ERC Starting Grant 2014
Research proposal [Part B1]

Making Scientific Inferences More Objective

OBJECTIVITY

Jan Sprenger

Tilburg University
Project duration: 60 months

What makes scientific inferences trustworthy? Why do we think that scientific knowledge is more than the subjective opinion of clever people at universities? When answering these questions, the notion of objectivity plays a crucial role: the label "objective" (1) marks an inference as unbiased and trustworthy and (2) grounds the authority of science in society. Conversely, any challenge to this image of objectivity undermines public trust in science. Sometimes these challenges consist in outright conflicts of interests, but sometimes, they are of a foundational epistemic nature. For instance, standard inference techniques in medicine and psychology have been shown to give a biased and misleading picture of reality.

My project addresses precisely those epistemic challenges and develops ways of making scientific inferences more objective. Our key move is to go beyond the traditional definition of objectivity as a "view from nowhere" and to calibrate the most recent philosophical accounts of objectivity (e.g., convergence of different inference methods) with the practice of scientific inference. The combination of normative and descriptive analysis is likely to break new ground in philosophy of science and beyond. In particular, we demonstrate how two salient features of scientific practice—methodological pluralism and subjective choices in inference—can be reconciled with the aim of objective knowledge.

The benefits of the proposed research are manifold. First and foremost, it will greatly enhance our understanding of the scope and limits of scientific objectivity. Second, it will improve standard forms of scientific inference, such as hypothesis testing and causal and explanatory reasoning. This will be highly useful for scientific practitioners from nearly all empirical disciplines. Third, we will apply our theoretical insights to ameliorating the design and interpretation of clinical trials, where objectivity and impartiality are *sine qua non* requirements.

Section A. Extended Synopsis of the Scientific Research Proposal.

Problem Description and State-of-the-Art

Why is scientific knowledge trustworthy? Why do we think that it is more than the subjective opinion of clever people with university jobs? In answering these questions, we inevitably have to deal with the **objectivity of scientific inferences**. Objectivity is generally considered to be an epistemic virtue: it marks an inference as unbiased and trustworthy and grounds the authority of science in society. For example, medical drugs will not be admitted to the market unless there is objective evidence that proves their efficacy. However, do these objectivity requirements really make science more trustworthy? As evidenced by longstanding methodological debates, "objective" inference methods in medicine, psychology and other disciplines lead to the publication of a lot of misleading research findings (e.g., Ioannidis 2005). Troubles with the objectivity of scientific inference may thus sketch a biased picture of reality, undermine the epistemic authority of science and lead to badly informed and inferior policy decisions.

This project focuses on the objectivity of scientific **inferences from evidence ("data") to theory**. It adopts a **philosophical, epistemological perspective**, asking the question of how we can eliminate bias and promote objectivity in scientific inference. We will focus on three ubiquitous types of scientific inference: statistical, causal and explanatory inference ("Inference to the Best Explanation").

The objectivity of all three modes of inference is challenged by (i) the existence of several well-founded and competing paradigms for making such inferences; (ii) the partial reliance on subjective factors in these inferences: Bayesian (=subjectivist), likelihoodist and frequentist methods compete for the "right" methodology for statistical inference (Sprenger 2014). Process theories of causation (Salmon 1984) compete with probabilistic networks for causal inference (Pearl 2000); subjective Bayesian, causal-mechanical and psychological approaches give different accounts of explanatory power (Schupbach and Sprenger 2011). All this seems to be incompatible with the received view on objectivity in science where objectivity is thought of as an impersonal, a-perspectival "view from nowhere" (Nagel 1986). According to that view, objectivity amounts to representing the world as it really is, unmediated by human minds and other distortions. This perspective leaves little room for subjective elements in inference or pluralism about inference methods, which are widespread characteristics of scientific practice. But if we aim at a realistic and practically useful logic of scientific inference, we have to spell out how objective knowledge can emerge from pluralist and partially subjective inference methods. This brings us to our principal research question:

How can we assess—and increase—the objectivity of scientific inferences in the light of the plurality of methods and the presence of subjective elements that characterizes scientific reasoning?

The key to making scientific inferences more objective lies in rethinking the concept of scientific objectivity and **applying philosophical innovations regarding scientific objectivity to these three types of inference**. The philosophical analysis of objectivity has rapidly advanced in the last decades: historians and philosophers of science have pointed out the limits of the a-perspectival idea of objectivity and thoroughly debunked it (e.g., Daston and Galison 2007). Nowadays, objectivity is rather located in a scientific culture of open transformative criticism (Longino 1990), an appropriate balancing of various individual perspectives and values (Douglas 2004, Giere 2006), or the convergence of different methods and approaches (ibid.). That is, the subjective elements in scientific reasoning are no more denied: they become part of a more realistic and refined image of scientific objectivity. Neither is the existence of different scientific perspectives seen as a threat to the objectivity and reliability of science.

While these new conceptions of scientific objectivity have been successfully applied to social and institutional aspects of scientific inquiry, the **implications for a logic of scientific inference** have not been explored so far. This is the salient gap that this project is going to fill. It calibrates advances in the philosophical analysis of objectivity with formal tools for making scientific inferences, and regulatory and institutional constraints. Especially, it aims at showing how pluralist and partially subjective inference methods can be fit into a refined philosophical account of objectivity in scientific inference. As part of this quest, we will answer the following research questions:

1. Can the idea of scientific objectivity as convergence of different methods and resistance to open, rational criticism be fruitfully applied to scientific inference?

2. Are there common grounds between the different statistical schools, and if yes, do they provide a foundation for objective statistical inference?
3. How can probabilistic models of causality with their undisputedly subjective elements account for supposedly objective cause-and-effect-relations?
4. Can we establish Inference to the Best Explanation as an objective form inference by developing a model of explanatory power that synthesizes causal and probabilistic models?
5. How can our new account of scientific objectivity be applied to improving the conduct of clinical trials, and to appraising evidence claims in medicine?

Research Plan and Methodology

A major challenge of the project consists in demonstrating how scientific inferences can be given an objective foundation and be defended against the charge of arbitrary subjectivity, without denying the inevitable role of subjective factors in scientific inference. To this end, we will transfer the idea of objectivity as convergence of different approaches and methods to different modes of scientific inference, and reconcile the different schools of statistical/causal/explanatory inference (Douglas 2004; Reiss and Sprenger 2014). Crucially, we will give a place to subjective elements (e.g., Bayesian reasoning) in a theory of objective scientific inference. By integrating our normative theoretical arguments with the state-of-the-art of scientific inference, we will give a novel and sustainable foundation to objectivity in scientific inference and improve currently used inference methods.

The research team consists of the PI, who is engaged into all subprojects, a postdoctoral researcher and two PhD students. Together, we work on four subprojects. Three subprojects (A-C) investigate statistical, causal and explanatory inference and a fourth one (D) applies the theoretical insights to the interpretation of clinical trials. Thereby we validate our results with the help of an empirical case study and demonstrate their practical relevance. Methodologically, the subprojects are tied together by the central role that **probabilistic reasoning** plays in scientific inference, on a theoretical as well as on an applied level. More precisely, the project combines several methods. As a main method of analytic philosophy, **conceptual analysis** naturally plays a prominent role in the project. The methodological arsenal is complemented by

(1) **formal modeling** and **explication** with the help of **probabilistic models** (all subprojects);
(2) the analysis of selected **case studies**, such as clinical trials in Subproject D;
(3) **psychological experiments** on explanatory reasoning and causal strength in Subproject B and C
(4) **integrating insights from statistical and philosophical literature** (Subproject A, B and D).

We combine these approaches by the method of **reflective equilibrium**, appropriately balancing conceptual desiderata, formal models and empirical findings. The fruitful interplay between conceptual and experimental work makes the project empirically calibrated and philosophically sensitive, and avoids the evident problems of relying on a single approach. Combining formal results with scientific practice and people's actual reasoning will ensure the ground-breaking impact of our research, especially when compared to earlier efforts that have not taken a multi-method approach. Moreover, the connection to medical statistics ensures that the research is not stuck in a narrow disciplinary perspective, but yields results that improve scientific inference and decision-making.

Description of the Subprojects

In the **preliminary phase**, we review various explications of scientific objectivity proposed in the philosophical literature and evaluate them with respect to their potential to be transferred to the nitty-gritty details of scientific inference and probabilistic inference in particular (Reiss and Sprenger 2014). Specifically, we zoom in on the thesis that objectivity in scientific inference is compatible with epistemic pluralism. We take Douglas' (2004) idea that objectivity can be located in the convergence and proper balancing of different methods and approaches, and argue that it has sufficient resources to integrate subjective choices and perspectives into objective scientific inference. Our principal case study is model selection in science (Sprenger 2013b), which naturally combines subjective choices with formal inference techniques. By spelling out this new conception of objectivity for probabilistic reasoning—the methodological umbrella of all subprojects—this preliminary investigation secures the philosophical foundations of the research project.

Subproject A focuses on generating **objective scientific knowledge through hypothesis testing**. To eliminate personal bias in inference, statisticians usually use **standardized inference procedures** such as null hypothesis significance tests (NHST). For example, medical drugs will not get approval by regulatory agencies unless their efficacy is backed up by statistically significant findings in NHST. However, recent criticisms of NHST elucidate that they are not as objective as they appear: they invite to misinterpretations, they do not provide objective guidelines for an appraisal of insignificant findings, and they imply tacit value judgments (e.g., Goodman 1999; Sprenger 2009). They also fail to compatible with influential theoretical paradigms such as Bayesian (=subjective probabilistic) reasoning and rational choice theory.
To address this challenge, we apply the idea of objectivity through convergence of different methods to two major schools of statistical inference, Bayesianism and frequentism (Sprenger 2014). That is, the subproject relates objectivity in statistical hypothesis testing to the **common denominator of Bayesianism and frequentism**, instead of continuing a deadlocked debate (cf. Sprenger 2013a). By examining whether objectivity can be established through measures of evidence that possess a valid Bayesian *and* a valid frequentist interpretation, we develop a new theory of objectivity in statistical inference and assess the prospects for a methodological unification of inductive reasoning.
Specifically, we provide an objective guideline for interpreting *insignificant results*. That is, we give a **probabilistic measure of corroboration**—the degree to which the null hypothesis has withstood attempts to refute it (Popper 1934/2002). This measure quantifies the epistemic relevance of statistically insignificant findings and cures a salient deficit of the logic of NHST. We derive a specific corroboration measure from a set of desiderata that capture core ideas of Bayesian *and* frequentist inference. Thereby, we stay neutral between both frameworks and achieve an objective appraisal of insignificant test results. Preliminary results in this direction have recently been presented at the *Philogica III* congress in Bogotá.

Subproject B engages in a **objective quantification of causal strength**. Scientists are, for all kinds of theoretical and practical purposes, excited about measuring the objective strength of a link between cause and effect. Any complete theory of causal inference needs to quantify the notion of causal strength. At the same time, such a theory needs to have an objective basis in order to rescue the powerful intuition that causal relations track features of the real world rather than psychological associations.
The probabilistic theory of causality (Suppes 1970; Pearl 2000), with the idea that causes raise the probabilities of their effects, provides an adequate formal model for conducting a **systematic normative comparison of measures of causal strength**. First, we develop adequacy conditions for measuring causal strength in terms of statistical relevance. Second, since it is unlikely that this search will converge on a single measure, we conduct psychological experiments for calibrating these measures with people's causal intuitions. Our experiments will transfer the methods of Tentori et al. (2007) in empirical work on inductive reasoning to causal judgments. Third, we assess the prospects for an objective notion of causal strength on the basis of our quantitative analysis, by investigating how probabilistic measures of causal strength with their inevitable subjective elements (choice of variables, assignments of probabilities) can be integrated with classical, process- and mechanism-based accounts of causation. Again, we build on the idea that objectivity can be established by making different research programs converge.
Finally, we transfer our analysis from the strength of causal links between *types* of events to the assessment of causal strength in *actual* events. Thus, we see whether general causation ("factor X tends to raise the value of factor Y") and singular causation ("event X is the cause of event Y") can be unified on a quantitative level. As a case study where objectivity is a crucial requirement, we explore whether our measure of causal strength can apportion blame and responsibility among several causes of an event in contract and tort law.

Subproject C investigates the **logic and objectivity of explanatory inference**. Present accounts of explanatory inference are very diverse and stress, inter alia, (i) a subjective sense of understanding, (ii) causal relevance, (iii) logical and statistical relations. These strands of research are all well-motivated, but normally pursued separately. As a consequence, the objectivity of explanatory inference is hard to assess. In this subproject, we combine probabilistic accounts of explanatory power (Schupbach and Sprenger 2011) with causal and phenomenological accounts in order to construct a comprehensive and objective model of Inference to the Best Explanation.
We examine how people reason from probabilistic information and contextual cues to judgments about the explanatory power of a hypothesis. In particular, we test a dual source model of explanatory reasoning, where causal plausibility and probabilistic relevance both determine the explanatory value of a hypothesis. Judgments are elicited using specific vignettes and lab/online surveys that figure in ongoing research on explanatory and probabilistic reasoning (Colombo, Sprenger and Postma-Nilsenová 2014). On this basis, we compare measures of explanatory power and causal strength (→ Subproject B), embed Inference to the Best

Explanation into probabilistic reasoning and assess whether it should be regarded as an independent mode of inference, as a useful heuristics, or as a special type of probabilistic reasoning.
The empirical parts of Subproject B and C are conducted in collaboration with scientists from the interdisciplinary (philosophy/psychology) DFG priority program #1516 "New Frameworks of Rationality" where I am one of the project leaders.

Subproject D applies the insights from Subproject A and B to **clinical trials in medicine**, a classical battleground for challenging the objectivity of scientific research findings. Clinical trials evaluate the adequacy of causal medical hypotheses by statistical means, but they are often charged of being severely biased (Goodman 1999; Ioannidis 2005; Montori et al. 2005). Therefore, we address the question of **how different forms of bias (publication bias, sampling bias) can be alleviated** with the help of the theoretical results in Subproject A & B. The challenge of this study consists in fitting the concepts of degree of corroboration and causal strength into the regulatory framework of medical research, and in characterizing those applications where these concepts are an especially important criterion for medical inference and decision-making. These contexts include, inter alia, tests of the safety of medical drugs where the corroboration of the tested null hypothesis is of substantial practical interest. We will use a case-study approach and apply our technical tools to a re-analysis of medical datasets from phase III and IV trials, with due considerations of the epistemic goals and regulatory constraints of such trials.
Furthermore, based on our advances in the logic of NHST (→ Subproject A), we elaborate how flexible design of medical trials enhances their efficiency without leading to a loss in objectivity (Nardini and Sprenger 2013). To this end, we demonstrate that perceptions of bias in evaluating effect size estimates disappear if the interpretation of NHST is put on a firmer epistemic basis. This novel evaluation of bias in clinical trials will ultimately lead to better and more efficient decisions in clinical care.
In getting access to relevant datasets, analyzing them and proliferating the final results to the medical community, we will be assisted by the medical psychology department of Tilburg University (Prof. Dr. J. Denollet, Prof. Dr. S. Pedersen) and the Istituto Europeo d'Oncologia in Milan (Prof. Dr. G. Boniolo).

Finally, we **synthesize our results into a book manuscript** "Objectivity in Scientific Inference" that tests and validates our theoretical conclusions with the help of the medicine case study. In this manuscript, we transfer novel philosophical conceptions of objectivity to statistical, causal and explanatory inference, we elucidate the nexus between those modes of inference, and we explore the implications of our achievements for the idea(l) of objectivity in science.

Innovation and Impact

The topic of objectivity in science is old, but the present project leaves the usual paths by combining normative and descriptive analysis. First, this project constitutes the **first systematic integration of conceptual philosophical groundwork on scientific objectivity with the practice of modern scientific reasoning, and formal models for describing it**. By doing so, it updates classical work on values and objectivity in statistical inference, and work on the objectivity of causal and explanatory relations, with empirical data and contemporary scientific practice. Second, in all three investigated types of inference, probabilistic reasoning plays a major role. However, the debate about objectivity in probabilistic reasoning is stuck in an ideological divide between Bayesians and frequentists. Therefore, this project provides the **first defense of how a pluralist methodology for probabilistic reasoning and scientific inference can yield objective knowledge**, instead of arguing that objectivity amounts to following specific research methods, or to having uniquely rational degrees of belief. It also opens **new horizons for gaining objective knowledge from hypothesis tests**. At the end of the project, we will achieve a better understanding of objectivity in scientific inference, obtain a sustainable defense of the ideal of objectivity in a pluralist setting, and explain why the a-perspectival idea of scientific objectivity was so influential for a long time. Such an explanation is only partially provided by research that focuses on historical/sociological aspects of objectivity, or the role of values and community standards.

Actually, the project transcends the boundaries of a purely philosophical investigation: its results are relevant for all scientists who have to make reliable inferences from data to theory. It is also valuable for all those decision-makers who have to routinely interpret scientific research findings. The impact of the acquired knowledge can be systematized in three dimensions:

1. **Better evaluation of statistical data.** Improving the **logic of NHST** (→ Subproject A) solves several problems of great practical interest, such as the appraisal of insignificant findings and reconciling Bayesian and frequentist analysis. These results contribute to **fighting publication bias** in science and close a salient gap that current standard methods (such as p-values or confidence intervals) leave open. Hence, almost all empirically working scientific disciplines can benefit from our innovations regarding the interpretation of NHST .
   On the policy side, expert findings such as the IPCC reports on climate change, are often criticized as not living up to their objectivity claims. However, these critiques are often based on outdated and misleading ideas about scientific objectivity. A more refined account of objectivity in statistical inference will help to respond to these critics, assist the writers of such reports in formulating their conclusions with the appropriate care, and support the authority of science in the public arena.
2. **Better understanding of human reasoning.** Statistical, causal and explanatory inferences are all cornerstones of human reasoning, and integrating them is one of the big challenges for experimental psychology. This project contributes, especially in Subproject B and C, to the theoretical foundations for such an integration. At the same time, it investigates empirically to what extent such inferences have an intersubjective basis.
3. **Fighting bias in evaluating medical trials, and better decisions in clinical care.** Our research analyzes several forms of bias in clinical trials and shows that some of them need not be a problem for the epistemic authority of medical research (e.g., sampling bias). We also show how modifying the logic of statistical inference in medicine, and the interpretational perspective on those trials (→ Subproject D), can lead to better appraisal of medical evidence and superior decisions in clinical care.

References

Colombo, M., Sprenger, J., and Postma-Nilsenová, M. (2014): *Explanatory Value, Probability and Abductive Reasoning*. Manuscript under review for the *Proceedings of the Conference of the Cognitive Science Society 2014*.

Daston, L., and Galison, P. (2007): *Objectivity*. New York: Zone Books.

Douglas, H. (2004): The Irreducible Complexity of Objectivity. *Synthese* 138: 453—473.

Giere, R. (2006): *Scientific Perspectivism*. Chicago: University of Chicago Press.

Goodman, S. (1999): Toward Evidence-Based Medical Statistics. 1: The *P* Value Fallacy, *Annals of Internal Medicine* 130: 995–1004.

Ioannidis, J.P.A. (2005): Why Most Published Research Findings Are False. *PLOS Medicine* 2: e124. doi:10.1371/journal.pmed.0020124.

Lipton, P. (2004): *Inference to the Best Explanation*. Cambridge: Cambridge University Press. Second Edition.

Longino, H. (1990): *Science as Social Knowledge. Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

Montori, V.M., et. al. (2005): "Randomized trials stopped early for benefit: A systematic review". *Journal of the American Medical Association* 294: 2203–2209.

Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.

Nardini, C., and J. Sprenger (2013): Bias and Conditioning in Sequential Medical Trials. *Philosophy of Science 80*: 1053—1064.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Popper, K.R. (1934/59), Logik der Forschung, Berlin: Akademie Verlag. English edition as *The Logic of Scientific Discovery*. London: Routledge.

Reiss, J., and Sprenger, J. (2014): Scientific Objectivity. Forthcoming lemma in the *Stanford Encyclopedia of Philosophy*.

Salmon, W.C. (1984): *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

Schupbach, J., and Sprenger, J. (2011): The Logic of Explanatory Power. *Philosophy of Science* 78: 105–127.

Sprenger, J. (2009): Evidence and Experimental Design in Sequential Trials. *Philosophy of Science* 76: 637—649,

Sprenger, J. (2013a): Testing a Precise Null Hypothesis: The Case of Lindley's Paradox. *Philosophy of Science*.

Sprenger, J. (2013b): The Role of Bayesian Philosophy within Bayesian Model Selection. *European Journal for the Philosophy of Science* 3: 101—114, 2013.

Sprenger, J. (2014): Bayesians vs. Frequentists in Statistical Inference. Forthcoming in *Oxford Handbook of the Philosophy of Probability*, eds. A. Hajek and C. Hitchcock. Oxford: Oxford University Press.

Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2007): Comparison of confirmation measures. *Cognition* 103: 107–119.