Philosophy of Science (2025), **92**, 1417–1426
doi:10.1017/psa.2025.10153

PSA PHILOSOPHY OF SCIENCE ASSOCIATION
PHILOSOPHY OF SCIENCE

CONTRIBUTED PAPER

Are There Experimental Virtues?

Peter Mättig[1] and Michael Stöltzner[2]

[1]Department of Physics, University of Bonn, Bonn, Germany and [2]Department of Philosophy, University of South Carolina, Columbia, South Carolina, USA
Corresponding author: Michael Stöltzner; Email: stoeltzn@sc.edu

(Received 21 January 2025; revised 23 April 2025; accepted 24 July 2025; first published online 27 August 2025)

Abstract

Philosophers have long discussed the virtues of scientific theory and the role they play in pursuing, preferring, or finally accepting a theory on the basis of internal criteria and experimental evidence. Based on two case studies from elementary particle physics, we propose a list of experimental virtues and discuss how they influence experimental strategies and whether a measurement is deemed conclusive—independently of its relation to theory.

1. Introduction

There are episodes in the history of science where multiple theories have provided equally good descriptions of empirical phenomena, sometimes over extended periods of time. Philosophers of science have long debated whether this underdetermination of theory by empirical data, first noted by Duhem, poses problems for the justification of scientific theories. Examining how this problem has been addressed in scientific practice, Duhem (1954) himself famously invoked "*bon sens*" as the judge of which hypotheses should be abandoned if a conflict with experiment occurs. He was quite optimistic: "The day arrives when good sense comes out so clearly in favor of one of the two sides such that the other side gives up the struggle even though pure logic would not forbid its continuation" (218). Today philosophers of science typically approach the difference between confirmation and acceptance by assuming that a theory accepted by the scientific community possesses virtues beyond its formal consistency and empirical adequacy. These theoretical virtues are sometimes understood as criteria of rational theory choice among empirically adequate theories. Those who, following Kuhn, stress their contextuality and primarily pragmatic function prefer to speak of values (or virtues) of theory preference. Kuhn's (1977, 322) influential list includes "accuracy, [internal and external] consistency, scope, simplicity, and fruitfulness." This list is often considered as the common core of the broad literature on theoretical virtues.

The present article argues that there are experimental virtues as well. Despite the long tradition of debates about theoretical virtues and the growing literature on the

© The Author(s), 2025. Published by Cambridge University Press on behalf of Philosophy of Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

philosophy of experiment, this topic has hardly been addressed to date. The question of experimental virtues must be distinguished from the question of "what makes a good experiment" (Franklin 2016) and concerns about experimental methods and technologies (cf. Franklin 2013).

The core function of the theoretical virtues lies in motivating which of two empirically equivalent theories should be accepted, whether scientists consequently should devise alternatives for empirically adequate theories with a bad score in these virtues, and whether the prospect to score well makes a research program worthy of pursuit. In the same vein, experimental virtues inform experimentalists, which among several potential ways of measuring the same process or object are easier to access and can more convincingly be defended against criticism. As such they are strongly correlated with minimizing systematic uncertainties. Whether experimental evidence is conclusive is affected by experimental virtues, which improve the potential of detecting a significant signal.[1] Although we will find some analogies between theoretical and experimental virtues, each is of specific epistemic value. Experimental virtues also apply to experimental research that is not motivated by, or even connected with, any theoretical hypothesis.

We illustrate our conception of experimental virtues with examples from elementary particle physics. As a matter of fact, this field has often been mentioned in the debates about theoretical virtues and the philosophy of experiment. Modern particle detectors can be seen as collaborative multipurpose experiments that allow experimentalists to probe a vast number of different questions, and for many of these questions, different signatures can be used. Physicists' decisions about what to measure and which measurements to accept as conclusive provide ample case studies for the function of experimental virtues.

We identify a set of experimental virtues that are shared across particle physics and can be distinguished from general methodological concerns and instrumental constraints. In the same way as the theoretical virtues in Kuhn's list, experimental virtues are not mutually exclusive, they are contextual, and different physicists may weigh them differently. Our nonexhaustive list includes (1) the uniqueness of a signature, (2) the precision of a signature, (3) the simplicity of the signature, (4) the independence of systematic uncertainties from model assumptions, and (5) the broadness of sensitivity in the sense that the experiment covers a broad range of phenomena.

Our article proceeds as follows. In section 2, we review the present debate about theoretical virtues to motivate our concept of experimental virtues. In section 3, we discuss two examples from the history of elementary particle physics. On the basis of these case studies, we suggest in section 4 a nonexhaustive list of experimental virtues. In section 5, we distinguish experimental virtues from other characteristics of good experiments. The final section concludes.

2. Theoretical virtues: An overview

Despite the considerable body of literature on theoretical virtues, the terminology remains unsettled. Shifting the focus from normative questions of rational choice to

[1] In particle physics, there is a convention to speak of 'observation' if a 5σ threshold is met, and of 'evidence' if a 3σ threshold is met.

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

factual investigations of scientific practice, Kuhn suggested speaking of values that are contextual, are sometimes idiosyncratic, and pull in different directions. Schindler (2018, 5n5) prefers virtues because they characterize theories rather than scientists and have no ethical connotations. Douglas (2013) follows Kuhn (1977) in talking about values but sets these cognitive values apart from the social values involved in decisions on how to implement them (Douglas 2017). We have decided to speak of virtues because experimental virtues function as dispositions to act; that is, they motivate how to proceed in an experiment and whether to consider some experimental evidence as conclusive. This terminological choice notwithstanding, we will not discuss whether experimental virtues can be interpreted as an instance of virtue epistemology (cf. Stump 2007; Ivanova 2010).

There is widespread agreement that theoretical virtues are part and parcel of what Peter Achinstein (1993) has aptly called the logic of pursuit. But they also influence the factual acceptance of theories. Virtues are thus applied retrospectively as justifications of theories and prospectively as motivations to develop them. This double function lies at the bottom of why Kuhn (1977) believed that they would allow us to maintain a notion of scientific progress across scientific revolutions. Philosophers of science have also debated whether theoretical virtues are epistemic or pragmatic, cognitive or social; how they are distinguished from other methodological rules; and whether they can be justified historically or a priori. The result is a surprising richness and variety in the lists of virtues proposed by philosophers. It is also controversial how different virtues should be weighted and whether they lend themselves to an objective procedure of rational choice (Kaes 2018) or are just informing individual scientists' actual choices and preferences (Kuhn 1977).

General philosophical orientations often influence the specific lists of theoretical virtues put forward in the literature. Constructive empiricists, such as van Fraassen (1980), hold that only epistemic virtues, empirical adequacy and consistency, are rationally compelling, while other virtues, among them simplicity and fertility, are essentially pragmatic. Scientific realists instead suggest a broader list of epistemic virtues and take them as indicators for a theory's truth (Schindler 2018). Psillos (1999, 171) lists "coherence with other established theories, [Whewellian] consilience, completeness, unifying power, lack of ad hoc features and capacity to generate novel predictions." Schindler (2018, 6) amends Kuhn's aforementioned "five standard virtues" by testability and not being ad hoc. Douglas (2013) separates the traditional list into minimal epistemic criteria, among them consistency and empirical adequacy and strategic or pragmatic desiderata, but emphasizes that by measuring only the strength of evidential support, "they don't help with deciding whether the evidence is sufficient" (Douglas 2017, 2) in a certain social and historical context. Even authors deliberately adopting a bottom-up approach (cf. Tulodziecki 2013) still have to disentangle the respective virtues from other, more general methodological principles.[2]

Kuhn viewed the five characteristics of a good scientific theory mentioned in the foregoing discussion as not mutually independent and influenced by contextual and idiosyncratic (psychological) factors. But theory choice was not a matter of taste. Kuhn's point was the factual nature of theory choice, not its arbitrariness. Historians

[2] See, for example, her analysis of John Snow's successful investigation on the origin of cholera.

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

often find an increasing unanimity of individual choices in a certain field—as Duhem put it, *bon sens* prevails in the long run.

More than other authors, McMullin (2014) emphasizes the role of diachronic theoretical virtues that concern a theory's past or future performance. Among them are the fertility to produce novel predictions or cope with anomalies and Whewellian consilience, which McMullin specifies as the power of unifying phenomena previously thought to be disparate.

Without attempting a synthesis of the broad debate concerning theoretical virtues, let us identify five joint characteristics:

1. Theoretical virtues are quality criteria of scientific theories. They guide the scientists' choice between competing theories without being rationally compelling.
2. While they are contextual and adaptable, they persist across foundational changes in scientific disciplines.
3. Even if some theoretical virtues can be seen to support a specific philosophical agenda, they have typically been shaped by the history of scientific practice.
4. Theoretical virtues guide scientists' approach to individual problems.
5. Theoretical virtues presuppose that the methodological standards of the respective science are properly followed.

We argue that along similar characteristics there exist experimental virtues.

3. Virtues in experimental arguments: Case studies from particle physics

Experimentalists are faced with the question of what problem to address with limited resources. Given the vast number of possible measurements and experimental strategies, how do they set their priorities, and how do they evaluate the eventual quality of a measurement? We have presented a first outline of these problems in Mättig and Stöltzner (2020). Notably, these strategies are to a large extent determined by the possibilities to isolate and control experimental signatures. Signatures are reconstructed from raw data of measurements and constitute stable and repeating patterns. Signatures are the direct input into experimental analyses (black-box theory) and are stable against theory change. Signatures introduce an order into the manifold of data without any high-level theory but are often conceptualized into phenomena. Such signatures may be seen as specific instances of a phenomenon, but they can be targets of model-independent measurements.[3]

3.1 The discovery of weak neutral currents

The discovery of neutral currents has often been discussed by philosophers in the context of the relationship between theory and experimental data (cf. Bogen and Woodward 1988; Schindler 2018). Our own analysis takes into account Galison's (1987) discussion of the Gargamelle experiment at CERN and the E1A experiment at (F)NAL,

[3] Compare Mättig and Stöltzner (2025) for a philosophical discussion of the relationship of signatures and phenomena.

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

today Fermilab, which both used neutrino beams and eventually discovered neutral currents (NC) in the process:

$$\nu_\mu/\bar{\nu}_\mu + \text{hadrons} \rightarrow \nu_\mu/\bar{\nu}_\mu + \text{hadrons}.$$

When the two experiments were planned, NC in themselves were considered of little interest; instead, many different experimental and theoretical "interests were woven together concisely into a broad experimental program" (Galison 1987, 159). The interest in weak NC grew significantly in the early 1970s in the context of the Salam–Weinberg model. The two experiments were in principle in a good position to search for NC because of their experimental virtues.

First, although NC were expected to be more frequent in hadronic interactions, the experiments were privileged to be particularly sensitive to weak interactions, the only interactions in which neutrinos participate. In contrast, hadrons interact mostly by the strong interaction making a possible signal of NC events being swamped by a huge number of background events: A signal of neutral currents becomes (almost) unmeasurable.[4] In contrast, the event signatures in neutrino interactions stand out. Second, the size of the detectors and their broad sensitivity to multiple signatures allowed them to address the double challenges of an expected low event rate and a significant background.

Third, this allowed both experiments to estimate the background with complementary data taken from other processes within their experiments and with minimal reliance on theoretical models. Even when first results made some experimentalists believe "that there was a real effect" (Galison 1987, 173), there were several possibilities of background processes, especially due to neutrons, that could fake NC. While this was first addressed using models, which were inherently uncertain, major progress was reached when the Gargamelle collaboration "presented a method for estimating the neutron background entirely from the characteristics of events in the visible volume of the chamber ... [which] avoided many assumptions about the distribution of matter around the machine or the flux of neutrinos into the experimental area" (189). A key strategy was "minimizing dependence on the calculated quantities of the computer simulation" (237), instead deriving the needed justification of the background estimates predominantly from data. In contrast, the parallel E1A experiment attempted to disentangle their analysis from the dependence on the Monte Carlo model and from the "not-yet secure parton model" (219).

Fourth, the detected signature consisting of just hadrons along the direction of the incoming neutrino, apparently without an additional particle, is fairly simple and thus easy to identify and convincing. In turn, the Gargamelle neutrino interactions allowed a precise measurement of the cross section and fundamental parameters of the Salam–Weinberg model. It should also be noted that the conclusion on the existence of NC was supported by Gargamelle's subsequent finding of a complementary process, $\bar{\nu}_\mu + e \rightarrow \bar{\nu}_\mu + e$, with an even simpler and more convincing signature but a lower event rate.

[4] Even some fifty years after their discovery, neutral currents in purely hadronic interactions are just barely visible, except for in special circumstances.

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

3.2. The discovery of the Higgs boson at the Large Hadron Collider

How experimental choices are made becomes especially transparent in modern multipurpose experiments, for which the Large Hadron Collider (LHC) stands prototypical. Let us briefly discuss the experimental pathways to arrive at the Higgs discovery. Theory predicted several decay modes for a Higgs boson at 125 GeV (bb, ZZ, WW, γγ, ττ, Zγ, ...).[5] Theoretically, the Higgs should yield, in all decay modes, a narrow enhancement in the invariant mass of its decay products. However, its existence was established mainly by its decay into two photons and two leptonically decaying Z-bosons, decay modes that in fact were expected to be rather rare. Why?

Each of the particles into which the Higgs decays is related to a characteristic "particle signature," which is operationally defined through the signals in the multilayered LHC detectors. For example, an electron yields a narrow signal in the electromagnetic calorimeter with signals in the tracking chamber of corresponding properties, for example, the momentum in the chamber should be the same as the energy in the calorimeter. A photon γ yields a narrow signal in the electromagnetic calorimeter but leaves no trace in the tracking chamber. A bottom quark b leads to a narrow bundle of particles in the tracking chamber with a displaced vertex a few millimeters away from the main pp interaction point. Additionally, it induces signals in the electromagnetic and hadronic calorimeters. To optimize the identification of bottom quarks, several parameters are used and integrated into a neural network. All these signatures are essential tools of LHC physicists in analyzing the debris of pp collisions. They differ in their efficiency, purity, and measurement precision. Experimental preferences for, say, certain Higgs decay modes are synonymous with preferences for certain event signatures, for which particle signatures are combined.

From a theoretical perspective, one might have expected that the dominant search for the Higgs would have been in the bb decay, because this was expected to be by far the dominant decay mode. In contrast, the actual discovery channels had an expected decay rate that was more than 100 times smaller than the bb. Similarly to the NC searches, the ease and clarity of the di-photon and di-Z channel with leptonic decays made these the discovery channels. The experimental preference for these signatures can be summarized as follows:

- The apparent broadness of the peak is small for the discovery channels but broad for the bb decay. This leads to a prominent signal in the di-photon and leptonic ZZ channels, while the enhancement in the bb channel is washed out. The reasons for the advantage of the discovery channels are, on one hand, the detection method—the narrow signatures of both photons and leptons, and, on the other hand, the physics of bottom particles that, as discussed, leads to a broad signal in the detector.
- Di-photons, di-Zs, and also pairs of bottom quarks at the LHC are produced independently of the Higgs decay. It is important to distinguish these from the Higgs decay and the continuum production (background). The di-photon background is rather high, and the ZZ decays into electrons/muons are very rare, however, bottom production is abundant at the LHC. Together with the

[5] Here b refers to the bottom quark; Z, γ, and W to the electrically neutral and charged electroweak bosons; and τ to the heaviest electrically charged lepton.

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

apparent broadness of the peak, the Higgs decay into two photons and in particular into ZZ leads to a clearly visible enhancement above the background —the Higgs decay into two bottom quarks is swamped by the background.
- The shape of the background beneath the γγ peak can be smoothly interpolated without even invoking a physics model. This is largely true also for the ZZ background, while the bb background sits on top of a tail of another particle (the Z-resonance), being thus a mixture of several processes and requiring special care.
- The different signatures are also different in terms of simplicity and transparency. Electrons, muons, and photons are single entities, easily identifiable for both signal and background. This is different for the bottom quark, which is identified in multivariate analyses at the price of less transparency.

These properties have little relation to theoretical expectations but predominantly reflect experimental qualities; they lead to experimental preferences and are thus reflections of experimental virtues. Even if theoretically the Higgs decay into bottom quarks is the most frequent decay, its poor experimental virtues are compensated for by the other rarer, but more pronounced, decay channels. Thus experimental virtues may be at odds with what a purely theoretical consideration suggests.

Let us conclude with two observations. First, after the announcement of the discovery of the Higgs boson in 2012, experimentalists continued to study the other channels in the context of investigations into whether the new boson had all the properties predicted by the Standard Model. Second, although the properties of experimental signatures can, to a certain extent, be better measured by improved and specialized technologies, they are largely inherent to the particles themselves. Owing to its nature, a bottom quark jet is much more difficult to identify than an electron. An electron can be measured independently in two different parts of the detector where it leaves narrow and redundant marks, allowing for important constraints. A bottom particle, in contrast, is identified through a whole bundle of some fifteen to twenty particles that spread out in space and have rather different momenta. These particles interact differently, and not redundantly, in the different detector components such that their momenta have to be combined in a complicated way.

4. A list of experimental virtues

In both examples, one sees a common set of experimental conditions that physicists strive for and cite when accepting a measurement as conclusive, although the concrete problems and technologies are different. We consider them as experimental virtues. They are relevant in planning the detector, devising a strategy of analysis, and accepting a result.

(a) *Uniqueness of a signature*: The signature should ideally be a signal of exclusively one phenomenon. However, in most cases, noise or background phenomena leading to the same signature cannot be avoided such that, instead, the signal-over-background ratio should be high: The signal should stand clearly above the background (noise).
(b) *Precision of the signature*: All elements of an analysis should be measurable with high precision, and the detector should lead only to minimal distortions of the

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

objects. The remaining (unavoidable) distortions should be well understood and their systematic uncertainties minimal.

(c) *Simplicity of the signature*: The signature should be simple, that is, its reconstruction should involve a minimal number of steps and only a minimum number of parameters. By such simplicity, the signature is transparent both for the analyses and for the scientific community in general.

(d) *Systematic uncertainties*: The systematic uncertainties should rely only minimally on model assumptions. The detector should be constructed such that most of the uncertainties can be derived from data themselves.

(e) *Broadness of sensitivity*: The experiment should allow one to cover a broad range of phenomena to help cross-checking, avoid bias, and thus be open to unexpected phenomena.

These experimental virtues are autonomous from theoretical considerations. Still, they match the general characteristics of theoretical virtues outlined at the end of section 2, which provides the motivation for denoting them as virtues: Experimental virtues guide experimentalists' choices without being rationally compelling. They were developed by experimental practice and persist across major changes in measurement devices and methods (although they may drive these developments). Albeit contextual, they are chosen on top of the standard practices of good experimentation.

We do not exclude that other virtues can be identified; however, we emphasize that experimental virtues are different from general experimental methods that define conditions that are mandatory for a trustworthy experiment. As are theoretical virtues, experimental virtues are not mutually exclusive. For example, the more precisely one can measure a signature, the better is the signal-to-background discrimination. This becomes apparent when comparing the narrow di-photon signal to the broad bb signal: The background is sizable for both, but the precisely measurable $\gamma\gamma$ signal stands out much more prominently than the broadly distorted bb mass. Furthermore, as for theoretical virtues, experimental ones may pull into different directions. Complicated signatures, for example, based on neutral networks (artificial intelligence) may lead to high efficiencies and reduce background, but can be opaque and render the estimation of the systematic uncertainty more difficult.

We can also diagnose some direct analogies between these experimental virtues and theoretical virtues. We find simplicity in analogy to (c) and breadth of scope in analogy to (e). One may also read (d) as a consistency of experimental practice in the sense that a measurement can eventually be justified without reliance on simulations or theoretical models.[6]

5. Distinguishing experimental virtues from other classifications

In conclusion, let us compare our experimental virtues to some related concepts that have been proposed in the analysis of similar experiments.

[6] Most philosophers count empirical adequacy among the epistemic virtues. We decided not to classify significance as an experimental virtue even though it plays some role if experimenters find that a theory is empirically adequate. The decisive point is that significance has an entirely different function than experimental virtues that does not depend on whether experiments test theories.

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

Galison (1987, 259–60) distinguishes two different axes of experimental strategies: directness and stability:

> By directness I mean all those laboratory moves that bring experimental reasoning another rung up the causal ladder: measurement of a background previously calculated ... or the separate measurement of two sources of an effect previously only measured together .... By "stability" I have in mind all those procedures that vary some feature of the experimental conditions: changes in the test substance, in the apparatus, or in the data analysis that leave the results basically unchanged .... Each variation makes it harder to postulate an alternative causal story that will satisfy all the observations.

Directness goes in the direction of our fourth virtue in the sense that it eliminates mediating models, the rungs in Galison's picture. Galison's stability is a standard experimental method to become sensitive to possible systematic uncertainties and to estimate them; it is not an experimental virtue.

In their analysis of epistemic superiority claims of experiments over observations, Boyd and Matthiessen (2024, 123) argue for "crosscutting" features: "higher signal clarity, better characterization of backgrounds, and higher discrimination and variability of precipitating conditions improve the epistemic outcomes of empirical research. Methods that better promote signal clarity increase the precision, accuracy, and confidence of an empirical result. Methods that better account for backgrounds prior to or after the recording of data will reduce systematic error." This parallels some of our experimental virtues, but Boyd and Matthiessen never interpret these features as such, nor do they distinguish them from other crosscutting methods, among them good measurements, or outline their common properties.

On the basis of a detailed analysis of particle physics experiments, Franklin (2013, 9; 2016, 4) lists eight and ten strategies for "good experiments," respectively. These include the elimination of possible errors and the exclusion of alternative explanations; but one also finds general inferential strategies, or basic conditions of experimental practice, such as blind analysis. Franklin lists standard methods and criteria for an experiment to be reliable and thus preconditions for trusted experimental results. We add also that the statistical conventions to deem a measurement significant can be seen as belonging to standard experimental practice. These practices are rather different from our virtues: Each measurement of the Higgs decay fulfills Franklin's strategies, but each has different experimental virtues.

We have developed the five experimental virtues by analyzing two case studies from elementary particle physics. But we believe that they are general enough to be found in other experimental fields as well, possibly with amendments and reformulations. Even if the dependence on models were stronger there than in particle physics, (d) would still remain a desideratum. Virtues are also important where they are not already achieved but provide valuable guidance.

6. Conclusion

We have argued that in the same way as there are theoretical virtues in science, one can identify a set of experimental virtues. They are key criteria for choosing

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press

experimental strategies and accepting measurements as conclusive. They are complements to other criteria, especially the perceived relevance and significance of the result. As such, they do not prevent experiments to measure signatures of lower experimental virtues as much as theories and models may be developed that score low in theoretical virtues. Experimental virtues underline the autonomy of experiments and may also be a good reason to embark on experimental studies, even without any theoretical motivation.

Acknowledgment. Our research was funded by the German Research Foundation (FOR 2063).

References

Achinstein, Peter. 1993. "How to Defend a Theory Without Testing It: Niels Bohr and the Logic of Pursuit." *Midwest Studies in Philosophy of Science* 18: 90–120. https://doi.org/10.1111/j.1475-975.1993.tb00259.x.

Bogen, James and James Woodward. 1988. "Saving the Phenomena." *The Philosophical Review* 97: 303–352.

Boyd, Nora M., and Dana Matthiessen. 2024. "Observations, Experiments, and Arguments for Epistemic Superiority in Scientific Methodology." *Philosophy of Science* 91: 111–31. https://doi.org/10.1017/psa.2023.101.

Douglas, Heather. 2013. "The Value of Cognitive Values." *Philosophy of Science* 80: 796–806. https://doi.org/10.1086/673716.

Douglas, Heather. 2017. "Why Inductive Risk Requires Values in Science." In *Current Controversies in Values and Science*, edited by Elliott Kevin and Steel Daniel, 81–93. Routledge.

Duhem, Pierre. 1954. *The Aim and Structure of Physical Theory*. Princeton University Press.

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

Franklin, Allan. 2016. *What Makes a Good Experiment? Reasons and Roles in Science*. University of Pittsburgh Press.

Galison, Peter. 1987. *How Experiments End*. University of Chicago Press.

Ivanova, Milena. 2010. "Pierre Duhem's Good Sense as a Guide to Theory Choice." *Studies in History and Philosophy of Science* 41: 58–64. https://doi.org/10.1016/j.shpsa.2009.12.009.

Kaes, Michael N. 2018. "Systematizing the Theoretical Virtues." *Synthese* 195: 2761–93. https://doi.org/10.1007/s11229-017-1355-6.

Kuhn, Thomas S. 1977. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension, 310–39*. University of Chicago Press.

Mättig, Peter, and Michael Stöltzner. 2020. "Searching for Signatures." *Philosophy of Science* 87: 1246–56. https://doi.org/10.1086/710618.

Mättig, Peter, and Michael Stöltzner. 2025. "Phenomana and Signatures." *European Journal for Philosophy of Science* 15: 34. https://doi.org/10.1007/s13194-025-00661-5.

McMullin, Ernan. 2014. "The Virtues of a Good Theory." In *The Routledge Companion to Philosophy of Science*, edited by Martin Curd and Stathis Psillos, 561–71. Routledge.

Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth*. Routledge.

Schindler, Samuel. 2018. *Theoretical Virtues: Uncovering Reality Through Theory*. Cambridge University Press.

Stump, David J. 2007. "Pierre Duhem's Virtue Epistemology." *Studies in History and Philosophy of Science* 38: 149–59. http://doi.org/10.1016/j.shpsa.2006.12.010.

Tulodziecki, Dana. 2013. "Underdetermination, Methodological Practices and Realism." *Synthese* 190: 3731–50. http://doi.org/10.1007/s11229-012-0221-9.

van Fraassen, Bas. 1980. *The Scientific Image*. Clarendon Press.

Cite this article: Mättig, Peter and Michael Stöltzner. 2025. "Are There Experimental Virtues?" *Philosophy of Science* 92 (5):1417–1426. https://doi.org/10.1017/psa.2025.10153

https://doi.org/10.1017/psa.2025.10153 Published online by Cambridge University Press