of Science's Desert Laboratory, ecologist D. T. MacDougal designed a sophisticated recording and measuring device still rugged enough for the field. MacDougal's collaborator on this project was Godfrey Sykes, a multitalented inventor, explorer, machinist, and engineer who fabricated innovative and unusual field instruments for Carnegie ecologists. MacDougal and Sykes' instrument, termed the dendrograph,was intended for MacDougal's own research, but also mass-produced and sold to other ecologists.

Promoting their own research as well as the instruments that they created, these ecologists pushed the discipline to focus on field experimentation rather than geographical surveys. As well, these innovative instruments show how freeing ecology from its static, descriptive roots required conceptualizing field experiments

focused on the very struggle at the heart of natural selection. These scientist-fabricators complicate recent historical depictions of a tension between field science and lab science, showing instead a history of scientific effort to closely link fieldwork to theory through innovation in instrumentation and measurement.

Thursday, 14 March • 20:00-21:00

Plenary Lecture

Simon Schaffer: *The Silent Trade: on the boundaries of Measurement*

The manifesto for this conference refers to measurement as a boundary object: measurement, it is claimed, cuts across disciplinary and social boundaries. This talk reflects on the roles of measures when boundaries are made and breached, drawing on historical material from cross-cultural encounters, when questions of measurement are peculiarly moot. According to Herodotus, in a story often repeated by early modern European voyagers and traders, travelers to the African coast in search of gold would leave a quantity of goods on the beach, then return to their ships: inland Africans would then approach the beach and match these goods with a given amount of the precious metal. No negotiation and no conversation took place between the two groups, yet perfect honesty in measurement took place on both sides. This tale of immediate, perfect, and silent exchange represents, perhaps, the utopian dimension of measurement, a theme of great interest for much more recent scholars of the history and sociology of measures as effective resources to resolve controversy and conflict. On the other hand, it is also argued that conflict drives the work of measurement: Brecht began 'Mother Courage and her Children' with a scene in which a group of soldiers complain that since a Swedish province has been so long at peace, its resources are unmeasured: "it's only where there's a war that you get proper lists and inventories". Study of measures taken at moments of boundary crossing may have lessons about this relation between irenic and polemic measurement.

Friday, 15 March • 9:00-11:00

Plenary Lectures

Nadine de Courtenay: *Connecting the Actual and the Virtual: The Double Interpretation of the Equations of Physics*

One seldom underlines how much the construction of a shared system of units has contributed to shape the way in which we represent physical relationships mathematically. I will first review the two paths that were followed in order to put physical relations into the form of mathematical equations and point out that they involved quite different conceptions of the role of units. I will go on to examine how these two paths become articulated in the double interpretation of the equations of physics advanced by Maxwell, now implicitly accepted by physicists. I will then turn to indicate how the previous discussion is connected with the forthcoming reform of the international system of units (SI), as well as with the new methods by which the traceability of measurement results is getting established so as to ensure the coordination of scientific, technical, economic and everyday activities.

Luca Mari: *Outline of the Current Status of Measurement Science: From the Point of View of the International Vocabulary of Metrology*

The International Vocabulary of Metrology ‧ Basic and General Concepts and Associated Terms (VIM) (JGCM 200:2012; http://www.bipm.org/en/publications/guides/vim.html) is produced by the Joint Committee for Guides in Metrology (JCGM), which currently gathers eight international organizations working in the field of metrology and aimed at addressing the general metrological needs of science and technology through the development of guidance documents (together with the VIM, the Evaluation of measurement data ‧ Guide to the expression of uncertainty in measurement (GUM) (JCGM 100:2008)).
A comparison of the three editions of the VIM (published in 1984, 1993, 2007 respectively) highlights that measurement science is a moving target, and some of its foundational topics, such as the concepts of quantity, measurement result, and measurand, have significantly changed even in a relatively short time.
With reference to some cogent examples, the lecture will emphasize the current understanding of measurement as a knowledge-based, pragmatic process, in which models play a primary role.

Friday, 15 March • 11:30-13:00

Measuring the Unmeasurable

You cannot compare apples and oranges. Everybody knows this proverb. In a sense, the statement is obviously wrong: If you measure the weight, you get two numbers. Voilà: perfectly comparable. Measuring things makes them comparable. The proverb, however, might be understood in a less trivial sense when apples and oranges stand in for items that cannot be validly compared. What if weight is not the critical feature in question, but rather color, flavor, or texture? That would suggest that measurements are impossible or misleading. Furthermore, the fact that the proverb is as common as it is indicates that this problem occurs quite commonly.
In the proposed symposium, we want to start out from this problem and probe a variety of settings where measuring seems to be impossible - but is done nonetheless. Under what circumstances and for which reasons do things or processes appear unmeasurable? With what strategies can and does one work around the apparent impossibility? More precisely, the three contributions deal with different types of reasons: conceptual inappropriateness - in the case of health care, operational inaccessibility - investigating so-called proxies, and logico-mathematical impossibility - a paradox of unmeasurability.The point is that in all three cases scientific practice has devised strategies that allow to work around the obstacles and to conduct measurements. However, these strategies also raise questions about the limitations of their validity.

Leah McClimans: *Can We Measure Quality of Life? You Decide*

In a recent conference on quality of life research, one paper entitled, 'Busting the Top Myths About Quality of Life Assessment in Clinical Practice' suggested that one of these myths is that 'quality of life' is a philosophical concept and thus not measurable. Wrong! The author argued. Quality of life can (and has!) been operationalized and validated in many, many measures. The author is right. Quality of life has been operationalized and 'validated' in hundreds (thousands!) of measures. Yet quality of life researchers are still haunted, perhaps rightly, with the worry that this construct cannot be measured. This worry leads scientists to improve their methods and currently there is an increasingly vocal minority advocating a move from traditional measures based on classical test theory to measures developed via RASCH analysis.
In this paper I begin by comparing philosophical definitions of quality of life with some definitions from the social sciences in order to highlight some differences in our understanding of this construct, differences that may affect our ability operationalize and thus measure it. I then move on to discuss different kinds of

measures, e.g. profile, single item and index measures as well as different ways to obtain a score, e.g. classical test theory and RASCH analysis. I end by discussing 1) the ways in which classical test theory and RASCH enable the measurement of quality of life and 2) ask whether or not these methods can lead to valid measures. In answering this second question I focus attention on one recurrent problem in quality of life research: response shift.

Ann Johnson: *Measuring Complex Phenomena: The Problem with Proxies*

Many of the phenomena that scientists would like to measure are inaccessible for a variety of reasons. Many of the phenomena or characteristics social scientists are interested in are too complex to measure, say intelligence, productivity, or creativity. One strategy to deal with phenomena that cannot be directly measured is to design a proxy that can be measured and use that in place of the phenomena. This paper examines the issue of proxies in social scientific measurement and focuses on when they break. A broken proxy is one that decouples from the phenomena; it no longer tracks with the phenomenon it is supposed to stand in for. Proxies break for a wide variety of reasons, but this paper focuses on proxies that are taken up by policy-makers and incentivized. For example, in the US colleges are often measured and ranked by the quality of their students; one proxy for student promise is the Scholastic Aptitude Test or SAT. When designed, the SAT tracked reasonably well with the students' success in their first year of college; in fact it tracked better than high school grades because students who went to poorer schools and were less well-prepared did worse on the test which tracked better with their college performance. But once colleges were rewarded for accepting students with higher SATs, they prioritized increasing the proxy (and not necessarily increasing the phenomenon the proxy stood in for). Since the 1990s most colleges have seen a marked increase in the SAT scores of their students, yet the first-year failure and drop-out rates have remained steady. Few teachers feel that the overall collective quality of their students is rising. Is the SAT a broken proxy? This paper takes up that question. It also examines the proxy of numbers of publications for faculty productivity and the use of impact factors to measure the dissemination of scientific ideas. Both of these are candidates for broken proxies due to their use as incentives for both colleges and individual faculty members. Do we need some awareness that proxies are not robust once they are incentivized—this paper argues that we do. Perhaps the reason physical science proxies appear more robust is simply that they are not used in policy?

Johannes Lenhard: *When a Pea is Equal to the Sun: The Mathematics of the Unmeasurable*

In dead earnest I want to attempt to describe, explain, and interpret a mathematical theorem. Not an arbitrary exemplar, but the most astonishing result of theoretical mathematics. It is the celebrated theorem of Stefan Banach and Alfred Tarski, published in 1924. It deals with set-theory and geometry and states that one can cut a (massive, three-dimensional) ball into finitely many pieces, re-arrange them, and put these pieces together again to get two massive balls of the same size as the original ball. One illustration of the theorem is as follows: One can start this mathematical operation with a pea to get two peas of the same size, then repeat the operation finitely often until one gets objects as big as the sun.
Of course, this kind of operation strikingly contradicts all intuitions of how measurement works. And in fact, the operation crucially employs sets - how the ball is decomposed - that are not "measurable", where being measurable is defined in a rigid mathematical way.
Ok, one might think, this is not the only cumbersome result in mathematics. Why should an audience interested in philosophy and measurement take care of set-theoretic geometry? The point is: The mathematical notion of measure relevant here can be seen as something like the least common denominator of all extant notions of measures in practical scientific use. And if there is no measure for how big finite sets are - in the sense that some exemplars are non-measurable - what kinds of things could be measured at all?
This talk will describe the result, analyze the conditions under which it holds, and discuss its ramifications. The lesson I argue for is that basic notions of scientific practice might offer surprises. Philosophically, it is important that we get this insight from mathematics.

Conceptual Issues

Flavia Padovani: *A Logical Space for Measurement*

Scientific theories typically consist of principles, laws, and equations, which include already sufficiently interpreted terms like velocity, pressure, time, temperature, etc. In most cases, in the history of science, their individuation as parameters is parallel to the creation of the corresponding measurement procedures and the development of the theory in which they occur. In line with van Fraassen's recent (Wittgensteinian-influenced) proposal, I will consider measuring as representing, or "locating an item in a logical space […] of possible measurement outcomes" (2008, p. 164). Without this space of pre-ordered possibilities, there can be no objects of representation for us. In this sense, measuring means framing, i.e., "constituting"

the measured quantities, thus allowing for the coordination of abstract, mathematical quantities to "pieces of reality".
Originally reinterpreting Ernst Cassirer's proposal, Hans Reichenbach was among the first to seriously tackle the issue of coordination, putting forward an account of constitutive principles of science in which the constitutive function of measurement was pivotal. In the past two decades, one of the leading interpretations of such principles has been proposed by Michael Friedman. However, in Friedman's account the constitutive function of measurement is completely neglected whereas, in my view, this function is crucial. The aim of this paper is to suggest an interpretation of the practice of measuring that combines the concept of "logical space" with Friedman's notion of "relativized a priori". This allows for a more dynamic and pragmatic account of the principles operating in the scientific practice.

Fabien Gregis: *Can we dispense with the notion of "true value" in metrology?*

Metrology is currently encountering an "epistemic turn", characterized by a focus on practical issues related to the application of measurement results. This turn comes with a shift in the interpretation of probability, from a frequentist view to a Bayesian one. Additionally, it is accompanied by a dismissal of the notion of the "true value" of a quantity.
"True value", as well as "error", have been classical concepts in measurement since the beginning of "theory error" in the XVIIIth century but are now dismissed as unobservable ideals, and thus dispensable for scientific practice. My presentation will investigate the features of the "epistemic turn" in metrology so as to interrogate the need for such a concept as "true value". My analysis will illustrate the way in which metaphysical and epistemological issues relate to each other within measurement science, particularly regarding the question of scientific realism.
To this end, I will make use of scientific material in the fields of metrology and history of metrology, in addition to philosophical insights about the foundations of measurement.

Christian Hennig: *Measurement as Constructive Act - a Statistician's View*

Processing measurements is central in statistics, but most statisticians take the measurements they work with as a given. The theory of measurement is treated only marginally in the statistical literature. This is all the more surprisinggiven that

part of the work of statisticians is to take part in the construction of measurements, such as national indexes of well-being, but also measurements of quality for the comparison of statistical methodology.
In this presentation I will sketch a concept of measurement as a constructive act with the aim to unify and communicate the perception of phenomena in a transparent way, which is of central importance to science. This perspective deviates strongly from the naive idea, held by many statisticians and implied by much of statistical modeling, that a measurement reflects an unobservable underlying truth of which it deviates by a "measurement error", which is often treated as quantifiable.
This view has implications on a number of issues meaningful to statisticians, namely the concept of "scale types" (nominal, ordinal, interval, ratio; Stevens, 1946), derived from representational measurement theory, and the idea that the scale type of a measurement restricts the statistical method that can be applied, the modeling of measurement errors, principles for the construction of indexes, and the concepts of validity and reliability of measurement.

Measurement and Engineering

Mieke Boon: *The Measurement of Properties in the Engineering Sciences*

The philosophy of science usually focuses on the epistemic role of measurements, in particular, in testing theories. Against this background, the theory-ladeness of empirical knowledge produced in measurements has been one of the major concerns.
From the perspective of the engineering sciences, some other philosophical issues are raised. Measurements provide qualitative and quantitative information about properties of the target system under study. However, most properties only become manifest by means of specific technological instruments and procedures. What is more, information on properties results from interactions between the measuring instrument and aspects of a target system. The more technological instruments we invent, the more properties and processes become manifest. We only have to take a look at the famous Handbook of Chemistry and Physics to see how many properties of materials have been established. Every physicist knows that the manifestation of these properties is dependent of the technological procedure for measuring them. Therefore, the description of such properties usually has the character of an operational definition, which means that it encompasses aspects of the measuring instrument and procedure.
In the engineering sciences we are interested in properties and processes for their roles in the functioning of technological artifacts. Put differently, the functioning or

dysfunctioning of a technological artifact is usually conceived in terms of properties or processes that determine its functioning. Also, conceiving of possible improvements of a technological function, or even, creating new functions often is in terms of properties or processes.
A simple example is scientific research for the development of paint. The technological function(s) of paint include qualities such as protecting a surface, workability in its application, durability and esthetic qualities. The manifestations of these technological functions involve perceivable and/or quantifiable properties of paint such as its color, its viscosity, and its fastness of drying, its adherence to a surface, its smoothness, its shininess, its hardness, and the stability of these properties. Hence, these are the properties that manifest (or display) the technological function. Examples of manifestations of technological dysfunctions of paint are properties and processes such as the tendency to maintain ripples, the increase of viscosity when applied at higher temperatures, the tendency to capture air-bubbles, the toxicity of the solvent, formation of cracks in hardened paint, and the tendency to turn yellowish under the influence of sun-light. Hence, for a technological artifact to perform its technological function(s), we aim at producing the properties or processes that are manifestations of its proper functioning, and prevent or change the occurrence of those that are manifestations of its improper functioning.
In this talk, I will explore how (the development of) instruments and procedures for measuring (and manifesting) properties and processes of a target-system is related to the development (production, improvement, etc) of technological functions.

Sjoerd D. Zwart: *The Foundations of Indirect Measurement and Model Laws*

The aim of this paper is to investigate the philosophical foundation of indirect measurement and 'scientific' scaling. Indirect measurement, here, refers to measuring the properties of scale models on a different scale than the target system with the purpose to find out the corresponding values of these properties for the target system of the real dimension. For instance, the drag of scale models of two or three meters of length are measured in towing tank laboratories and scaled up using model laws to find out the drag of a vessel of the real world dimensions. Traditionally in engineering curricula, the foundation of indirect measurement and the application of model laws are taken to be dimensional analysis. In this paper I argue that the empirical hypothesis that some physical property is ratio-scale measurable lays the foundations of indirect measurement and 'scientific' scaling rather than dimensional analysis. Regarding the empirical content of the geometrical, the kinematic, the dynamic, and even Froude's scaling laws I conclude the following. The scaling laws derive all their content deductively

from classical mechanics, Euclidean geometry and the ratio-scale measurability assumption. Consequently, within the context of classical mechanics these laws cannot be false, and have therefore an analytic character. I will show that if all ceteris paribus conditions of a scaling law hold, the antecedent of the law deductively implies the consequent. Finally, I will show that this conclusion holds mutatis mutandis for any other law of similitude.

Pablo Schyfter: *Measuring Promoters to Make the Field: Metrology as a Tool for Building Knowledge, Building Artifacts, and Building Fields*

Synthetic biology is a nascent field of biological science and technology. Its practitioners currently include researchers from sciences such as molecular biology and engineering fields such as control engineering and informatics. Many of those engaged in 'synbio' are committed to making of the field a discipline of 'real' biological engineering. That is, they hope to make of synthetic biology a discipline that practices knowledge-making and artifact construction as do existing disciplines of engineering. So-called genetic 'engineering,' these individuals claim, is nothing of the sort, and synthetic biology will succeed where this previous attempt at engineering with the living world has failed. Thus far, practitioners have focused on the use of so-called 'engineering principles'—techniques such as standardisation and functional compartmentalisation. Recently, greater attention has been dedicated to metrology. The engineering faction within synthetic biology hopes to take existing 'standardised' parts and characterise them through systematic processes of measurement. Many in the field now want to measure like engineers measure. The metrological tools and techniques under study are intended to enable 'rational' and 'predictive' design of biological technologies—they are intended to enable what is taken to be 'true' engineering design.

My talk presents an epistemological analysis of synthetic biology's attempts to systematically measure transcriptional promoter 'strength'. Transcriptional promoters are genetic elements that initiate the 'reading' of DNA and the 'writing' of RNA— effectively, they serve to activate genes. In synthetic biology, promoters are employed as ON/OFF switches for artificial genetic constructs. Drawing lessons from Walter Vincenti's historical studies of aeronautics, I discuss contemporary measurement practices in synthetic biology. Specifically, I discuss how researchers have developed a 'rough and ready' empirical measure of relative promoter 'strength', one similar to the empirical testing of air-propeller thrust carried out by W.F. Durand and E.P. Lesley in the period 1916-1926. Just as with Vincenti's example, metrological work in synthetic biology demonstrates the fundamental links between knowledge-making and the design and fabrication of engineered artifacts. I contribute to Vincenti's observations a concern for discipline formation.

I discuss the practices by which metrology is used to produce 'valid' engineering knowledge, define the character of biological objects, and justify the status of synthetic biology as a 'true' discipline of engineering. That is, metrology in synthetic biology has epistemic, ontological, and disciplinary uses. Knowledge-making, object-making, and discipline-making are practices linked by way of acts and tools of measurement. My analysis is based on extensive sociological fieldwork of synthetic biology laboratories in the United States, a comprehensive review of the technical literature, and an epistemological analysis of the kind undertaken by Vincenti.

Friday, 15 March • 14:00-16:00

The Politics of Measurement

Anne Harrington: *Nuclear Values*

How do we measure the value of a nuclear weapon? Disagreement over how many weapons are necessary to maintain a credible nuclear deterrent has fueled debate among deterrence advocates for more than a half a century. At the level of policy this debate plays out as a disagreement over the number of nuclear weapons in the US nuclear arsenal and the size of their yield. However, at the level of theory, the disagreement is a more fundamental difference of opinion over the nature of the relationship between military means and political ends. It is a disagreement over the limits of what can be accomplished through violence. Some argue that there is a point at which possessing relatively more weapons will not confer any additional advantage in a crisis because there is no longer any measurable material effect. Others argue that numerical superiority is always a significant factor because regardless of the military value there is a measurable political effect. In this paper I trace the evolution of the measurement of a nuclear weapon's value from a basis in its physical effects to an attempt to quantify its political value, arguing that the evolving language of measurement and value sustains the current normative order against challenges that attempt to "devalue" nuclear weapons.

Carlo Martini: *"Subjective" Measurement in Economics, but not Arbitrary*

The debate on whether economics is a science, on a par with the natural sciences, and with the same level of objectivity, has often been centered around the problem of a lack of consensus over economic core principles and methods. Of course, among the core tenets of a science are its principles and methodology of measurement. Measurement in economics is an extremely complicated matter,

because defining economic quantities is seen as both an empirical and a social (or conventional) process.

Let's take one example: the measurement of ``wealth'' in the best Smithian tradition. On the one hand, we would like to measure wealth; not just for fun, but, for instance, to judge our politicians, or decide whether socialism is better than capitalism at satisfying a nation's needs. But there is extremely little consensus as to what wealth is, and that is true even if we forget about trying to measure wealth as ``happiness'', and focus on material wealth only, say, GDP. The US Bureau of Economic Analysis, in its definition of the ``production boundary'' (that is, what gets and what doesn't get measured), states that ``certain household activities -- such as housework, do-it-yourself projects and care of family members -- are excluded, partly because by nature these activities tend to be self-contained and have limited impact on the rest of the economy [...]''. But that is clearly an understatement: Numerous ``toy-examples'' can shows that the Bureau of Economic Analysis is understating the ``impact on the rest of the economy.''

Imagine two households (the Smiths and the Trumps) with two members each (this time, for political correctness, we'll call them X and Y). In both the Smiths and the Trumps, X earns 50 euros, which they use for mortgage and vacations. In the Smiths Y earns no income, and takes care of the children, as well as cooking and cleaning. In the Trumps, instead, Y has a paid job and earns 50 euros, but doesn't have time to take care of the children and of other household activities. Thus, in the Trumps, Y pays for children day care, meals catering, and house cleaners. The Smiths and the Trumps, intuitively, produce about the same amount of total wealth, even discounting for the extra amount of quality that the Trumps may gain from outsourcing most household activities to ``professionals''. For the purposes of GDP calculation, however, the Trumps produce twice as much as the Smiths. If the foregoing example were to be used in the context of whole populations, we could imagine a traditional society with a very large number of Smiths-like households, next to a much different, and by some criteria more advanced, society that has predominantly Trumps-like households. It should be obvious that the bias resulting from the method of calculation of the BEA has, quite definitely, not a limited impact, but rather a large one.

In fact, the problems with GDP-calculation, cost-of-life index, etc. run deep into the discussion of whether economic measures are empirically grounded or just conventional. Reiss writes that ``the answer is, in my view, that this question [how to measure economic quantities] should be settled in the exact same way as other questions regarding the aims of a society are settled. Societies must make up their minds about such decisions all the time: What's the age of consent? Should there be capital punishment? Should we allow smoking in bars? [...] From the point of view of the methodology of economics it is important to see that such a question cannot be addressed by scientific means alone''.

In this paper I develop on the work of Reiss (2008) and others, and show that much of the consensus on foundational economic matters (in this case the focus will be on measurement) should not be sought on supposed objective grounds, but rather as subjective agreement, where the subjects of economics are the ``economic experts''. But `subjective' is at times synonym for `arbitrary'; therefore, the challenge I will undertake is to develop a concept of subjective agreement that is a rational one, and that is grounded on the principles of objectivity and replicability. Such a project is well under way in engineering, where subjective expert judgment is an essential methodological tool, and other sciences, but is only moving its first steps in economics and the social sciences.

Sharon Crasnow: *The Measure of Democracy: Coding in Political Science*

Challenges for coding, a way of measuring latent concepts in political science, include questions of conceptualization, operationalization, and aggregation (Munck 2009). In turn, these concerns each generate theoretical, observational, methodological, and pragmatic questions. The different ways in which these concerns about coding are resolved can affect which inferences are legitimately supported by resultant datasets. Those generated for one knowledge project may not support conclusions for another project and so are not productively transported to another context. Consequently, different goals may require different datasets and so the “correct” coding of democracy (or other latent concepts) or constructing a standardized measure as in the case of physical concepts may not possible. Whereas a convergence on purpose played a role in standardizing the measurement of temperature (Chang 2004), it is not clear that convergence is possible in the case of measuring concepts like democracy. For example, it may well be that the observational features that are suited for consideration of causal hypotheses about the democratic peace hypothesis are not appropriate for understanding mechanisms of democratic transitions. Differences in coding have implications for the testing of hypotheses. Minimally, this suggests that coding practices should be highly transparent since testing hypotheses is also testing of the dataset – generating a circularity of sorts. I argue that this circularity is benign and necessary and that Hasok Chang’s notion of epistemic iteration can help support an argument that measurement (of democracy) in political science should be goal specific and pluralistic.

Cheryce von Xylander: *The Metrics of Tact*

This paper probes the limits of managerial management in relation to process engineering and explores why quality control defies measuring techniques. The aim is to illustrate, both historically and systematically, conceptual tensions between management theory, on the one hand, and measurement practice, on the other. Topics presented in this paper revolve around the notion of Taktzeit, a German term with numerous mutually reinforcing connotations. First, it refers to the beat of music, to musical measure, as interpreted by a performer or conductor. Here the emphasis is on interpretative specificity and varying a given beat and deviating from a fixed norm. Second, Taktzeit names the periodicity of time tables in public transport systems, especially in relation to coordinating train schedules. In this context, synchronization of related events matters. Third, Taktzeit operates in the background of modern communications networks as an incremental measure of duration for calculating usage fees. This application of the term describes an ordering of time units that is entirely automated and rigidly consistent. In the postwar period, the specific range of meanings associated with the German Takt enters the global arena: In English, "Takt Time" designates an industrial planning tool used to effect lean management approaches and just-in-time delivery services worldwide. By exploring various phases and applications of the notion of Takt in applied engineering settings across cultural contexts, this paper investigates both how and why this German term with its dual resonances of "socially appropriate behavior" (in English "tact") and "rigid temporal ordering" (in English "beat") has entered corporate management speak. A technoscientific deconstruction of Taktzeit reveals that the coordination of industrial labor relies on notions of efficiency and rationalization whose measurement implies a process of social construction whereby regulatory measures express contingencies brought about by managerial interventions rendered invisible in a controlled environment.

Standards in Context

Shaul Katzir: *Frequency and Time Standards from Acoustics to Radio - the Road to the Quartz Clock*

The rapid development of electromagnetic communication during WWI and the 1920s, both in the techniques employed and in their use and spread suggested a need for increasingly more accurate standards for increasingly higher radio frequencies. Already during the war, physicists and electrical engineers in France and Britain combined the tuning fork, a well-studied measurement instrument in acoustics, in a new valve-electronic circuit, making it a basis for standards of higher frequencies for the new needs of radio. For accurate measurements an electrically

driven but mechanically vibrating tuning fork had a few advantages on purely electromagnetic system. On top of its high stability, it offered an interconnection to the more established and highly accurate standards of time. Vibrating at an audible period of about 1,000 Hz, the tuning fork could drive a clock mechanism and so be calibrated and checked by a pendulum clock. This combination seemed promising to national research institutes in Britain, France and the USA, as means for standardizing wireless communication, to coordinate the growing number of its users. It also deemed useful for the internal needs of integrating the various communication channels used in the vast communication network of American Telephone and Telegraph Company.

The tuning fork, however, was impractical for routine measurements of radio frequencies. Novel findings on the electric effect of the quartz resonator and its stability suggested its employment as accurate standards of radio frequency, first to its discovered and developer of methods for its practical use, Walter Cady, and then to a growing circle of physicists and engineers. Based on the piezoelectric effect, the quartz oscillator showed a high stability and narrow band of frequency, making it a preferable frequency standard. Moreover, it could be easily carried from one place to another, suggesting both a way of distributing the value of the primary standard to small engineering laboratories, and a means for linking national standards into a united global one. Vibrating tens of thousands times a second it did not offer, however, a direct method for comparing its frequency to a standard clock.

Seeking a highly accurate and continuously available standard of radio frequencies, researchers at Bell Labs and at the British National Physical Laboratory reached, circa 1927, methods for reducing the high frequencies to a known lower one, allowing the oscillating quartz to drive a clock mechanism. This quartz clock combined the standards of frequency and time, on their roots in electromagnetism, acoustics and astronomy. Other researchers, however, reached similar methods by different aims. A research at MIT, sought a simple and inexpensive means to furnish small laboratories with reliable frequency standards. A researcher at Philips led to the method following his theoretical study of oscillations in electronic tube. The accomplishments of these and further researchers resulted in an electric communication network of higher coordination, and in exacter clock on a novel principle, which later would contribute to the popularization of precision.

Sharon Ku and Frederick Klaessig: *A Matter of Size Does Not Matter: The Social Construction and Application of Gold Nanoparticle Standard*

This paper challenges the ideology of exact measurement and atom-by-atom manipulation in nanotechnology, by examining the ontology and social

construction of a nanosize standard. Our analysis is based on a longitudinal study of the Gold Nanoparticle Reference Material (Gold RM) certified by the US National Institute of Standards and technology (NIST). Tracing the biographies of the Gold RM—from the original size debate in nanobio communities, the production within NIST, to its application for the inter-laboratory studies in American Society of Testing Material (ASTM), we demonstrate that atomic precision cannot travel from angstrom physics to nanotechnology, producing the "exact" measurement to harmonize nanosize characterization. Instead of being a "harmonizing device", the Gold RM functioned as an "irritating object" at each step of its life-time to trigger confrontations among different research paradigms in nanotechnology communities. We argue that these irritations constitute the ontology of standardization: The Gold RM acts as a proxy to perturb the superficial harmony under partnership building, generating "differences" embedded in heterogeneous nanotechnology communities while concurrently offering organizational platforms that allows consensus over the "tolerance of differences" to be negotiated.
Gold RM's dual identity as a controversial scientific subject at the research frontier, and a mundane calibration device for coordinating routine laboratory practices, raises theoretical challenges of reconstructing biographies of scientific objects for technoscience standardization. Our empirical data suggest that an integrated framework that combines: 1) ontological indifference between representing and intervening; 2) dynamic co-production of material/organizational structure, is essential to capture the coming-into-being of technoscientific standard objects.

Lara Huber: *Measuring by which Standard? Plurality Challenges Epistemic Singularity*

How to access the validity of a measuring device without referring to the authority of a widely accepted and scientific approved standard? As regards strategies of measurement in particular, as much as issues of comparing scientific data in general, both endeavours are highly connected to the normative power of standards (i.e. scales of measurement). The agreement on international units of measurement in 1960 is commonly regarded as a transcendental means not only to reduce transaction costs but also to ensure scientific progress as such.
Given their genuine authority, standards are often identified with the very epistemic goals that are tended to be achieved referring to them (i.e. accuracy). By introducing the epistemic singularity of a standard I like to stress its specific rank as a scientific norm. Due to technical innovation and scientific progress standards remain object of improvement or even displacement. New standards are regularly introduced if they have proved to be more potent with regard to the very epistemic goals of a scientific endeavour, such as the measurement of temperature. And, what is more, the epistemic rank of standards might differ significantly. For

example, as the paper highlights, in the case of reference data (i.e. reference genome, reference brain) epistemic singularity regularly is aspired but not achieved: In the Life Sciences different standards could be consulted to tackle a given goal of measurement or comparison respectively. Scientists are confronted with a plurality of standards (i.e. different standardized formats of processing and representing data). As a matter of course, reference data exhibit a quite limited authority with regard to the epistemic goals that are associated with standardized formats (i.e. representativeness).

Wolfgang Pietsch: *A Revolution Without Tooth and Claw: Redefining the Physical Base Units*

A case study is presented of a recent proposal by major metrology institutes to redefine four of the physical base units, namely kilogram, ampere, mole, and kelvin. The episode shows a number of features that are unusual for progress in an objective science: for example, the development is not triggered by experimental discoveries or theoretical innovations; also, the new definitions are eventually implemented by means of a voting process. In the philosophical analysis, I will first argue that the episode provides considerable evidence for confirmation holism; second, that the episode satisfies many of the criteria which Kuhn requires for scientific revolutions even though one would naturally classify it as normal science. These two observations are interrelated since holism can provide within normal science a possible source of future revolutionary periods.

Human Dimensions

Donna J. Drucker: *The Klein Sexual Orientation Grid and the Fluid Measurement of Sexual Identity*

Like the Kinsey scale, the Klein Sexual Orientation Grid has played an important role in sexual identity discovery for many Internet users. The Kinsey scale, first published in Alfred Kinsey's Sexual Behavior in the Human Male in 1948, presented readers with a scale from 0 to 6 (strictly heterosexual to strictly homosexual), with an "X" category off the grid for those with no sexual tendencies. The Kinsey scale showed that many of Kinsey's 18,000 interviewees had sexual feelings, desires, and behaviors that placed them in between the standard identity categories of "heterosexual" and "homosexual." First published in 1978 in The Bisexual Option, the Klein Sexual Orientation Grid was author Fritz Klein's attempt to improve the Kinsey scale by giving it more specificity. The Klein Grid has users rate seven personal factors on a scale of 1 (heterosexual) to 7 (homosexual): sexual attraction,

sexual behavior, sexual fantasies, emotional preference, social preferences, heterosexual/homosexual lifestyle, and self identification. Users rate these factors in three columns: past, present, and ideal. A mathematical combination of these factors gives the user an immediate score from 0.00 to 6.00, equivalent to a point on the Kinsey scale. Klein himself intended that the scale be used as part of an individual's sexual identification process, and specifically as part of the bisexual "coming out" process. However, mathematical analysis of the Klein Grid showed that it was no more or less accurate as an identity classification tool than the non-interactive Kinsey scale (Weinrich et al., "A Factor Analysis of the Klein Sexual Orientation Grid in Two Disparate Samples," Archives of Sexual Behavior, 22 [1993], 157-68).

The Klein Grid has proven interesting to Internet users interested in exploring their sexuality. Unlike the Kinsey scale-- which users simply study in order to find their place on the scale-- users can take the Klein Grid in an online quiz form and get an instant number. The Klein Grid is especially popular among groups for queer and questioning young people, and GLBT youth service organizations often encourage their participants to take the quiz to help them sort through their thoughts and feelings. This paper undertakes a qualitative textual analysis of ten individual bloggers and ten online communities, for both youths and adults, that have used the Klein Grid to understand their sexual identities. Some users prefer the Kinsey scale over the Klein Grid, some think both tools are a waste of time, and others think that the Klein Grid is good for giving a snapshot-- but far from the whole picture-- of an individual's sexual identity. Those who like the Klein Grid especially like the way that it breaks out behaviors, desires, and identities over time, and allows them to take their ideal sexual life into consideration. However, it is clear from even the best reviews of the Klein Grid that it is inadequate as a measure for a person's whole sexual life. Simple measures of human sexuality illuminate some aspects of a person's life, but remain inadequate for representing its complexity.

Erik Angner: *The Problem with Happiness Measurement*

Subjective measures of well-being are measures of well-being based on questions such as: “Taking things all together, how would you say things are these days – would you say you’re very happy, pretty happy, or not too happy these days?” Using the term “subjective well-being” to denote that which they are supposed to represent, subjective measures of well-being are frequently referred to as measures of subjective well-being. Subjective measures are frequently presented as substitutes for, or complements to, more widely used economic welfare indicators like Gross National Product. Subjective measures have been subjected to harsh criticism from philosophers, psychologists, and economists alike. In this paper, I argue that most of these criticisms are based either on a misunderstanding

of the nature of subjective measures or on a misunderstanding about the nature of measurement. One problem associated with the ubiquity of bad arguments is that they may have obscured other, very real problems associated with happiness measurement. Chief among them, I will suggest, is the problem of reification. In the literature on psychometric measurement this is a well-known problem to which happiness researchers are obviously not immune, and its effects can be pernicious. I conclude that the central problem associated with the measurement of happiness as a subjectively experienced mental state is not that it is *too hard* to measure, but rather that it is *too easy* to measure.

Zed Adams: *Man a Measure?*

The problem of color realism is standardly formulated in terms of whether man is the measure of what it is for something to be colored. Realism is taken to be the position that although humans might be in a privileged position to know which things are blue, what it is for something to be blue does not constitutively depend upon human experience; and anti-realism is taken to be the position that what it is for something to be blue does constitutively depend upon human experience.
In "Objectivity and Modern Idealism" (1994), Gideon Rosen argues that the standard formulation of this problem rests upon a false presupposition: namely, that the reality of some type of thing is impugned if it turns out that man is the measure of what it is to be that type of thing. Rosen argues against this presupposition on the grounds that it rests upon transcendental conception of subjectivity, according to which experience is outside the natural, physical world.
In this paper, I criticize Rosen's argument, but not by defending either the standard formulation of realism/anti-realism debates or a transcendental conception of subjectivity. Rather, I focus on an aspect of Rosen's argument that he does not himself explore: namely, what, exactly, would it be for man to be the measure of something? I explore this question as it relates to color, with a specific emphasis on the history of the development of the 1931 CIE color space. I argue that this history reveals the ontological priority of standardized forms of measurement and calibrated measuring devices. In short, Rosen's argument overlooks the way in which inter- and intra-personal variation in color experience implies there is a very real sense in which such experiences do not measure anything, because there is no unity to the putative units of measurement involved.

Hans Radder: *Benjamin Libet's Measurement of Freely Willed Decisions: A Critical Analysis*

In the extensive, recent debates on free will, the pioneering experiments by Benjamin Libet continue to play a significant role. It is often claimed that these experiments demonstrate the illusoriness of freely willed actions. In this paper, we provide a detailed analysis and evaluation of Libet's experiments from a philosophy of science perspective. Our analysis focuses on Libet's central notion of the "initiation" of freely willed processes by the brain. We examine four interpretations of the notion of initiation: in terms of a cause, a necessary condition, a correlation, and a regular succession. We argue that none of these four interpretations can be supported by the design and results of Libet's experiments.
The paper draws on:
Hans Radder and Gerben Meynen, 'Does the brain "initiate" freely willed processes? A philosophy of science critique of Libet-type experiments and their interpretation', Theory and psychology, 23 (1), 2013, 3-21; freely available at http://tap.sagepub.com/content/23/1.toc.

Friday, 15 March • 16:30-18:30

Plenary Lectures

Laura Dassow Walls: *With Compass, Chains, and Sounding Line: Taking Thoreau's Measure*

"The human imagination is released by fact."
--Stanley Cavell, Senses of Walden

"Measure" most clearly differentiates Thoreau from his mentor and rival, Emerson. Where Emerson seeks correspondences, hence analogies, between the mind and the material world, Thoreau seeks to make mind and matter "commensurate" through the practices and concepts of measurement. His youthful speculation that strength comes when "the body marches to the measure of the soul" ultimately became one of his most famous lines: "If a man does not keep pace with his companions, perhaps it is because he hears a different drummer. Let him step to the music which he hears, however measured or far away." Along this dimension, "measure" commensurates body and mind by the practices of poetry and music: the measured line, the deliberate or measured step, bodies literally thinking through measures. But the aging Thoreau also literalized this metaphor via the scientific practices of measurement, a tendency long dismissed as a scientific pathology symptomatic of a waning poetic imagination. On the contrary, Thoreau

was modeling his measuring practices on those of Alexander von Humboldt, to identify patterns—conceptual, temporal, and spatial—otherwise invisible. The best-known instance is Thoreau's survey, "with compass, chains, and sounding line," of Walden Pond, which was the inception of both his career as a professional surveyor and also, not coincidentally, of his most famous book, Walden, a demonstration project in the commensuration of nature, soul, and society. As he wrote, a lake—Walden—"is earth's eye; looking into which the beholder measures the depth of his own nature." Walden's governing figure of "living deliberately" is another metaphorization of measurement, that is, to weigh justly, as in Libra's scales. A third dimension concerns nature's self-registration through a range of writing practices, from foxes leaving tracks in the snow to grasses tracing arcs in the sand, which, with the intermediation of such devices as sundials, drums, and water wheels, becomes a further figure for recording inscriptions both human and natural. The upshot of these several epistemological sequences was the transformation of measure from an absolute standard into modes of relationality, whether perspectival, as in differential views of an object-in-common; scalar, as in fractal movements up and down scale levels; or temporal, as a way to make changes visible—for instance, as Thoreau's measures are being used by ecologists today to make climate change visible. Measurement turns out not to delimit boundaries but to open them up, such that one can indeed step not to one overruling "drummer" but to the music which one hears, however measured, and however far away.

Martin Kusch: *"A Branch of Human Natural History": Wittgenstein and Metrology*

This paper focuses on two interrelated topics: (a) Wittgenstein's use of metrological ideas as models or analogues for certain phenomena or issues in the philosophy of language, mathematics, colour, or epistemology; and (b) Wittgenstein's own rudimentary philosophy of metrology. The central elements of (b) are the relationship between metrology and the rule-following considerations, metrological relativism, and the problem of coordination. Some of these elements will be related to Bas van Fraassen's recent work.

Saturday, 16 March • 9:00-11:00

Plenary Lectures

Mary Morgan: *Accounting Reasons - Reasoning Accounts*

Accounting forms one of the most widespread systems of measurements - not only homely and businesslike, but also a craft of the state. Yet, despite its dull exterior, accounting is not just a system of measurement. National accounting, for example, not only defines the health and wealth of an economy, but provides means of reasoning about an economy, and bounds the policy interventions which may take place in that society. This powerful combination of roles - representing, reasoning and intervening - plus the contingencies of history, lie behind the extensive ways in which this system of quantification has become "performative" in modern economic life.

Marcel Boumans: *Clinical Measurement*

Measurement is the assignment of numerals to a property of objects or events – according to a rule. This paper aims at discussing solutions to the following deceivingly simple looking problem: What are the rules that make that these numerals provide reliable information about these objects or events? The solutions that will be discussed are the ones that are developed in the 'field sciences'. A field science studies phenomena that cannot for practical, technical or ethical reasons be studied in a laboratory, which means that these phenomena cannot be isolated from their environment and cannot be investigated by manipulation or intervention.

The reason for the addition of the adjective 'clinical' to measurement is because of the ambiguous meaning of the term 'clinical'. On the one hand it means "coldly detached and dispassionate", so pertains to unemotional, scientific, objective, analytic, impersonal, clean, disinterested and emotionless. On the other hand, it pertains to the bedside, and pertains to or is founded on "actual observation and treatment of patients, as distinguished from theoretical or experimental". This second meaning becomes even more significant when taken as adjective to 'judgment', that is, in 'clinical judgment': "the critical decisions made on the basis of scientific observations but with the added skill provided by long experience of similar cases". In current debates in relation to Evidence-Based Medicine (EBM), clinical (as an adjective to judgment) even receives a meaning opposite to the first meaning above: biased, preconceived, prejudiced and subjective.

This dual meaning of 'clinical' illustrates nicely the problems that are dealt with in this paper. The subject under investigation is what I call 'measurement outside the laboratory'; a label which denotes measurement of phenomena when one cannot

systematically intervene on the measurand and control its environment. When the measurand is simple, usually there are no additional problems compared to measurement in a laboratory. But most phenomena are not simple, an almost infinite amount of factors are involved of which not all can be measured or are even known. Moreover, whatever measurement procedure is developed, it always has to assume similar baseline characteristics, that is, it always has to abstract from idiosyncratic circumstances, which have repercussions for the reliability of the measurement of a field phenomenon.
It is generally acknowledged that 'data underdetermine theory', but this yields in the same vein for method and procedure. Data will not tell you how they should be treated. The choice of the 'calculus of observations' is based on the assumed level of reliability each treatment can provide, but reliability has different facets – precision, accuracy, standardization, stability, certainty, unbiasedness – that usually conflict with each other. These different facets express different epistemological values. Someone has to make a choice; there exists no 'golden rule' that without intervening human judgment can be applied mechanically. Measurement is the combination of mechanical procedures and human judgment, or in other words, measurement is a combination of objective procedures and subjective judgments. This duality is captured by the ambiguous term 'clinical'. Hence, science outside the laboratory requires clinical measurement, measurement that combines mechanical measurement procedures and clinical judgments.

Saturday, 16 March • 11:30-13:00

Measurement of Mind

Andreas Kaminski: *Measuring Intelligence and Temperature: Is it the Same?*

The debate about psychometrics is dominated by the question of its epistemological possibility. The focus is on the question: Is the relation of measurement and subject (mind) appropriate? This debate is important, but it also has a blind spot. In short: It's not just about the possibility, but also the effectiveness of psychometrics. The thesis of my talk is: Psychometrics changes the forms of thought. To explain this, I am referring to a structure which is fundamental to this matter: The subject (mind) itself is related to this relation of measurement and mind. In this case, the subject is oriented to the scale that is applied to it. This distinguishes it from classical practices like the measurement of temperature or length (no other entity would try to increase its size if a ruler is applied to it). Therefore the debate concerning psychometrics is relevant regardless of its epistemological implications - psychometrics changes the forms of thought.

Andreas Gelhard: *Measuring Competence Rather than Intelligence? McClelland's Claim*

David McClellands Essay „Testing for competence rather than for ‚intelligence'" (1973) is widely acclaimed as the beginning of a new testing movement which hopes to avoid the pitfalls of traditional intelligence testing. My paper gives a short sketch of McClellands criticism regarding IQ-tests and his own attempt "to be tough-minded (i.e., experimental) about a tender-minded subject (i.e., human motivation)". Keeping in mind current attempts to measure competence it might be particularly revealing to re-examine McClellands discussion of operant and respondent measures.

Malte Bachem: *Matching Personality to Vocation. A Historical Perspective on an Epistemological Model around 1920*

Around 1920, the matching of personality to vocation and vice versa was established as an epistemological model within Swiss vocational counseling. The conceptual and practical combination of personality and vocation aimed at the production of social and individual stability. The presentation focuses on the two essential formats of occupational counseling, the general and the individual counseling. Each format presented its own way to deal with personality. The general guidance can be described as a mighty and precise calculator for personality. The individual guidance dealt with the problem of measuring the immeasurable personality.

Simulations, Models and Measurements

Annamaria Carusi: *Modelling Measurements and Constitutive Realism*

In order to be successful as a research programme in the life sciences, modelling and simulation need to be meaningfully connected to experiments. Indeed, the realism of the model depends upon successfully making this connection. This paper discusses the crucial role of measurement in bringing about this interconnection.
Models and simulations are empty of empirical content without being parameterised by data acquired from experiments; and they further require connection with the experiments in order to be validated. There is still not consensus in the field regarding how validation should be considered, and this is exacerbated when the models shift outside of disciplinary domains, for example, when they shift from settings dominated by mathematics and engineering to

settings dominated by clinical aims and concerns. In Carusi, Burrage and Rodriguez (2012 and 2013), we showed that experiment, model and simulation should be considered to be a hybrid system of interconnected processes in order to interpret the results of any validation test. Since it cannot be assumed in advance that the results of laboratory experiments and computational simulations can be meaningfully compared to each other in a validation test, we proposed that the iterative relations between these stages of the process go towards establishing grounds of comparability. This paper discusses two examples of the way in which comparability is dealt with in the cardiac modelling research programme, one taken from the inception of the cardiac modelling research programme experimental system when the Hodgkin-Huxley model of electrical excitation in nerve cells was first adapted for cardiac cells, and one taken from a current development in this research programme, which is developing a population of models approach to exploring the interconnections between models and experiments. These examples will show how quantitative results of modelling and experimenting are interpreted, established as significant, and result in further articulations of both modelling source and target domain: these are all aspects in which the emerging grounds of comparability are manifested in the modelling and experimenting process.

The second part of the paper considers what philosophical accounts might be given of the way in which grounds of comparability come to be established, and draws upon accounts of art and literature for this. Grounds of comparability have to do with a relationship at the heart of 'realist' art and literature, that is the ways in which symbolic systems relate in a generative way to objects in the world. Two accounts are particularly promising: The first is Joseph Rouse's description of experimental systems as 'materialized fictional "worlds"' which are domain constituting in that they 'help constitute the fields of possible judgment and the conceptual norms that allow [conceptualizable] features to show themselves intelligibly' (51). The second is Merleau-Ponty's idea of the coherent deformations brought about by style, and the way in which these institute (rather than constitute) systems of equivalence which articulate a field of interactions in which things can be experienced or counted as equivalent. These two accounts – the constitutive and the institutive – approach the process of establishing measurements that count from different directions: Rouse's constitutive approach from a conceptual standpoint, and Merleau-Ponty's from a non-conceptual standpoint.

The presentation will focus on the constitutive account, and point to the gaps in it which call for something closer to the institutive account.

Wendy Parker: *Data Assimilation, Measurement and the Construction of Global Climate Datasets*

Computer simulation has brought "disturbances" (Keller 2003, p.202) to traditional concepts in philosophy of science. The most salient example of this is scientists' frequent description of simulation studies as computer experiments. These simulation studies differ in some obvious ways from the laboratory and field studies that are traditional exemplars of experiment. Do they nevertheless constitute a new kind of experiment, made possible by the digital computer? In addressing questions like this, philosophers of science have been prompted to re-examine existing ideas and intuitions about experiment.

This paper explores some of the ways in which a new practice involving computer simulation is bringing further disturbances to traditional notions. This new practice – known as data assimilation – has undergone rapid development in the geosciences over the last few decades, especially in the field of atmospheric science. In data assimilation, computer simulation plays an integral role in the production of "observational" data sets – datasets used in the roles of traditional observations. In fact, many key "observational" datasets used in the study of weather and climate today are in fact composed entirely of simulation output. Here, it is not the concepts of theory and experiment that come under pressure, but rather those of observation, observing instrument, and measurement.

After providing a non-technical introduction to data assimilation (and a closely-related practice known as reanalysis), I briefly illustrate some of the ways in which these practices are putting pressure on traditional concepts – in particular the concepts of observation and observing instrument/system. Most notably, it is sometimes claimed that computer simulation models used in data assimilation serve as independent observing systems, supplementing the thermometers and satellites of the conventional observing system. I will reject such a characterization but will argue that (i) data assimilation systems as a whole can be classified as complex observing systems and (ii) simulation results produced via data assimilation might well be considered measurements of atmospheric properties.

In making a case for these conclusions, I will identify three core features of measurement; in particular, I will suggest that measurement is a procedure that is empirical, backward-looking and informative. While most computer simulation studies lack one or more of these core features of measurement, some data assimilation methods can be argued to have all three. These features are consonant with the view of measurement recently advocated by van Fraassen (2008) and, indeed, I suggest that the results of data assimilation could qualify as measurement outcomes on his view. I also consider the objection that data assimilation as a whole should not be considered a measuring procedure but rather can be neatly partitioned into a part that is measurement proper (involving conventional observing instruments like thermometers) and a part that is

subsequent calculation. I resist this objection, arguing that data assimilation occupies one end of a spectrum of measuring practices that involve models and calculation in various ways.

Teru Miyake: *Uncertainty and Modeling in Seismology*

Philosophical discussion of measurement is usually about the measurement of discrete values. But in some cases, we want to determine something more complicated. One example is the density of the interior of the earth. If you assume spherical symmetry, the density of the interior of the earth is given by a continuous function of a single variable r (distance from the center of the earth). This density function is taken to be very complicated, with discontinuities and anomalies at various depths.

Geophysicists have tried determining it by measurement of the travel times of seismic waves and the frequencies of the normal modes of oscillation of the earth. The density distribution of the interior of the earth is modeled in a certain way, and the parameters of that model are varied until a model is found that can account for the observed frequencies. Even if you find a density function that can account very well for the observed frequencies of the normal modes, however, you cannot discount the possibility of a radically different density distribution being able to account for those frequencies just as well.

One reason for this underdetermination is simply because of the mathematics—a finite number of observations are used to try to determine a model with a huge number of degrees of freedom. But uncertainties can arise for other reasons. One reason is that the models are always idealized (parameterized) in certain ways, and there are different ways in which you could idealize the models. Different models, using different parameters, might come up with very different density distributions.

Another reason is that the method depends on assumptions that must be made in order to apply the theory of waves in elastic media. For example, we might assume that the earth is elastic and isotropice. We might get worried, then, about how such assumptions would affect the determination of the density distribution.

Another problem in geophysics, the determination of the source geometry of earthquakes from seismic waves, gives rise to similar difficulties, for similar reasons. The function to be determined, the source geometry of an earthquake, is very complicated, and different ways of modeling it could potentially give rise to very different solutions.

In this paper, I examine the kinds of strategies that have been developed by geophysicists to try to address these problems, particularly how to deal with the uncertainties arising from modeling. In particular, a look at the history of the determination of the interior density of the earth reveals a piecemeal approach,

perhaps best characterized as bootstrapping. I will inquire into whether such a piecemeal approach could be used in seismic source determination.

History of Measurement and Measurement Theory

Francesca Biagioli: *Empirical and Formal Conditions in Helmholtz's Theory of Measurement*

Helmholtz (1887) has been acknowledged as one of the forefathers of measurement theory. However, his conception of measurement differs from the modern, representational conception in several aspects. Firstly, Helmholtz does not clearly distinguish numbers from numerals, and uses considerations regarding numerals to support an empiricist philosophy of arithmetic. Secondly, his analysis of measurement entails something more than a study of conditions for using numbers in modelling measurement situations, because it presupposes that mathematical structures are common to subjective experiences and objective ones. I suggest that Helmholtz's supposition rests upon a Kantian line of argument that can be summarized as follows: additive principles can be formulated independently of the entities to be measured, and, nevertheless, they are necessary for judgements about quantities to be generally valid. That is to say, mathematical structures have a constitutive function in Kant's sense. Helmholtz therefore needs formal conditions of measurement. At the same time, his inquiry has an empirical aspect because mathematical structures are supposed to have a constitutive function only relative to scientific theories and their advancements. For the same reason, Helmholtz maintains that the extension of the laws of addition to all known physical processes should proceed progressively and independently of supposedly external limits such as incommensurable magnitudes.

Stéphanie Dupouy: *How Meaningful is it to Measure Sensation? Discussions of Fechner in France in the 1870's and 1880's*

Psychophysics, the domain of psychology that studies the relationship between the physical stimuli and the psychological events that they cause in the mind, is grounded, since its foundation by Gustav Theodor Fechner (1801-1887), in the idea that sensation is accessible to measurement—a proposition that has been attacked by various philosophers such as Bergson and James, contesting the idea that sensations can be decomposed into homogenous, elementary units that could be numbered and added to one another. The meaningfulness that can be attributed to such an operation of measurement is closely related to the degree of conventionality one is willing to incorporate into the notion of measurement itself,

i.e., to how far one is ready to depart from an intuitive, essentalist view of measurement. Focusing on this bundle of problems, this paper examines the reception of Fechner's psychophysics in France (and among French-speaking psychologists), studying in particular the epistemological debates on the measurement of sensations that took place in philosophical and scientific journals (Revue scientifique, Revue Philosophique) in the 1870's and 1880's. More precisely, I examine a number of objections against the very possibility of measuring sensations which were initially generated by the French mathematician Jules Tannery (1848-1910) in an article published in 1875. In particular, I will focus on the critiques of Fechner's method formulated by the Belgian mathematician, philosopher and psychologist Joseph Delbœuf (1831-1896). Fechner's method of extending measurement to the psychological realm was based on the principle that any measurement is, in any case, "indirect," i.e., mediated by realities other than the quantity to be measured (such as the mediation of space in the case of the measurement of time, for instance). I will examine how Delbœuf, without being opposed to the measurement of sensation per se, was progressively led, by contrast, to promote another kind of psychophysical measurement, a "direct" measurement, supposedly more intuitive and more appropriate to the nature of the nervous system than Fechner's "indirect" measurement. The paper discusses the relevance of this concept of "direct" measurement, applied to sensation.

Jochen F Mayer: *On the 'Mathematisation' of Measurement in (West) German Academic and Official Statistics c.1920-1950*

The application of mathematical calculus to social phenomena is a defining but contested feature of statistics as a science and tool of government (Porter 1987). This paper explores German statistical discourse at a crucial moment when, especially after 1945, mathematically trained statisticians increasingly challenged the authority of then dominant social and economic statistics. Drawing on archival material of the German Statistical Association's annual meetings and statistical textbooks, I argue that 'mathematisation' in statistics was a particular discourse in the language of contemporary statisticians. The first part of the paper utilises the idea of 'boundary disputes' (Gieryn 2001) to show how statisticians used the 'mathematisation' discourse to (re-)order knowledge within their discipline in intellectual (sampling theories, formulas) and institutional (education and training of statisticians) terms. Social statisticians (the 'Frankfurt School') defended their factual logic but struggled to keep their discipline 'pure' from advancements in probability theory. Mathematical statisticians (the 'Munich School'), by contrast, translated 'mathematisation' into matters of methodical proficiency with the aim of expanding the epistemic authority of their field.

The main part of this paper will focus on how 'mathematisation' played out differently with regard to measurement issues, their cognitive boundaries and to claims for statistical objectivity. I will show how 'Frankfurt' statisticians defended an autonomous space for the measurement of the social world essentially demanding the logical definition of elements to be counted; their 'social arithmetic' involved conceptual work, similar to that in contemporary social sciences. Social statistics were seen as a method of mass observation essential for the study of society precisely because of the diversity of individuals. Society, however, ultimately remained lawless and unanalysable. By contrast, for mathematically trained statisticians, 'mathematisation' first and foremost helped to demarcate a space for statistical practice based on probability calculus, which, in contrast to 'social arithmetic', served as a unifying, overarching perspective that could be applied infinitely to any research field. The mathematical language thus functioned as a formula to foster agreement among all those who had mastered its procedures while excluding all those who did not. This language – almost entirely incomprehensible to old school social statisticians – helped set numerical 'rules' that established when and how judgments could be made on the basis of partial or uncertain information; a matter for which 'Frankfurt' statisticians had nothing but their 'sure instincts'.

Saturday, 16 March • 14:00-16:00

Plenary Lectures

Hasok Chang: *Operationalism: Old Lessons and New Challenges*

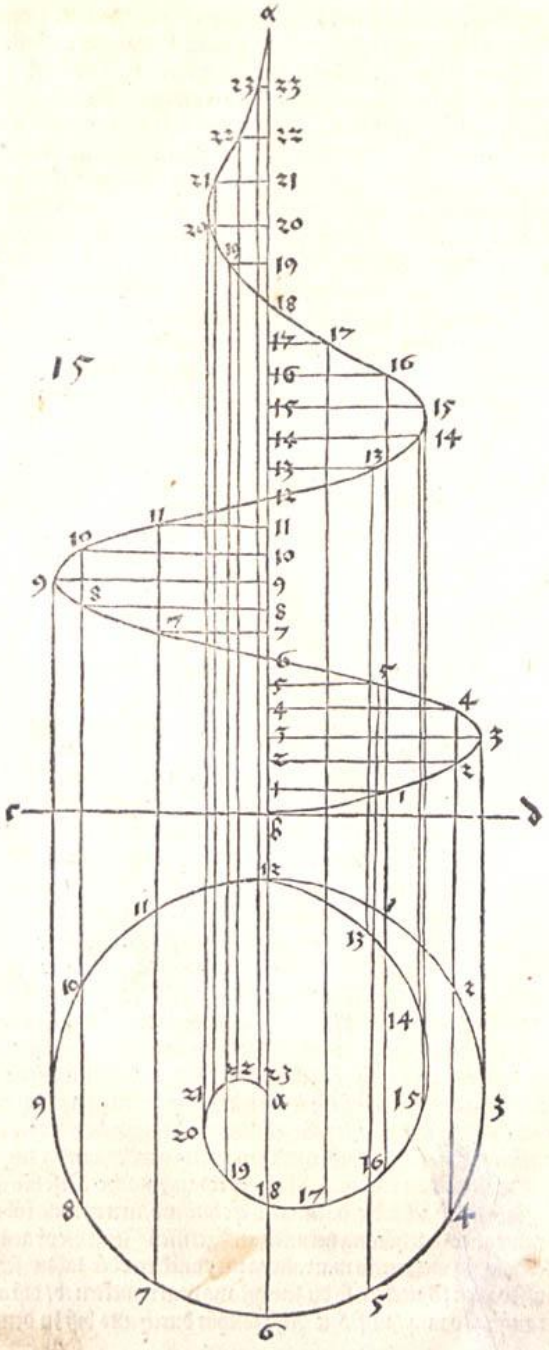
Even though operationalism has long been rejected by most philosophers and scientists, there are valuable insights in Bridgman's original reflections worth recovering and developing further. I propose taking operationalism not as a theory of meaning, but as a directive to seek (higher degrees of) meaningfulness in concepts employed in empirical inquiry. The demand that concepts be backed up by measurement operations should be coupled with a recognition that meaning can only be gained through the hard work of finding or creating operations linked to the concepts in question. On the basis of the late-Wittgensteinian notion of meaning-as-use, it is easy to see that the more uses a concept has, the more meaningful it is. According to my doctrine of "active realism", operationalism is the striving to create more opportunities for learning about reality. Bridgman's skeptical worries were valid concerns about the loss of meaningfulness arising from careless empirical and mathematical extensions of concepts. From this perspective also arises a new solution to the problem of the multiplicity of measurement methods that avoids both Bridgman's timid conscience and Hempel's

deference to theory: the multiplicity should be celebrated as richness of meaning, as long as some plausible links can be established between each measurement method and the concept in question.

Eran Tal: *A Model-Based Epistemology of Measurement*

The epistemology of measurement is an interdisciplinary area of research concerned with the conditions under which measurement and standardization methods produce knowledge, the nature, scope, and limits of this knowledge and the sources of its reliability. In recent years it has become increasingly recognized that the content and quality of knowledge obtained by measuring depends on the theoretical and statistical assumptions with which a measurement process is modeled. These assumptions often involve idealization, that is, intentional distortion of aspects of the measuring instrument, measured object and environment. The presence of idealizations may appear to threat the accuracy and objectivity of knowledge gained through measurement. Here I show that the opposite is the case: idealization is a necessary precondition for obtaining accurate and objective measurement outcomes at all. A measurement outcome, I argue, is a value range assigned to a parameter in a model in a way that allows the model to coherently predict the final states ('indications') of a process. Idealizations are necessary for identifying the measured parameter with a particular object, for distinguishing genuine effects from errors and for comparing measurement outcomes to each other. These claims are exemplified with a study of the contemporary evaluation and comparison of atomic clocks across national metrological laboratories. Building on these insights, I conclude by highlighting the promise held by model-based approaches for further research in the epistemology of measurement.

Picture: Albrecht Dürer: ‚Underweysung der Messung mit dem Zirckel und Richtscheyt.' Nürnberg 1538, p. 7v.



Diß iſt der ſchneck aus dem grůd auf gezogen / mit allen nottürfftigen linien daraus er gemacht wirdet.