Degrees of Value-Ladenness and Signal-to-Noise Ratio

Torsten Wilholt

Institut für Philosophie and Leibniz Center for Science and Society, Leibniz Universität Hannover, Hannover, Germany

torsten.wilholt@philos.uni-hannover.de

Abstract

Although fundamental arguments have been presented to support the value-laden nature of all scientific research, they appear to be difficult to apply to basic research in physics. To explain this, I argue that basic research in physics is, in a very specific respect, often value-laden to a lesser degree. To spell this out, I refer to the different signal-to-noise ratios that can be achieved in different fields of research. I also argue that having a very low degree of value-ladenness in the very specific respect that I identify does not mean that the research is not value-laden at all.

1. Introduction

An emerging consensus in the philosophy of science holds that non-epistemic values inevitably have an influence on key parts of scientific processes. Philosophers have shown this using examples from the life sciences, the environmental sciences, climate science, or the social sciences (Alexandrova 2018, Douglas 2000, Elliott 2011, Steel and Whyte 2012, Wilholt 2009, Winsberg 2012). Such demonstrations have not yet been put forward for research that deals with questions of physics or chemistry that have no direct application relevance. Although very fundamental and general arguments have been presented to support the value-laden nature of all scientific research, they appear to be difficult to apply to these areas. In this paper, I explain why this is the case. I will argue that 'basic' research in physics is actually, in a very specific respect, value-laden to a lesser degree than the well-known examples from the life and environmental sciences. To explain this, I will refer to the different signal-to-noise ratios that can be achieved in different fields of research. In the end, however, I will also argue that this finding should not be confused with the alleged fact that basic research on physics is not value-laden at all.

In order to make the question manageable, I limit myself in this paper to one fundamental argument for the value-ladenness of all research, namely the argument from inductive risk (Douglas 2000, 2009; Elliott and Richards 2017). It is well suited to show the difficulty of tracing value influences in classic cases of basic research. To give a historical example: What trade-offs of risk of error did Robert Boyle have to make when he measured the volume of air in various states of pressure and very cautiously determined that the pairs of numbers he recorded were consistent with "the hypothesis, that supposes the pressures and expansions to be in reciprocal proportion"? (Boyle 1744 [1662], 101) According to the argument from inductive risk, accepting the hypothesis and communicating it are based not only on assumptions about the hypothesis itself and about the available evidence, but also about how high a tolerable inductive risk is permitted to be in the case at hand, and that question in turn is one that cannot be answered in a non-arbitrary manner without considering how severe the consequences would be if one were to falsely accept and communicate a finding. This is a very fundamental argument that should be applicable to any form of inductive science. Even in research that has no recognizable relevance to practical questions, judgments about acceptable risks of error must have some value basis. The puzzle cannot therefore be solved simply by pointing out that an

episode of basic research has no relevance for our ethical and political concerns (cf. Dupré 2007, 31)—not to mention the fact that historians of science point out that, on closer inspection, this is often not true in cases like Boyle's (Shapin and Schaffer 1985). Nevertheless, Boyle's neat and simple experiments, his very accurate measurement data, which consistently show less than 1% deviation from a hyperbolic function, raise the question of where the risks and, consequently, the value-laden decisions are hidden. This is especially true when compared to paradigmatic cases of value-ladenness, like Heather Douglas' (2000) example of laboratory rats that were exposed to dioxin for two years at various levels and then autopsied. Scientists examined slides of the rats' livers to determine if there were any tumors and, if so, whether they were malignant. As Douglas convincingly argued, this ultimately required them to balance the risk of false positives and the risk of false negatives in light of their choices' implications for subsequent research and public health policies. Are Boyle's experiments and his analysis of their results not quite clearly value-laden to a lesser degree in some sense? And if so, in what sense?

2. An abstract framework

I will start to approach this question by asking why scientific research needs to take significant inductive risks at all. To achieve the desired generality, so that it becomes possible to compare radically different areas of scientific research, I avail myself of a framework introduced by James Bogen and James Woodward (Bogen and Woodward 1988). In this framework, scientists seek to infer features of phenomena. Phenomena are understood to be events, processes, relations, and manifestations of properties that have stable and repeatable characteristics. The basis for scientists' inferences about phenomena is provided by the results of measurement and observation processes, which Bogen and Woodward refer to as "data."

Of course, inferences from data to phenomena are always subject to uncertainty. One very important source of this uncertainty is the presence of noise. Following Woodward (2010, 793), I will refer to features of the data caused by the phenomena as "signal" and features caused by other factors as "noise." I deliberately accept the somewhat imprecise way of speaking of "features" in order to preserve the generality I am aiming for. "Noise" in this sense can be many things: causal processes that have nothing to do with tumor formation but affect rat liver tissue in ways that visually resemble a malignant tumor; impurities and inaccuracies in the manufacture of a Torricellian tube—the simple type of barometer Boyle used in his experiments—which have

the effect that changes in the height of the mercury column are not exactly proportional to changes in atmospheric pressure.

A critical factor for the uncertainty in data to phenomena inference is how much the causal factors subsumed under "signal" affect the data in relation to the causal influence of the noise. In the following I will refer to this relation as "signal-to-noise ratio". For the purposes of this paper, it suffices to stipulate that we call the signal-to-noise ratio "more favorable" when the signal's influence is more prominent in the data, holding the noise level fixed, and "less favorable" when it is less so.

Of course, it is not the signal-to-noise ratio alone that determines how uncertain conclusions about the phenomena are. Even with noisy data, the reliability of inference can in principle be improved through more effort, for example by increasing the amount of data used. This may suggest that no matter the signal-to-noise ratio, it is a matter of choice which inductive risks science is willing to accept. If one applies very strict standards as to how probable it is supposed to be that one's conclusions from data-to-phenomena inference actually correspond to the facts, then one just abstains from drawing any conclusions until one has sufficiently broadened the data base to be able to achieve the desired reliability for a given signal-to-noise ratio.

3. The relevance of productivity: Is the level of reliability that scientific research aims for a matter of choice?

The fact that the business of strict standards is not always so simple can be illustrated by the example of agricultural research, which provides the background for Ronald Fisher's pioneering contributions to statistics. In his study of the genesis of Fisher's significance test method, Cornelis Menke (2016) has emphasized the relevance of two factors: first, that Fisher was strongly moved by considerations of research economics, and second, that at the time that he first developed his method he was employed as a statistician at an agricultural research institute. In order to draw conclusions from a comparative agronomic field trial with an experimental plot and a control plot, one must make assumptions about the comparability of the soil and other conditions of the plots. Ideally, these must be based on many years of experience with the yields of the plots. In terms of research economics, it makes no sense to keep the risk of statistical error

lower than the uncertainties about the comparability of the experimental plots. To push these in turn to a very low level, one would theoretically need hundreds of years worth of documented planting and yield cycles. Against this background, it is to be understood that Fisher says of the p-value of 0.05 that "[i]t is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not." (Fisher 1925, 47; Menke 2016, 134-135). A significant result is thus characterized as one in which we can clearly state that either the null hypothesis is false, or the data are a random result that is less than 5% likely.

A five percent probability that the data can occur if the null hypothesis is true is not exactly low. A practice of inferring phenomena from data that would be guided by Fisher's recommendation for identifying significant results is accordingly not very strict in its standards. Indeed, this is one of the reasons why the practice of using the p-value of 0.05 as the standard for publishable results is heavily criticized today. The context of agronomic research nevertheless makes Fisher's choice of this convention understandable. In agricultural research, there are limits to the application of ever stricter standards that are rooted in the subject matter of research itself. Increasing the sample size and thus the acreage exacerbates the problem of providing homogeneity of soil and growing conditions. Repeating experiments generally requires waiting a year.

In discussing inductive risks, it is commonly emphasized that methodological decisions involve balancing the desired reliability of positive outcomes against the desired reliability of negative outcomes. The example of Fisher and the feasible significance level for agronomic field research makes it clear that, in addition to this, a trade-off against another factor, which I call productivity, is also required. By productivity I mean the rate at which a given research effort produces any results at all in a given time—whether positive or negative (Wilholt 2022). A superficial consideration might lead to think that the willingness to accept a reduced reliability of results in return for higher productivity would mean a limited orientation towards the goal of truth. This is not the case, as productivity is a dimension of the search for truth just as the reliability of positive results and the reliability of negative results are. It is trivial to achieve high reliability if one is prepared to sacrifice productivity—for example by adopting a research strategy that only deals with obvious cases and suspends judgment indefinitely on all the rest.

In agricultural research, there are sources of noise that can only be controlled to a very limited extent, such as the many inhomogeneities of a crop field, or the distinct characteristics of individual organisms. In order to achieve minimal productivity at all, researchers have to accept noticeable inductive risks. In practically relevant contexts, of course, researchers must not only achieve minimal productivity, but have specific goals for the productivity of their research that are dictated by the practical needs of an application context: Application-relevant research has a greater or lesser degree of urgency. Productivity, reliability of positive results, and reliability of negative results must therefore be traded off against each other in a way that allows for noticeable and costly shifts between these parameters. Since all three of these are dimensions of science's orientation toward truth, there are no purely epistemic considerations that allow us to determine what exactly the trade-off must be. Only non-epistemic values can be the determining factor.

4. Improving the signal-to-noise ratio

For a fixed signal-to-noise ratio, epistemic risks can only be shifted between productivity, reliability of positive results and reliability of negative results—from one dimension of an investigation's truth-orientation to the other. However, there are areas of scientific research where it is quite possible to aim for a more favorable signal-to-noise ratio. Examples of this can be found in all laboratory sciences and especially in physics.

In some cases, it is quite simple steps by which the ratio of signal to noise is optimized in experimental practice, and the separation of signal from noise is made possible. In Boyle's (1744 [1662], 101-103) famous experiments that he used to establish the law now named after him, a certain small amount of air is contained within a glass tube partly filled with mercury, and is thus included in an enclosure bounded on one side by a mercury surface and on all other sides by rigid barriers. This confinement allows both experimentation with a specific, fixed amount of air and shielding against a whole range of causal influences on that portion of gas. The atmospheric pressure acting on the apparatus from the outside is controlled by being measured independently for each individual experiment in close temporal and spatial proximity. Manipulations of the system that cause changes in pressure and volume are performed extremely slowly and in small steps. This preparation creates a tiny, well-controlled experimental space. The phenomenon of air pressure is made controllable and manipulable, it is isolated from external influences in a

laboratory environment. Ernan McMullin has summarized such aspects of experimental practice, which amount to a foundational strategy of the modern experimental method, under the apt label of "causal idealization" (McMullin 1985).

A good decade after McMullin's reflections, Nancy Cartwright accentuated the observation that, especially in physics, experimentation often begins with extensive material preparation of the phenomenon by coining the term "nomological machine". The experimental apparatus facilitates "the repeated operation of a system of components with stable capacities in particularly fortunate circumstances" (Cartwright 1997, 65), which is what makes the exact application of mathematically formulated laws possible in the first place.

Even in physics, the application of causal idealization does not inevitably lead to inductive risks disappearing or becoming vanishingly small. But experimenters in physics often have the advantage that the way to further improve the signal-to-noise ratio by further causal idealization is open to them. In contrast, in the life and social sciences, there are often reasons lying in the subject matter that stand in the way of further improving the signal-to-noise ratio. Of course, the life sciences have fully embraced the experimental method, and in their case, too, causal idealization is evidently part of this method to some extent. But to take the shielding of a phenomenon against extraneous causal factors to extremes is often incompatible with the integrity of a studied organism. Sources of noise are often rooted in the causal complexity of the object under study itself and cannot be isolated or shielded without disturbing or even destroying the object.

This is at least sometimes different in physics. Here, under favorable circumstances, it may be possible to unravel the "tangle of causal lines", to use one of McMullin's (1985: 264) metaphors for causal idealization, by material preparation, and in the end to isolate a single or a few among them as a signal, with little or almost no noise. We always have implicit or explicit value-infused expectations about how reliable positive and negative results of a study need to be, and how quickly results need to come (i.e., in my terminology, how productive the study should be). If causal idealization has been very successful and the signal to noise ratio is very favorable, it may be possible to meet all of these expectations simultaneously with ease. Some methodological choices would then theoretically still shift the balance between productivity and reliability, but if these shifts occur in areas of reliability and productivity that far exceed our need

for reliability and productivity anyway, then they do not make a relevant difference and do not impose hard choices on scientists that require comparative judgments about competing value-infused goals. In this sense, a very crucial source of value-ladenness disappears for the investigations in question, and in this respect, it can be said that the more favorable the signal-to-noise ratio that can be achieved in an investigation, the less value-laden it is, ceteris paribus.

5. Some observations and clarifications

It is not only noise in the data and the resulting inductive risks in data-to-phenomena inference that pose a threat of false results in physics. Philosophers of science have been aware, at least since the times of Pierre Duhem, that a large number of assumptions and presuppositions go into physical experiments. The theoretical interpretation of the experimental setup could contain errors. Initial conditions or parameter values could be wrong. The experimental setup could contain an undetected technical mistake. Arguably, historians of science have shown how at least sometimes the kinds of assumptions I just mentioned are influenced by scientists' political, ethical, or social values. However, it would be quite hard to demonstrate that such assumptions *must* be influenced by values. By contrast, when it comes to weighing inductive risks—to the question of how precisely, say, the reliability of positive results ought to be traded off against the desired productivity—there simply *is* no rational alternative to starting from a judgment about how severe the consequences of a false-positive result would be. Therefore, this potential source of value-ladenness plays a special role, and therefore scientific research that enjoys the luxury of being able to work with a favorable signal-to-noise ratio is, in this respect, at least potentially value laden to a lesser degree. Of course, a favorable signal-to-noise ratio does not apply to all physical experiments, nor is it fundamentally limited to physics. But the most characteristic examples of such successful causal idealization come from physics, which is why I stick to physical examples in this paper.

The fact that the advantages of physics in this respect are due to its subject matter is not to be understood in such a way that they are inherent to the material stuff itself that physicists investigate. Rather, they depend on how well this stuff can be prepared for experimentation. Take, for example, the hydrodynamics of the earth's atmosphere. It is made up of the same stuff that is the subject of Boyle's experiments: Air under different pressure conditions extending into various regions of physical space. However, in the study of the actual hydrodynamics of our

atmosphere, one can by no means speak of a favorable signal to noise ratio, whereas in Boyle's experiments one can.

It is important to note that the same material manipulations that optimize the signal-to-noise ratio often widen the gap between internal and external validity, making near-value-free results possible in principle but often render them limited in their practical applicability at the same time. Internal validity in the sense that is relevant in this context is the truthful description of the operation of the nomological machine in the laboratory, while external validity means the (approximate) correctness of the result when applied to a practically relevant problem outside the laboratory.

Another important clarification is that everything really turns on the ability to infer the signal well and reliably from the background noise while avoiding overfitting to the data. The term signal-to-noise ratio, which I am using for simplicity, describes this a bit imprecisely. In particle physics, for example, there is a lot of noise. The particles that are caused by collisions are extremely short-lived. The detectors register the products of millions of decay events, and the task is to filter out from this abundance the traces of decay of rare, particularly interesting particles. The fact that in particle physics it is nevertheless possible to filter out the signal with very high reliability is due to the sheer gigantic amount of data that can be collected over a long time in the colliders. The diagram below shows how data from two-photon collisions were collected over a period of about a year and a half in the ATLAS Experiment at CERN. The visible bump in the fitted red curve that appears in the aggregated data at the end corresponds to the mass of the Higgs boson.
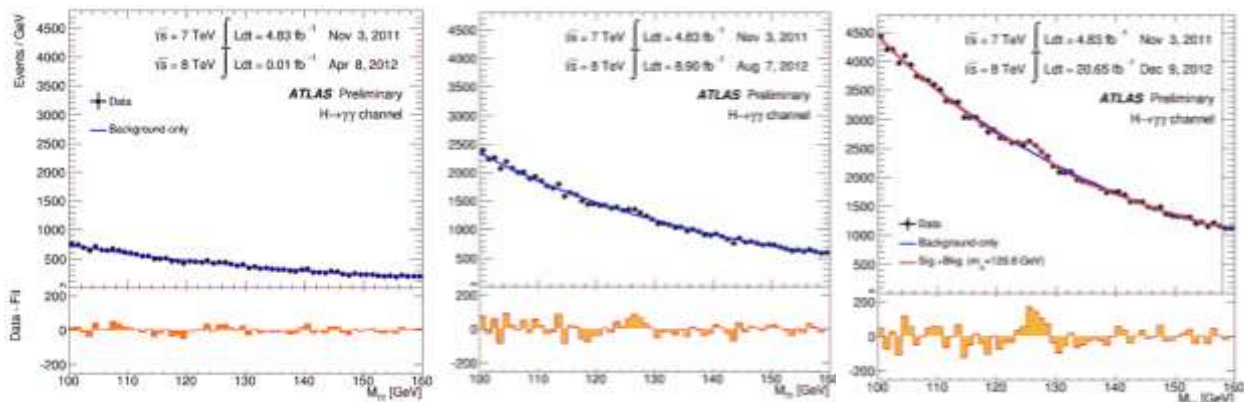
Snapshot from April 2012   Snapshot from August 2012   Final data from December 2012

Fig. 1: Aggregated data (mass spectra) from two-photon collisions at the ATLAS experiment, CERN.
Source: https://home.cern/resources/faqs/five-sigma. Used with permission.

ATLAS waited to publish the "discovery" of the Higgs boson until the data showed statistical significance of five standard deviations—a convention also known as "five sigma", or 5σ. Converted, this means a p-value of 0.0000003. This value indicates the probability with which, under the assumption that *no* Higgs bosons were created in the collisions, the bump in the curve could arise as a result of mere fluctuations in the background noise. The five sigma standard is conventionally required in particle physics to talk about a discovery; journals, for example, typically do not allow the use of the term "discovery" in title or abstract unless the data reach at least this level of statistical significance.

From the perspective of virtually any other discipline, especially any field from the life or social sciences, the five sigma standard marks an unattainable and almost fantastic level of rigor in the demands for reliability of data-to-phenomena inference. The fact that it is at all possible to meet such strict standards in physical experiments is what makes it seem to us as if there were no trade-offs or compromises required at all, and no room for value-ladenness. Had the ATLAS physicists waited a few months longer, they would have realized, in effect, even higher standards for the reliability of their discovery, at the cost of somewhat reduced productivity. But we do not perceive this as a value-laden choice, since our standards for productivity and reliability are not affected in any *relevant* way by one or two months' worth of data. Either way, our expectations are overfulfilled.

The example shows that, strictly speaking, a distinction can be made. The question of signal to noise ratio, i.e., according to our definition, how much the data are causally influenced

by the phenomenon in relation to other causal influences, can be asked once in relation to individual measurements and once in relation to the accumulated data that the circumstances under which the scientists work allow them to collect. The relevant sense in our context is always the latter—for data-to-phenomena inference is always made on the basis of the total, accumulated data. Where noise is unsystematic, the large-scale accumulation of data can make the signal stand out from it. In this sense, it can also be said of the ATLAS experiment that the huge number of measurements that are made makes it possible to realize a favorable signal to noise ratio, even though noise is considerable for each individual measurement.

6. Why a favorable signal-to-noise ratio does not mean value free science

Finally, I would like to warn against confusing a very low degree of value-ladenness in this very specific sense with the complete lack of relevance of non-epistemic values. This would be a serious mistake for two reasons. Firstly, the fact that trading off between reliability of positive outcomes, reliability of negative outcomes and productivity does not require us to make decisions where we have to accept falling short of the value-laden expectations we have set in one respect or the other, does not mean that such trade-offs do not need to take place at all. The choice of five sigma, for example, is not merely a matter of social coordination; its pros and cons can be discussed (see e.g. Lyons 2013, Staley 2017, and, for the history of the standard, Franklin 2013). What ultimately plays a role in such trade-offs, when they do not cut into markedly value-laden expectations one way or the other, are considerations of research economics (cf. Staley 2017, 50-51). If the cost to the research community of being led astray by a mistake, even a very unlikely one, is potentially extremely high, then it may be worth investing a few more resources (i.e. sacrificing productivity) to make the mistake even less likely. Such questions of research economics are too closely interwoven with economic issues in a more general sense to allow the values in question to be clearly marked as "purely epistemic".

The second and arguably even more important reason is that the possibility of achieving such favorable epistemic conditions depends on presuppositions that are themselves relevant to questions of value in the context of discussions of the ends and means of science. For in order to enjoy favorable signal to noise ratios one must be able to afford the necessary reduction in complexity and the time and material effort required to construct nomological machines. In particular, one must be able to afford to focus on phenomena that can be captured in such

nomological machines and thus may be far removed from the phenomena that matter in application contexts. Whether one studies more directly application-relevant questions that force one to make hard choices in weighing inductive risks, or application-remote nomological machines that do not, but on the other hand may take one further away from practically relevant questions, is itself a value-laden trade-off at the level of the choice of research question and amounts to its own form of value-ladenness. This form of value-ladenness is concealed when one focusses solely on trade-offs between different types of inductive risks, as such trade-offs always presuppose a given, very specific subject of research and cannot represent value-laden decisions that involve subtle or not so subtle shifts in the subject of research.

The false conclusion that research with favorable signal-to-noise ratios would no longer require any values at all would ultimately mean failing to recognize that there are many other forms of epistemic risks in research besides inductive ones (Biddle 2016; Biddle and Kukla 2017).

7. Conclusions

I have found that some kinds of research endeavors, because of the favorable signal-to-noise ratio they can achieve, do not require researchers to make hard choices in the sense of requiring them to compromise on reliability of positive results, reliability of negative results, or productivity relative to levels of these objectives that seem worth pursuing. In this particular respect, these episodes of research exhibit lower degrees of value-ladenness, and at least in this sense, then, degrees of value-ladenness may be said to exist. This explains why there are examples of scientific research where we find it extremely difficult to identify risks of error that need to be weighed against each other in value-laden decisions. I have also pointed out that achieving such advantageous signal-to-noise ratios requires the prior material preparation of experimental systems, and that in the course of these preparations, trade-offs involving shifts in the subject of research are implicitly made at another level, such as those between internal and external validity. That research which benefits from a favorable signal to noise ratio is in a certain respect value-laden to a lesser degree does therefore not mean that it is completely free of non-epistemic values.

References

Alexandrova, Anna. 2018. "Can the Science of Well-Being Be Objective?" *British Journal for the Philosophy of Science* 69:421-445. https://doi.org/10.1093/bjps/axw027

Biddle, Justin. 2016. "Inductive Risk, Epistemic Risk, and Overdiagnosis of Disease." *Perspectives on Science* 24 (2):192-205. https://doi.org/10.1162/posc_a_00200

Biddle, Justin, and [Quill] R. Kukla. 2017. "The Geography of Epistemic Risk". In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards, 215-237. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467715.003.0011

Bogen, James, and James Woodward. 1988. "Saving the Phenomena." *The Philosophical Review* 97:303-52. https://doi.org/10.2307/2185445

Boyle, Robert. 1744 [1662]. "A Defence of the Doctrine Touching the Spring and Weight of the Air". In *The Works of the Honorable Robert Boyle,* edited by Thomas Birch, Vol.1, 76-117. London: Millar 1744.

Cartwright, Nancy. 1997. "Where Do Laws of Nature Come From?" *Dialectica* 51(1): 65-78. https://doi.org/10.1111/j.1746-8361.1997.tb00021.x

Douglas, Heather E. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4):559-579. https://doi.org/10.1086/392855

Douglas, Heather E. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.

Dupré, John. 2007. "Fact and Value". In *Value-Free Science? Ideas and Illusions*, edited by Harold Kincaid, John Dupré, and Alison Wylie, 27-41. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195308969.003.0003

Elliott, Kevin C. 2011. *Is a Little Pollution Good for You? Incorporating Societal Values in Environmental Research*. New York: Oxford University Press.

Elliott, Kevin C., and Ted Richards, eds. 2017. *Exploring Inductive Risk: Case Studies of Values in Science*. New York: Oxford University Press.

Fisher, Ronald A. 1925. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Franklin, Allan. 2013. *Shifting Standards: Experiments in Particle Physics in the Twentieth Century.* Pittsburgh: University of Pittsburgh Press.

Lyons, Louis. 2013. "Discovering the Significance of 5 sigma." *arXiv preprint* arXiv:1310.1284. https://doi.org/10.48550/arXiv.1310.1284

McMullin, Ernan. 1985: "Galilean Idealization." *Studies in History and Philosophy of Science* 16 (3):247-273. https://doi.org/10.1016/0039-3681(85)90003-2

Menke, Cornelis. 2016. "Die Ökonomie der Forschung." Habilitation thesis. Universität Bielefeld.

Shapin, Steven und Simon Schaffer. 1985. *Leviathan and the Air-Pump: Hobbes, Boyle and the Experimental Life.* Princeton: Princeton University Press.

Staley, Kent W. 2017. "Decisions, Decisions: Inductive Risk and the Higgs Boson." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards, 37-56. https://doi.org/10.1093/acprof:oso/9780190467715.003.0003

Steel, Daniel, and Kyle Powys Whyte. 2012. "Environmental Justice, Values, and Scientific Expertise." *Kennedy Institute of Ethics Journal* 22:163-182. https://doi.org/10.1353/ken.2012.0010

Wilholt, Torsten. 2009. "Bias and Values in Scientific Research." *Studies in History and Philosophy of Science* 40:92-101. https://doi.org/10.1016/j.shpsa.2010.03.003

Wilholt, Torsten. 2022. "Epistemic Interests and the Objectivity of Inquiry." *Studies in History and Philosophy of Science* 91:86-93. https://doi.org/10.1016/j.shpsa.2021.11.009

Winsberg, Eric. 2012. "Values and Uncertainties in the Predictions of Global Climate Models." *Kennedy Institute of Ethics Journal* 22:111-137. https://doi.org/10.1353/ken.2012.0008

Woodward, James. 1988. "Data, Phenomena, Signal, and Noise." *Philosophy of Science* 77:792-803. https://doi.org/10.1086/656554