



Reliability: an introduction

Stefano Bonzio¹ · Jürgen Landes² · Barbara Osimani³

© The Author(s) 2020

1 Introduction

How we can reliably draw inferences from data, evidence and/or experience has been and continues to be a pressing question in everyday life, the sciences, politics and a number of branches in philosophy (traditional epistemology, social epistemology, formal epistemology, logic and philosophy of the sciences). In a world in which we can now longer fully rely on our experiences, interlocutors, measurement instruments, data collection and storage systems and even news outlets to draw reliable inferences, the issue becomes even more pressing.

While we were working on this question using a formal epistemology approach Landes and Osimani (2020); De Pretis et al. (2019); Osimani and Landes (2020); Osimani (2020), we realised that the width of current active interests in the notion of reliability was much broader than we initially thought. Given the breadth of approaches and angles present in philosophy (even in this journal Schubert 2012; Avigad 2021;

Barbara Osimani is the PI of the European Research Council-founded project PhilPharm and gratefully acknowledges being fully funded by the project (Grant 639276). Stefano Bonzio gratefully acknowledges funding from the European Research Council ('PhilPharm' grant 639276) and the PRIN 2017 project "From models to decisions" funded by the Italian Ministry for University, Education and Research (MIUR). Jürgen Landes gratefully acknowledges funding from the European Research Council ('PhilPharm' Grant 639276) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-405961989 and 432308570.

✉ Stefano Bonzio
stefano.bonzio@unito.it

Jürgen Landes
juergen_landes@yahoo.de

Barbara Osimani
barbaraosimani@gmail.com

¹ Department of Philosophy and Education Sciences, University of Turin, via Sant'Ottavio 20, 10124 Turin, Italy

² Munich Center for Mathematical Philosophy, LMU Munich, Ludwigstr. 31, Room 129, 80539 Munich, Germany

³ Department of Biomedical Sciences and Public Health, Marche Polytechnic University, 60126 Ancona, Italy

Claveau and Grenier 2019; Kummerfeld and Danks 2014; Landes 2021; Trpin et al. 2021; Schippers 2014; Schindler 2011; Kelly et al. 2016; Mayo-Wilson 2014; Olsson and Schubert 2007; Pittard 2017), we thought that it would be beneficial to provide a forum for an open exchange of ideas, in which philosophers working in different paradigms could come together. Our call for expression of interest received a great variety of promised manuscripts, the variety is reflected in the published papers. They range from fields far away from our own interests such as quantum probabilities de Ronde et al. (2021), evolvable software systems Primiero et al. (2021), to topics closer to our own research in the philosophy of medicine Lalumera et al. (2020), psychology Dutilh et al. (2021), traditional epistemology Dunn (2021); Tolly (2021) to finally close shared interests in formal epistemology Romero and Sprenger (2021) even within our own department Merdes et al. (2021).

Our job is now to reliably inform you about all the contributions in the papers in this special issue. Unfortunately, that task is beyond our capabilities. What we can and instead do is to summarise the contributed papers to inform your reliable inference to read them all in great detail.

2 Contributions

2.1 Traditional epistemology

In line with the focus on the semantic analysis of knowledge claims, typical of contemporary analytic epistemology, Tolly (2021) shifts the focus from the definition of knowledge, to that of reliable knowledge, or, more precisely, knowledge-enabling reliability. Tolly draws on the notion of multi-type evidential reliabilism (MTE) and the standard toolset of counterfactual semantics, in order to identify the truth conditions for a learning process to be defined as reliable: knowledge-enabling reliability requires a process token to be reliable with respect to multiple content-evidence pairs (i.e., the propositional content of the knowledge items, and the related evidential basis), each with varying degrees of specificity. By using the notion of similarity across worlds and developing a model that assigns multiple types to any given token of knowledge process, MTE describes reliability as a concentric multi-type structure: every relevant type that belongs to a given token has a different degree of specificity. These differing degrees of specificity are defined by a similarity relation and measurements of conceptual difference. A token's specific content-evidence pair is something akin to an 'anchor-point' for determining all of the other relevant types. By describing the multiple relevant types that determine whether a given token generates knowledge-enabling reliability, MTE successfully meets the specificity challenge, a formidable objection to reliabilism Feldman (1998), while at the same time retaining Comesaña's insight Comesaña (2006) that reliability crucially depends on a token's precise content-evidence pair. Readers interested in the context-sensitivity of causation and the related reference class problem will find interesting family resemblances with the debate sketched in this contribution.

The question of how to aggregate judgements of a plurality of agents into a rational collective judgement represents an intriguing and relevant problem in epistemology

and in the theory of social choice. Apart from the epistemological nature of the intrinsic theoretical challenges arising in the discipline of judgement aggregation [see for instance List and Puppe (2009)], its pragmatism is mainly due to the high stake of decisions that may involve certain groups. Example of such group-decisions include, for instance, those taken by boards of experts called to decide whether a certain drug has to be approved for the market (after testing) or not, or the Federal Reserve Board believing whether or not the interest rates should be raised.

Dunn (2021) focuses on the question whether group beliefs are justified and reliable, assuming that one can speak about group beliefs. Indeed, there is no agreement concerning how group beliefs are formed. For this reason, Dunn (2021) relies on the idea proposed by Goldman (2014) that “group beliefs are formed according to some *belief aggregation function*”. In particular, Goldman’s account can be sketched via the so-called Social Process Reliabilism (SPR), which is founded on the conditional reliability of the aggregation procedure (one can think at “simple majority” as an example of aggregation rule). Conditional reliability of the aggregation rule can be rephrased as follows: “when enough of input beliefs have a high probability of being true, the output group belief is usually true”. In words, this means that the reliability of group beliefs depends substantially more on the reliability of the single judgements than on that of the aggregation rule that has been used. A point that indeed leaves room for criticism. The new position defended by Dunn (2021) is opposite: “a group belief is justified just in case the group belief-forming procedure as used by the group in question is reliable.” Roughly speaking, it is the reliability of the aggregation operators that becomes more important than single agents’ judgements.

2.2 Formal epistemology and the replication crisis

There are two main formal models of source reliability and confirmation in the current literature on Bayesian epistemology Bovens and Hartmann (2003); Olsson (2011). Merdes et al. here provide an analysis of normative components of the models and they investigate the accuracy of the models by means of large-scale computer simulations. While the model in Bovens and Hartmann (2003) only has reliable agents (perfect inquirers) and “randomizing” agents (reporting the outcome of a coin flip disconnected from the issue at stake) the Olsson (2011) model incorporates a continuum of agent reliability types including the types in Bovens and Hartmann (2003).

Despite this and further differences between both models, Merdes et al. (2021) report surprising similarities, they discovered in their simulations. Of particular interest might be how well the models learn that a source is lying (reporting falsehoods). Although, the Olsson (2011) model has a build-in capacity to learn that a source is lying, it fails in certain scenarios to learn that the source is indeed lying and produces credences which are only marginally less inaccurate than the credences produced by the Bovens and Hartmann (2003) model.

Merdes et al. conclude that “neither model is able to fundamentally solve the problem of source reliability in circumstances where there is no reference class of relevant past predictive success that enables outcome-based estimates of the relevant

likelihoods.” This pessimistic conclusion points to an exciting area for future work connection (formal) epistemology, statistics and/or computer simulations.

In the wake of the replication crisis, Romero and Sprenger (2021) aim to increase the reliability of published research. They investigate the suggested statistical reform to replace frequentist inference (null hypothesis significance testing) by Bayesian statistics. They compare how accurate these two statistical methodologies fare under a variety of reporting scenarios and biases operating in the sciences.

Romero and Sprenger extend the computer simulations of Romero (2016) and find that there is not much difference between both inference methods, if the inconclusive evidence is published. However, under the more realistic assumption of a present file drawer effect (inconclusive evidence is not published) they find that Bayesian inference is considerably more accurate than frequentist inference. They explain this finding pointing out that under a Bayesian regime strong evidence supporting the null hypothesis is regularly published, which is not the case under the current frequentist paradigm. As a result, the Bayesian approach increases accuracy, if the null hypothesis is true.

Romero and Sprenger conclude that the choice of the statistical framework plays an important role in increasing the reliability of published research and favour the Bayesian approach over the frequentist paradigm.

Dutilh et al. (2021) consider the implications of preregistration of statistical analyses in the face of the replication crisis in the sciences. They argue that while preregistration is a powerful and increasingly popular method to raise the reliability of empirical results, it imposes an unwelcome lack of flexibility of statistical analysis. They point to two recent high-profile studies in which unexpected circumstances forced researchers to change the preregistered statistical analysis. As a result, the researchers could only label the analysis as exploratory but not as a confirmatory.

To give statisticians the required freedom to select appropriate statistical analyses while safeguarding against p-hacking, Dutilh et al. discuss advantages and disadvantages of six methods of analysis blinding by inserting a data manager, who blinds the data for the analyst. They then give their recommendation of an analysis blinding method for three different common experimental designs. Since data blinding can only do the intended job, if the data manager and data analyst follow the rule, the authors design a simple online blinding protocol.

While recognising limitations of analysis blinding, Dutilh et al. believe, on the basis of available evidence, that analysis blinding contributes to improving the reliability of science. They also draw on their personal experiences reporting their thrill at the stage of unblinding.

In the very unlikely event this primer has not yet enticed you to read their paper, you nevertheless must check out their paper to read their fantastic cartoon (three pages long).

2.3 Medicine

Because of the high stakes at play and the increasing awareness of vested interests strongly impacting the reliability of medical evidence, stricter and stricter evidential

standards have been developed in view of helping doctors, health professionals and policy makers orient themselves in such a confusing epistemic environment Osimani (2020). This is especially valid for pharmaceutical products, whose marketing incentivizes manipulation of evidence and its disclosure at any stage of the product life cycle. This generates “information wars” and strong uncertainty. How should the epistemic community behave in the face of such manipulation strategies?

Holman (2021) addresses this topic by extending the literature on epistemic injustice from dyadic situations—where relevant agents of the knowledge exchange are essentially divided in two camps—to epistemic communities, which comprehends two or more, possibly conflicting, sources of evidence, and an audience of agents, whose behaviour may be more or less strongly impacted by the received information (e.g. doctors prescribing a given drug). In his analysis he draws on the notion of testimonial injustice, which Fricker (1998, 2007) identified as a specific kind of epistemic injustice, relating to situations where an agent endowed with rational authority (i.e. both competent and trustworthy) is denied credibility, or suffers a credibility deficit, and is therefore inhibited in their capacity to deliver knowledge, on the basis of identity prejudices.

Expanding the notion of epistemic injustice from dyadic cases to group settings allows one to see that the counterpart of credibility deficit, that is excess credibility may be a possible generator of epistemic injustice too. This sort of injustice emerges out of the fact that receivers of the message, qua hearers have unlimited credibility to assign to the source (hence, as such, credibility is non-finite good), but, qua believers, they must weigh the *relative credibility* of conflicting sources (and possibly discard some of them) in order to update their beliefs in light of contradictory claims. Hence, apportioning undue credibility to a source of knowledge automatically entails denying trustworthy competitors to be assigned their due portion of credibility. The sort of epistemic injustice that excess of credibility generates is denoted by Holman as “collateral epistemic injustice”. This automatically springs from the audience wrongly apportioning an excess of credibility to a source of information, thereby automatically undercutting the capacity of other comparatively more reliable sources of evidence of being duly recognised as such.

Holman hence distinguishes the notion of credibility from that of influence; that is the degree to which each one of conflicting sources of information may impact the receivers beliefs: while the former is not finite, the latter is. Hence hearers have the ethical duty to apportion it judiciously. Holman exemplifies his points by illustrating the behaviour of the community of doctors faced by conflicting information from scientific publication, expert opinion, personal experience and promotional material in evaluating the efficacy and safety of the hormonal treatment DES during the 50s-60s. By apportioning undue credibility to the information received by pharmaceutical detailers, doctors automatically dismissed any other conflicting evidence from much more authoritative sources and discarded them.

A further consequence of these considerations is that the value-apt ideal John (2018a, b, 2019) is insufficient. Scientists must not only disclose their non-epistemic values, whenever these are not shared by the listeners, in order not to infringe their autonomy. They must also communicate the extent to which, and how, such values impacted on their inferences and conclusions.

Related to the value-ladenness of data is also the other contribution in the special issue devoted to medical evidence Lalumera et al. (2020), which is rooted in the tradition of Science and Technology Studies. The paper investigates the reliability of evidence from molecular imaging diagnostics by analysing its use in the medical practice. This perspective is particularly cogent in that the use of advanced medical imaging such as PET (positron emission tomography), fMRI (functional magnetic resonance imaging), or CT (compute tomography) is increasingly prescribed by doctors and solicited by patients in daily health care. Although the information provided by such diagnostic tests is material to many important therapeutic, surgical and prophylactic decisions, the “illusion of immediacy” that such imaging techniques bring with them may induce a false sense of knowledge and, relatedly, a tendency to overutilization.

Lalumera et al. distinguish three notions of reliability: one concerns the nature of knowledge and the truth-conduciveness of learning processes (epistemological reliability); another one relates to the accuracy and precision of the instruments put into place in order to learn from data, with respect to the investigated phenomenon; the commonsense notion of such dimension of reliability is the proportion of true vs. false positives and negatives, as measured for instance by test sensitivity or specificity (methodological reliability); a third notion of reliability is related to the repeatability and reproducibility of study results (procedural reliability). They then go on to assess the status of molecular imaging techniques with respect to these dimensions of reliability. With respect to the first one, PET imaging is heavily theory mediated in that the pixels that form the output image, are the results of a numerical conversion of photon collision events, associated with hyperglycolysis, itself explained as an effect of various malignant tumors, according to Warburg’s theory of cancer. The translation of hyperglycolysis phenomena into pixels is mediated by normalization and correction algorithms; finally the resulting image needs the interpretation of an expert specialised in reading such diagnostics, who needs to assign a clinical interpretation to the specific instantiations of such imaging techniques.

One could say that, in the first step, one goes from theory to expected evidence, in the second one, focus is on the link between the two, in the third one, one goes back from evidence to theory, as it were, and abduction enters into play in order to subsume the received evidence into the most plausible diagnostic slot. This step is impacted not only by observer-dependence, but also by the reference-class dependency of test accuracy. The authors conclude by suggesting three avenues for the improvement of imaging techniques utilization. Helping the interpretation step by recurring to AI as an aid to diagnosis; in turn, fostering the harmonization of AI standards through consensus conferences and Delphi studies; enhancing patient communication, with a special focus on the limitations, the specific epistemic characteristics of such tests, and their cost-efficacy profile, with respect to issues of fairness and life values.

2.4 Quantum mechanics, computer science and reliability

As evidenced, by the diversity of contributions presented so far in this introduction, the heterogeneity of contributions appearing in this volume offers a rich panorama on the notion of reliability, which extends beyond its traditional boundaries. Such

extension is particularly evident in de Ronde et al. (2021) and Primiero et al. (2021). The mentioned papers present indeed the role of reliability in (the foundation) of Quantum Mechanics and in the theory of change of software systems, respectively.

de Ronde et al. (2021) approaches reliability in the context of one of the most puzzling philosophical issues raised by quantum mechanics, which is central in the discussion around the interpretation of the theory: the role and the interpretation of probability. To sum up the whole problem in a question: which notion of probability comes out from quantum theory? Interestingly enough, the answer depends on the interpretation of quantum theory that one is more likely to embrace, and this leads to different views of reliability, relevant for our debate. The key role played by probability in quantum mechanics relies on the famous Born rule: fixing the quantity $\langle \psi | H | \psi \rangle$ allows one to predict the average value of an observable H with respect to a given state of the system, represented by a vector ψ in a Hilbert space (while H is an Hermitian operator over the same space). It turns out that for the supporters of the subjective approach to probability, that are found within some of the most popular interpretations of quantum mechanics, namely the quantum Bayesianism and the many worlds interpretation, quantum probability represents a “*reliable predictive tool* used by agents in order to compute measurement outcomes” (de Ronde et al. (2021), page 2. Italic is ours). On a different interpretative strand, the objective reading of quantum probability, substantially endorsed in de Ronde et al. (2021), “understands it as providing a *reliable informational source* of a real state of affairs—as theoretically described by QM.” (de Ronde et al. (2021), page 2. Italic is ours). The analysis of the different interpretations of quantum probability, and, consequently of reliability in QM, is finally contextualized within some relevant applications—both for the theory itself and for its applications—of the theory, namely quantum computation and quantum computational logics [see Freytes and Sergioli (2014), Dalla Chiara et al. (2014)]. Although the vision of probability, understood as a reliable tool for an agent seems to be grounded in these applications, the new look given to the objective approach discussed in de Ronde et al. (2021) appears very convincing in the context of quantum computation.

Slightly changing the field, from quantum to classical computation, an operative question arises naturally in (theoretical) computer science: when can a software be considered reliable? Primiero et al. (2021) moves from the consideration that the answer is necessarily grounded on two concepts: *resilience* and *evolvability*. In order to provide a sketch, on the one hand, resilience of a computational system mirrors its capability to “preserve a working implementation under changed specifications”. On the other hand, evolvability represents the “ability to successfully accommodate changes, the capacity to generate adaptive variability in tandem with continued persistence and more generally the system’s ability to survive changes in its environment, requirements and implementation technologies” [see Ciraci and van den Broek (2006)]. Primiero et al. (2021) carries out an innovative logical analysis. In particular, the authors introduce a logical system where formulas trigger change operations and an order relation regulates priorities between functionalities. It turns out that resilience in a software can be defined via logical consequence. Moreover, this kind of “inferential” approach (to the notion of resilience) also brings a characterisation of fault-tolerance. In this context,

evolvability is defined “as the property of a software to be updated to fulfill a newly prioritised set of functionalities”.

3 Wrapping up

We hope that this Special Issue will trigger further discussion on the intriguing notion of reliability. We considerably increased our understanding of such multifaceted and fundamental epistemological notion, and will treasure all contributions in our future work. All papers are worth engaging with and we wish them all great impact. We do, however, want to caution against only using impact factors to reliably measure (their) impact Greenwood (2007).

Acknowledgements Open Access funding provided by Projekt DEAL. We want to thank Wiebe van der Hoek and the team at Synthese for their continued and very timely support. Furthermore, we would like to thank the anonymous reviewers for their diligent work. Finally, many thanks to the authors of all the contributed manuscripts, whether their manuscripts were accepted for publication or not.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Avigad, J. (2021). Reliability of mathematical inference. *Synthese*, <https://doi.org/10.1007/s11229-019-02524-y>, forthcoming.
- Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Oxford University Press.
- Ciraci, S., van den Broek, P. (2006). Evolvability as a quality attribute of software architectures. In L. Duchien, M. D’Hondt, T. Mens (Eds.), *Proceedings of the International ERCIM Workshop on Software Evolution 2006*, pp. 29–31.
- Claveau, F., & Grenier, O. (2019). The variety-of-evidence thesis: a Bayesian exploration of its surprising failures. *Synthese*, 196, 3001–3028. <https://doi.org/10.1007/s11229-017-1607-5>.
- Comesaña, J. (2006). A well-founded solution to the generality problem. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 129(1), 27–47. <https://doi.org/10.1007/s11098-005-3020-z>.
- Dalla Chiara, M., Giuntini, R., & Sergioli, G. (2014). Probability in quantum computation and in quantum computational logics. *Mathematical Structures in Computer Science*, 14(3), e240306. <https://doi.org/10.1017/S0960129512000734>.
- de Ronde, C., Freytes, H., & Sergioli, G. (2021). Quantum probability: a reliable tool for an agent or a reliable source of reality? *Synthese*, <https://doi.org/10.1007/s11229-019-02177-x> forthcoming.
- De Pretis, F., Landes, J., & Osimani, B. (2019). E-Synthesis: a Bayesian Framework for Causal Assessment in Pharmacosurveillance. *Frontiers in Pharmacology* 10, URL <https://doi.org/10.3389/fphar.2019.01317>, 17 pages supplementary material available at <https://www.frontiersin.org/articles/10.3389/fphar.2019.01317/full#supplementary-material>
- Dunn, J. (2021). Reliable group belief. *Synthese*, <https://doi.org/10.1007/s11229-018-02075-8>, forthcoming.

- Dutilh, G., Sarafoglou, A., & Wagenmakers, E. J. (2021). Flexible yet fair: blinding analyses in experimental psychology. *Synthese*. <https://doi.org/10.1007/s11229-019-02456-7>, forthcoming.
- Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 89(1), 1–29. <https://doi.org/10.1023/A:1004243308503>.
- Freytes, H., & Sergioli, G. (2014). Fuzzy approach for toffoli gate in quantum computation with mixed states. *Reports on Mathematical Physics*, 74(2), 159–180. [https://doi.org/10.1016/S0034-4877\(15\)60014-3](https://doi.org/10.1016/S0034-4877(15)60014-3).
- Fricker, M. (1998). Rational authority and social power: Towards a truly social epistemology. *Proceedings of the Aristotelian Society*, 98, 159–177. <https://doi.org/10.1111/1467-9264.00030>.
- Fricker, M. (2007). *Epistemic injustice: power and the ethics of knowing*. Oxford: Oxford University Press.
- Goldman, A. (2014). Social process reliabilism: solving justification problems in collective epistemology. In J. Lackey (Ed.), *Essays in collective epistemology*. Oxford: Oxford University Press.
- Greenwood, D. C. (2007). Reliability of journal impact factor rankings. *BMC Medical Research Methodology*, 7(1), 48. <https://doi.org/10.1186/1471-2288-7-48>.
- Holman, B. (2021). An ethical obligation to ignore the unreliable. *Synthese*. <https://doi.org/10.1007/s11229-019-02483-4>, forthcoming.
- John, S. (2018a). Epistemic trust and the ethics of science communication: Against transparency, openness, sincerity and honesty. *Social Epistemology*, 32(2), 75–87. <https://doi.org/10.1080/02691728.2017.1410864>.
- John, S. (2018b). Scientific deceit. *Synthese*. <https://doi.org/10.1007/s11229-018-02017-4>.
- John, S. (2019). Science, truth and dictatorship: wishful thinking or wishful speaking? *Studies in History and Philosophy of Science Part A*, 78, 64–72. <https://doi.org/10.1016/j.shpsa.2018.12.003>.
- Kelly, K. T., Genin, K., & Lin, H. (2016). Realism, rhetoric, and reliability. *Synthese*, 193(4), 1191–1223. <https://doi.org/10.1007/s11229-015-0993-9>.
- Kummerfeld, E., & Danks, D. (2014). Model change and reliability in scientific inference. *Synthese*, 191(12), 2673–2693. <https://doi.org/10.1007/s11229-014-0408-3>.
- Lalumera, E., Fanti, S., & Boniolo, G. (2020). Reliability of molecular imaging diagnostics. *Synthese*. <https://doi.org/10.1007/s11229-019-02419-y>, forthcoming.
- Landes, J., & Osimani, B. (2020). On the assessed strength of agents' bias. *Journal for General Philosophy of Science*. <https://doi.org/10.1007/s10838-020-09508-4>.
- Landes, J. (2021). The variety of evidence thesis and its independence of degrees of independence. *Synthese*. <https://doi.org/10.1007/s11229-020-02738-5>, forthcoming.
- List, C., Puppe, C. (2009). Judgment aggregation: a survey. In P. Anand, P. Pattanaik, C. Puppe (Eds.), *The handbook of rational and social choice*.
- Mayo-Wilson, C. (2014). Reliability of testimonial norms in scientific communities. *Synthese*, 191(1), 55–78. <https://doi.org/10.1007/s11229-013-0320-2>.
- Merdes, C., von Sydow, M., & Hahn, U. (2021). Formal models of source reliability. *Synthese*. <https://doi.org/10.1007/s11229-020-02595-2>, forthcoming.
- Olsson, E. J. (2011). A simulation approach to veritistic social epistemology. *Episteme*, 8(2), 127–143. <https://doi.org/10.3366/epi.2011.0012>.
- Olsson, E. J., & Schubert, S. (2007). Reliability conducive measures of coherence. *Synthese*, 157(3), 297–308. <https://doi.org/10.1007/s11229-006-9056-6>.
- Osimani, B. (2020). Epistemic gains and epistemic games: reliability and higher order evidence in medicine and pharmacology. In Boston studies in the philosophy and history of science, vol. 338. *Springer, Cham, chap. 15*, 345–372. https://doi.org/10.1007/978-3-030-29179-2_15.
- Osimani, B., & Landes, J. (2020). Varieties of error and varieties of evidence. *British Journal for the Philosophy of Science*, forthcoming.
- Pittard, J. (2017). Disagreement, reliability, and resilience. *Synthese*, 194(11), 4389–4409. <https://doi.org/10.1007/s11229-016-1137-6>.
- Primiero, G., Raimondi, F., & Chen, T. (2021). A theory of change for prioritised resilient and evolvable software systems. *Synthese*. <https://doi.org/10.1007/s11229-019-02305-7>, forthcoming.
- Romero, F. (2016). Can the behavioral sciences self-correct? A social epistemic study. *Studies in History and Philosophy of Science Part A*, 60, 55–69. <https://doi.org/10.1016/j.shpsa.2016.10.002>.
- Romero, F., & Sprenger, J. (2021). Scientific self-correction: the Bayesian way. *Synthese*.
- Schindler, S. (2011). Bogen and woodward's data-phenomena distinction, forms of theory-ladenness, and the reliability of data. *Synthese*, 182(1), 39–55. <https://doi.org/10.1007/s11229-009-9615-8>.
- Schippers, M. (2014). Coherence, striking agreement, and reliability. *Synthese*, 191(15), 3661–3684. <https://doi.org/10.1007/s11229-014-0488-0>.

- Schubert, S. (2012). Is coherence conducive to reliability? *Synthese*, 187(2), 607–621. <https://doi.org/10.1007/s11229-010-9865-5>.
- Tolly, J. (2021). Knowledge, evidence, and multiple process types. *Synthese*, <https://doi.org/10.1007/s11229-019-02146-4>, forthcoming.
- Trpin, B., Dobrosovstnova, A., & Otzendorfer, S. J. (2021). Lying, more or less: A computer simulation study of graded lies and trust dynamic. *Synthese*, forthcoming.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.