The Reasoner

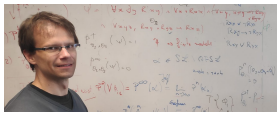
Volume 18, Number 4
July 2024

thereasoner.org
ISSN 1757-0522

Contents

Guest Editorial

We live in a world which never stays still. We've all (had to) become used to change. It could be said that the only constant in our lives is a continued change, which is progressing at an ever faster pace. One of the few remaining fixtures in my life used to be *The Reasoner*. Now changes are coming to *The Reasoner*, too. I'm taking this opportunity to talk to the founder, who is also subject to changes.



Jürgen Landes
Munich Center for Mathematical Philosophy

Features

Interview with Jon Williamson

Jon Williamson is currently Professor of Reasoning, Inference and Scientific Method at the University of Kent.

Jürgen Landes: Can you tell us a bit about your research interests?

Jon Williamson: Sure. I work on topics in epistemology, metaphysics, logic, philosophy of science, philosophy of medicine and philosophy of AI—particularly topics connected in one way or another to causality or probability.

For example, I've been working for a long time on three specific theories:

Evidential Pluralism is a theory of the epistemology of causality. It holds that causation is established by establishing both the presence of an appropriate correlation and the existence of a suitable mechanism between the putative cause and effect. So, to evaluate causation, we need to scrutinise both association studies, which test for the presence of the appropriate correlation, and mechanistic studies, which test for the presence of key features of hypothesised mechanisms. Orthodox evidence-based evaluation methods tend to focus just on association studies. Evidential Pluralism urges considering mechanistic studies too, leading to a new approach to evidence-based medicine, called EBM+, and a new approach to evidence-based policy, called EBP+.

Objective Bayesianism is a theory that says something about exactly how strongly one should believe propositions of interest, given available evidence. According to objective Bayesianism as I conceive of it, strengths of belief should be probabilities, calibrated to empirical probabilities insofar as one has evidence of them, and equivocal unless evidence warrants strong

belief or disbelief.

Epistemic Causality is a theory about the nature of causality: a theory that understands causal relationships in terms of rational causal beliefs. It is analogous to the Bayesian account of probability. The Bayesian is interested in probabilistic beliefs as a kind of belief—degrees of belief—rather than beliefs about some putative non-epistemic probabilities. Similarly, according to epistemic causality, causal beliefs are a kind of belief, not beliefs about causal relationships that are 'out there' in the world. Our causal beliefs are tools to help us predict, explain and control the world. This leads to a view of causal relationships as epistemic, but objective.

Jürgen Landes: What are your current projects?

Jon Williamson: I have a Leverhulme Trust project, working with with Alexandra Trofimov on using Evidential Pluralism to motivate a new approach to evidence-based law, called EBL+. It's important to assess whether our laws are doing what we want them to do—for example, to assess whether banning mobile phone use when driving reduces the number of road accidents. But how do we assess the effectiveness of laws? Orthodox evidence-based evaluation would tell us to carry out a randomised controlled trial of the intervention. But that's almost never possible in the case of laws: one can't randomise individuals to a group that is bound by a law and a control group that instead receives a 'placebo' law indistinguishable from the law itself. We're going to have to appeal a much more diverse range of evidence, and Evidential Pluralism can tell us how.

I'm also very excited about a project I have with you and Soroush Rafiee Rad on inductive logic. Following the demise of Carnap's programme for inductive logic, inductive logic has become a bit of a niche area, but this is undeserved, we think. Our plan is to ground inductive logic in objective Bayesianism: i.e., to use the norms of objective Bayesianism to tell us how strongly we should believe a conclusion, given some premisses which they themselves be less than certain. I think it's fair to say that we've been taken aback by how well things work out. For example, we were surprised to find that a very large class of inferences in predicate inductive logic is decidable. We're just writing everything up now and hopefully will have a book to show for it before too long.

I'm also in the early stages of planning an interdisciplinary project that will further develop EBP+. I hope to have the opportunity to go ahead with this project and report back in a future issue of *The Reasoner*!

Jürgen Landes: Tell us how *The Reasoner* was born.

Jon Williamson: Two concerns led to *The Reasoner*. First, the community of researchers interested in reasoning, inference and scientific method was fragmented across many disciplines, with no natural forum for interdisciplinary communication. Second, there was no real medium of communication that lay between the medium of journal publication, which requires a considerable time commitment from authors, reviewers and readers, and that of blog posting, which can be quite sporadic and short-lived. Hence the idea of a regular 'gazette,' containing short articles, for the benefit of the interdisciplinary reasoning community. This came to fruition in May 2007.

Jürgen Landes: Outline connections to the Centre for Reasoning.

Jon Williamson: Around the same time, my colleague David Corfield and I developed plans for an interdisciplinary Centre to provide a hub for work on reasoning and methodology at the University of Kent. The Centre for Reasoning quickly came to fruition and had about 50 members, from all corners of the university. We hosted *The Reasoner*, as well as many research projects and conferences, and we ran weekly seminars.

Jürgen Landes: You edited *The Reasoner* for its first decade. Can you tell us about the transfer to the current editor?

Jon Williamson: Thanks to the efforts and enthusiasm of our editorial board, and in particular our news, features and production editors, the editorial process went very smoothly. So much so that I didn't change much when I was editor. Ten years in, I felt it was time for some fresh ideas and Hykel Hosni took over as editor and made some great changes. The main effort has always been encouraging people to contribute. I would strongly encourage readers to submit. This is a unique forum for the community and it requires regular community involvement.

Jürgen Landes: What is currently happening at the Centre for Reasoning and your department?

Jon Williamson: Unfortunately, the University of Kent is cutting a range of subjects across the humanities and social sciences, including Philosophy, which runs the Centre for Reasoning. This means the Centre is also being cut, and the philosophy staff are being made redundant. This is a great disappointment and rather unexpected because our department was doing so well. Recently we were 5th in the UK research assessment, 3rd in our national student satisfaction survey, and we are one of only three subjects at the university ranked in the top 100 globally by QS. We are also financially viable, even by the exacting financial demands of the university.

So what explains the cuts? Many readers will have already heard that UK higher education is in a desperate situation. All UK universities are feeling poor, due to student fees remaining static throughout a period of high inflation. Added to this, there are enormous problems caused by structural changes. There used to be quotas on student numbers in the UK: this meant that there were limits on the numbers of students that universities higher up the league tables could admit, leaving plenty of students for universities lower down, including provincial universities like mine. This resulted in a very wide geographical spread of top-quality education and research, including in deprived areas of the UK. But during 2012–2016, the UK government scrapped these quotas. This allowed the large city universities to expand without limit and has now left universities like mine with deficits caused by decreasing student numbers. The new system provides poor value for money for almost all students. The large city universities are overcrowded and offer little individual attention to students, while smaller and provincial universities are facing widespread staff cuts, leading to poor regional provision and, again, less individual attention for students.

So, universities like mine have to make widespread cuts. Exactly which areas get cut is then a matter of internal politics. Smaller departments with less representation at the meetings that matter are often easiest to cut. Reason doesn't seem to enter the picture, ironically.

Jürgen Landes: Is there something you want to say to our readers?

Jon Williamson: The closure of my department and the Centre for Reasoning means that Kent can no longer host *The Reasoner*. I'm pleased to say that Milan University Press has agreed to take it on. This will lead to certain advantages, including a unique doi number for each published article. Please submit!

Evidential Pluralism as a methodology for Evidence-Based Law

I am currently working on a Leverhulme funded project with Professor Jon Williamson. The aim of our project is to develop a new approach to evidence-based law using the principles of Evidential Pluralism, called EBL+. Evidence-based law (EBL) is an emerging approach to law that seeks to make use of the best available evidence to ensure that legislations and regulations effectively achieve their aims (EU Commission, 2023; UK Government, 2023; Westminster Foundation for Democracy). This raises the question, 'what evidence should be considered?'

On orthodox evidence-based approaches, randomised controlled trials (RCTs) are the gold standard of evidence. Appreciation of the limitations of orthodox evidence-based approaches have led to calls for a more inclusive approach to evidence in other domains, including medicine and policy.

When we turn to law, the limitations of RCTs are even greater. Firstly, there might be ethical challenges to insisting that individuals in an intervention group must abide by a law that other individuals in the same jurisdiction do not have to abide by. Secondly, it is not possible to properly blind and randomise a law to individuals. This is because participants need to know that they are subject to a law in order to comply with it and enforcers need to know who is subject to a law in order to enforce compliance. Thirdly, there might be spillover effects, such that those in the control group abide by the law because those in the intervention group are abiding by it.

Given the limitations of orthodox evidence-based approaches, it is necessary to adopt a more inclusive approach to evidence when evaluating laws. Evidential Pluralism offers such an approach.

Evidential Pluralism is a philosophical account of causal enquiry. According to Evidential Pluralism, to establish that A is a cause of B requires establishing:

> (i) That A and B are appropriately correlated, and
>
> (ii) That there is some mechanism connecting A and B and which can account for the extent of the identified association.

Evidential Pluralism has previously been applied to develop a more inclusive evidence-based approach in medicine, called EBM+, and policy, called EBP+. A similar application to evidence-based law provides a needed methodology for systematically integrating different kinds of evidence to evaluate the effects of laws, called EBL+.

Covid-19 face mask mandates provide a good proof of concept case study to illustrate the need for and benefits of an EBL+ evaluation. During the Covid-19 pandemic, uncertainty and controversy concerning the effectiveness of public health interventions, including public face mask mandates, resulted from a narrow focus on experimental studies. This prompted calls for a more inclusive approach to evidence in responding to the novel, complex and rapidly changing problem of Covid-19 (Aronson et al. 2020; Greenhalgh et al., 2022).

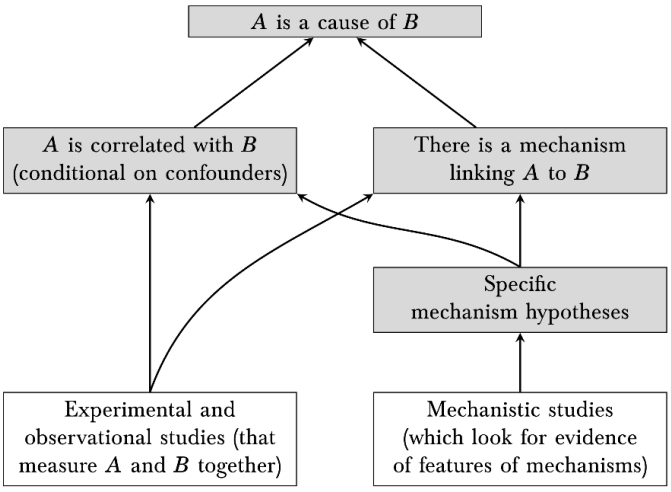


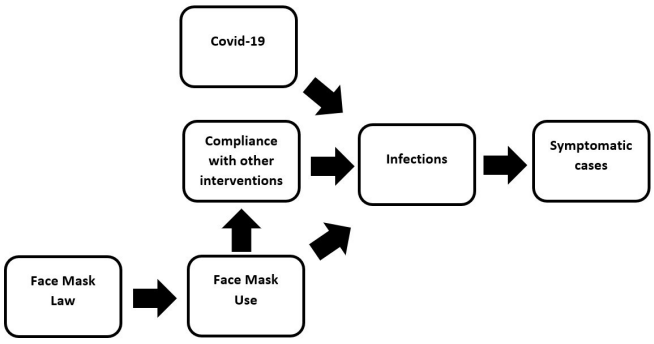
Figure 1: Evidential Pluralism

An EBL+ evaluation begins by specifying the claims of interest. In this case, the causal claim of interest is that a legal requirement to wear a face mask in public reduces the prevalence of symptomatic Covid-19 infections and thereby reduces the number of hospitalisations and deaths.

The correlation claim is that a legal requirement to wear a face mask in public is negatively correlated with symptomatic infections, conditional on potential confounders.

A plausible mechanism hypothesis is that a legal requirement to wear a face mask in public increases the use of face masks which in turn reduces the prevalence of covid-19 which reduces the prevalence of symptomatic infections and thereby the number of hospitalisations and deaths.

A plausible hypothesised counteracting mechanism is that a legal requirement to wear a face mask in public will decrease compliance with other public health interventions, such as social distancing. This, in turn, would result in an increase in the number of symptomatic infections compared to the number that would have occurred if the legal requirement to wear a face mask had not been introduced.



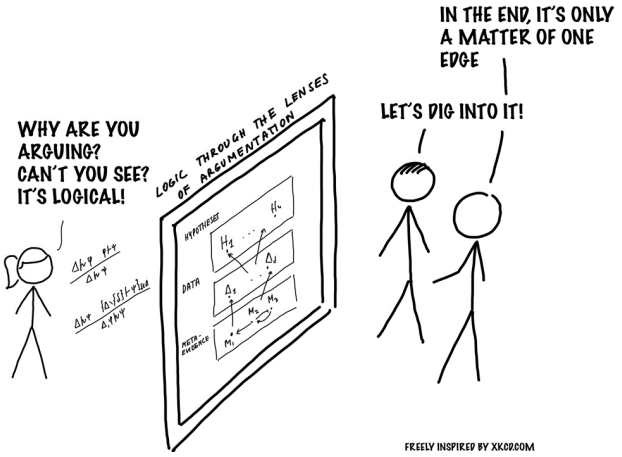
Taking account of available evidence, we found that experimental and observational studies detect a robust correlation across contexts. We also found that each stage of the mechanism hypothesis is supported by a range of studies and that there is evidence against the hypothesised counteracting mechanism. Overall, we conclude that the combination of evidence of correlation and evidence of mechanisms establishes the effectiveness of face mask mandates (Trofimov and Williamson, forthcoming).

As illustrated through the proof of concept case study of Covid-19 face mask mandates, Evidential Pluralism provides a much-needed methodology for systematically incorporating a range of evidence to evaluate laws.

Alexandra Trofimov
Philosophy, University of Kent

How Argumentation Theory and Antisequents Can Shed Light on the Scientific Debate

The objective of logic is to formalize correct reasoning. However, the valid rules of inference are contingent upon the circumstances. In the context of scientific reasoning, the scientific community has repeatedly engaged in significant debates (the Ptolemaic vs. Copernican system, the expansion of the universe or, more recently, the efficacy of certain vaccines). Such debates are not just about the collection and interpretation of data but also about the logical framework through which scientists understand and communicate their findings. Keeping implicit the methodological aspect of the scientific research can lead into several problems in terms of transparency and evaluation of results. Introducing new logical systems that formalize scientific methodologies appears to be a reasonable approach to tackle these issues. However, it's important to note that the scientific community may not be accustomed to the specific language and formalisms of logic, and an intermediate level of abstraction could bridge this gap. One potential solution is to adopt Dung's style argumentation theory. As Dung (1995: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence, 77(2):321-357) proposed in his seminal paper, argumentation frameworks can be visualized as a directed graph where the nodes represent the arguments and the edges a relation between the arguments, intuitively understood as 'attack'. These frameworks can be enriched in several ways: adding weights to the attack relations, introducing a support relation, instantiating the arguments, etc. Furthermore, with different definitions of semantics, different sets of arguments can be justified in various ways. In essence, argumentation frameworks have a significant expressive power and their structure is easy to understand. Consequently, by instantiating arguments using logical formulas (and this can be done in several ways) and possibly also the relations among the arguments, we can make explicit the actual practice scientists use and the implicit logic they use. It is important to note that the formalization of the scientific methodology in terms of logical terms will not prescribe scientists' actions; rather, it will enhance the comprehension of where and why scientists agree or disagree. If the motivation for building this bridge is clear, many are the ways to do it. We could work on a fully abstract level by simply distinguishing three types of arguments: hypotheses, data, and meta-evidence. However, if we want to see which logic is at work we should instantiate the arguments and the relations among them using logical formulas. In the literature of logical argumentation theory, it has been explored how to use sequents to instantiate arguments, see e.g. Arieli and Straßer (2019: Logical argumentation by dynamic proof systems. Theoretical Computer Science, 781:63-91). In a recent paper by Piazza, Pulcini and Sabatini (2023: Abduction as Deductive Saturation: a Proof-Theoretic Inquiry, Journal of Philosophical Logic, 52(6):1575-1602) it is explored the concept of abduction and its relationship with deductive saturation from a proof-theoretic perspective. Abduction, as a form of reasoning, involves generating hypotheses that are able to explain empirical evidence or phenomena, it is often described as "inference to the best explanation" and, as the authors say in Piazza et al. (2023, 1576): "the ultimate goal of a rational agent in abductive reasoning can be described as the search for the missing premise of an "unsaturated" deductive inference". Furthermore, they introduce a hybrid system where the rules are defined in terms of both sequents and antisequents that, in the context of refutation calculi, are introduced to denote sequents that assert their own invalidity. Given the central role of the attack relation in argumentation theory and the rejection of hypotheses in the scientific practice, the use of a system defined in terms of antisequents seems a new and potentially fruitful connection. Following the approach of Arieli and Straßer, a new dynamic proof system could be defined and possibly simplify the process of arguments evaluation. Then, starting from a real case – perhaps from the field of life sciences – we could investigate which abductive algorithm is justified by the actual scientific practice. In addition, always having the real scientific practice as justification method, we could investigate if it is possible to identify some patterns, that in the literature are referred to as attack principles, among the arguments instantiates either with sequents or antisequents once they share in their supports (the antecedent) or in their claims (the consequent) some propositional formula.



By employing this comprehensive approach we should be able to let the lab scientists and the logicians talk. Providing a framework to make the logical structure of scientific reasoning explicit, scientists can then better communicate their methodologies and findings both to the broader scientific community and society

Esther Anna Corsi
LUCI Lab, University of Milan

The Reasoner Speculates

Role of heuristics in diagnostic reasoning in practice

In last 50 to 60 years, experimental studies in cognitive psychology have revealed, that heuristics, which are mental shortcuts in reasoning, cause errors in making judgments under uncertainty in everyday reasoning (Gilovich T et al. 2002. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge: Cambridge University Press). These findings have been applied to diagnosis in recent years to claim heuristics cause similar errors in diagnostic reasoning, which lead to diagnostic errors (Croskerry P. 2003. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med 78: 775-80). I shall argue, this application is incorrect as diagnostic reasoning is very different from everyday reasoning and therefore heuristics do not cause diagnostic errors. I shall base my argument on different roles of the heuristic of representativeness, which is the most extensively studied heuristic in cognitive psychology, in everyday and in diagnostic reasoning.

I shall start by looking at the well-known engineer-lawyer experiment (Tversky A. et al.1974. Judgment under uncertainty: Heuristics and Biases. Science 185: 1124-31), which is one of several experiments about this heuristic, to learn about the experimental finding about this heuristic which has been applied to diagnosis. In this experiment, subjects are provided with a personality description of an individual which is typical of an engineer, and asked to judge probability of this individual being an engineer in two conditions. In one condition, they are told, the individual is drawn from a group of 70 engineers and 30 lawyers, while in another condition he is drawn from a group of 30 engineers and 70 lawyers. The subjects are found to judge probability of this individual being an engineer to be the same in both conditions, which is an error, as the base rate or prior probability of this individual being an engineer is neglected in making the probability judgment. This error is attributed to heuristic of representativeness by which this probability judgment is made solely from resemblance of personality description to stereotype of an engineer, which leads to cognitive bias of base rate neglect.

This experimental finding, that the heuristic of representativeness causes an error in making a probability judgment in everyday reasoning, appears to have been applied to diagnostic reasoning to claim it causes a similar error in judging probability of a disease, which leads to a diagnostic error. I shall now look at diagnostic reasoning in a clinical situation analogous to the engineer-lawyer experiment to see if this application is correct. Let us consider a patient, who presents with chest pain typically seen in acute myocardial infarction (MI). This application implies, probability of acute MI in this patient during diagnosis will be judged due to heuristic of representativeness solely from resemblance of this presentation to stereotype of acute MI, while ignoring its base rate (whether patient is an old man or young woman), which would be an error. But if we look at diagnostic reasoning in this patient in practice, this is not how probability of acute MI is actually judged in this patient. In practice, acute MI is merely suspected in this patient, I suggest, from resemblance of this presentation to stereotype of acute MI and formulated as a hypothesis, which is evaluated by a test. The prior probability of acute MI is estimated from its prevalence and combined with likelihood ratio (LR) of test result to generate (posterior) probability of acute MI, from which it is diagnosed (Kassirer JP et al. 2009. Learning Clinical Reasoning. Baltimore: Lippincott Williams and Wilkins).

We note (posterior) probability of acute MI is judged in this patient in practice only after testing, in a process of hypothesis generation and testing. The role of the heuristic of representativeness in this process is only to suspect acute MI and formulate it as a hypothesis and not to judge its probability. In this role, it does not cause a diagnostic error, for if this hypothesis is found to be incorrect after testing, it is discarded, and another hypothesis generated and tested.

In addition, there is no credible published report about a real patient in whom this heuristic has caused a diagnostic error. The example given in the literature of this heuristic causing diagnostic error is of failure to suspect a disease with an atypical presentation due to lack of resemblance of this presentation to stereotype of disease (Croskerry P. 2003. 775-80). I find this diagnostic error being attributed to heuristic of representativeness to be strange, for it is resemblance, not lack of resemblance which causes an error in making a probability judgment in the engineer-lawyer experiment. I suggest, a more likely cause of this diagnostic error seen in practice is lack of awareness of atypical presentations of a disease due to lack of experience.

It will not escape notice that diagnostic reasoning, with its hypothesis generation and testing, is essentially scientific in nature (Jain BP. 2017. The scientific nature of diagnosis. Diagnosis 4: 17-19), while everyday reasoning, which lacks hypothesis generation and testing, is unscientific. The heuristic of representativeness plays different roles, as we have discussed above, in unscientific everyday reasoning and in scientific diagnostic reasoning. In the former it causes an error in making a probability judgment, while in the latter, it generates a hypothesis only. Due to these different roles, the experimental finding of this heuristic causing an error in making a probability judgment in unscientific everyday reasoning is not applicable to scientific diagnostic reasoning and thus this heuristic does not cause a diagnostic error.

Bimal P Jain MD
Mass General Brigham/Salem Hospital

Dissemination Corner

SMARTEST
The Epistemology of DT Simulations

As introduced inVolume 18, Issue 2 of *The Reasoner*, the project SMARTEST aims at developing an ontological, epistemological, and logical analysis of Digital Twin (DT) simulations. A DT is a digital replica of an industrial artefact, a physical system under development. DT simulations are becoming crucial to many industrial processes as long as they are able to provide predictions of correct and incorrect behaviours of the artefact while it is being developed, as well as of interested properties of the system during its lifecycle. SMARTEST intends to inquire the reliability of those predictions on the basis of an analysis of the simulation processes involved. The unit at the Department of Cognitive, Psychological, Pedagogical Sciences and Cultural Studies of the University of Messina (UNIME), is involved in the epistemological examinations of DT simulations. This will be done on the basis of the ontology

of DTs and their relation with the simulated systems, to be developed by the LOA unit at CNR Trento (see Volume 18, Issue 3).

Computer simulations have been the object of epistemological investigations since, at least, the pioneering work on simulative artificial intelligence by Herbert Simon and Allen Newell in the 1950's. Today, the Epistemology of Computer Simulations (EOCS) is a well established subfield of the philosophy of science, dealing with such methodological problems as the definition of what a computer simulation is, the ontology of models involved in a simulation, their empirical adequacy, their verification and validation, and the epistemological status of computer simulations, to mention some. These issues have been developed mainly for equation-based simulations, wherein a dynamical system is taken as a *mathematical model* of a physical *target system*, and a *computational model* approximating the involved differential equations is implemented in a *computational artefact* to provide numerical solutions to the differential equations,thereby mimicking the temporal evolution of the target system.

DT simulations differ in, at least, two epistemologically significant aspects. Firstly, they are powered by deep learning (DL) models, fed with data collected from the industrial artefact and used to draw new correlations concerning potential behaviours of the latter. DL models are known for being epistemically opaque, that is, not fully interpretable; consequently, they are often represented by probabilistic structures, such as probabilistic state transition systems, to achieve formal verification. Secondly, there is a continuous data flow from the artefact to the DL model and back: DL models are constantly updated from the artefact data, and the artefact is developed following the regular predictions produced by the DL model. The epistemological scenario depicted by EOCS therefore needs to be carefully modified for DT simulations. This is the *second* objective of SMARTEST, which will be mainly accomplished by the UNIME unit. UNIME contributes with competence in the fields of philosophy of computing, epistemology of computer simulation, and formal verification. Members of UNIME have strong expertise on DL, covering both formal, engineering, and philosophical aspects.

DT is not the only simulation context wherein DL is employed. Current trends of DL applications involve simulative contexts wherein there is the implicit emergence, inside a model exposed to data of a natural system, of structures that bear some correspondences with structural features of that system. Examples include convolutional DL models used to simulate parton shower, neural models simulating the Hénon-Heiles potential, or the MetNet-2 (Meteorological Neural Network 2) DL model, the first featuring a weather forecasting range of up to 12 hours of lead time. The UNIME unit will first develop an epistemological analysis of DL simulations in the scientific context, such as weather forecasting, where EOCS has been traditionally developed. This will include defining whether DL simulations satisfy the definition of simulation provided by EOCS; identifying the different kinds of models involved in a DT simulation, such as a deterministic model of the target system, the DL model, the program executing the DL model, and the probabilistic state transition used for verification purposes; verifying whether ideal properties of simulation relations hold for DL simulations, such as full isomorphisms between mathematical models and the target system, correct implementation relations between computational models and their simulative programs, and bisimulation relations (or simulation relation in one of the two directions) between mathematical and computational models. This will allow to identify which model (or models) is validated with respect to which system (or systems), to which extent representational adequacy can be achieved, to which extent verification can be achieved, and how validation is related to the formal verification of the probabilistic state transition system. The analysis of the verification/validation problem for DL models will define the reliability of predictions of DL simulations over the simulated target systems.

In a second phase, the unit will extend such a framework for the case of DT simulations, wherein artefacts, rather than natural systems, are being simulated. DT simulations are therefore characterized by a specific scenario in which both the simulated and the simulating systems are artefacts. Industrial artefacts are developed implementing, among others, specifications coming from their DTs; at the same time, they also specify properties for their DTs. Another, crucial, feature of DT simulations is that artefacts and their DTs simulate each other, each predicting behavioural properties about the other. This is due to the fact that the systems, by being both artefacts, need to be both verified against each other. Accordingly, verification deserves a careful treatment for the case of DT simulations.

The epistemological and methodological analysis of the verification problem for DT simulations will serve as the basis for the logical work on the formal relations holding between the various different models involved, which will be carried out by the leading unit at LUCI Lab, University of Milan.

Nicola Angius
Department of Cognitive Science, University of Messina

BRIO Goes Into the Real World.

The BRIO project has produced extensive theoretical research addressing the formal, ontological, and technical issues related to bias, risk, and opacity in machine learning and deep learning systems. For an overview, see our publications page.

However, the project did not stop at conceptual contributions. One major deliverable was the development of a prototype software that implements the main strategies for bias detection and mitigation, as well as risk evaluation, proposed in our formal and philosophical contributions. This achievement was made possible with the help of Alkemy, our industrial partner. The open-source code for this implementation is available in a GIT repository .

BRIO is a tool designed to analyze algorithmic models to identify biases and risks and provide methods to mitigate them. It is intended for developers and data scientists, enabling them to define algorithms based on probabilistic learning mechanisms, identify incorrect behaviors related to biases, and gather data on these biases. The ultimate goal is to provide researchers with useful information and data to improve their artificial intelligence systems.

The tool takes as input the output of an AI model and a set of parameters selected by the user. The first input is encoded as a set of datapoints with their associated characteristics, and the second includes the designation of a sensitive characteristic. BRIO then outputs an evaluation of the likelihood that the examined AI model is biased concerning the user-designated sensitive characteristics.

The system carefully guides the user through the parameter-

setting process, explaining the conceptual meaning in detail while remaining customizable regarding the mathematical details of the analysis.

Almost all existing services aiming at bias and risk evaluation for AI systems fall into one or more of the following categories:

- Products that offer legal support.
- Products that certify the quality of the model based on the quality of the data (e.g., data obtained with consent).
- Products based on an analysis of the weight of each feature used (e.g., Shapley values, often statistical).
- Products that certify the quality of a known and accessible model.

BRIO distinguishes itself in two main aspects:

1. Model-Agnostic Approach: BRIO is independent of whether the person conducting the analysis has access to the model or the data it is trained on. This is crucial because observers (e.g., regulatory bodies) or users (e.g., customers) are rarely allowed to know the details of the model due to industrial secrecy. This feature makes BRIO very versatile, as it works independently of the model and is compatible with all models.
2. Logical-Formal Foundations: BRIO is based not only on statistical foundations but also on logical-formal ones. This allows for greater control over the structure and has been shown to reduce the exploration space of the analysis, providing a significant computational advantage.

Early testing of BRIO on credit data has been very successful, as reported in two recent publications, available here and here[here]. Encouraged by these results, we have decided to bring BRIO to the real world.

MIRAI is a spin-off of the Department of Philosophy of the University of Milan. Its main objective is to support responsible and trustworthy uses of AI systems by developing digital ecosystems for the verification, control, and supervision of data-driven and machine learning technologies, with a particular focus on fairness, bias, and transparency in compliance with legal and ethical criteria. The first and main product of MIRAI is the services provided through an enhanced version of the BRIO - Algorithmic Bias and Risk Detector.

The service will involve:

- Collecting information about the current AI models and data used in the client organization.
- Identifying and prioritizing the models and scenarios that need assessment.
- Inspecting data and models with appropriate MIRAI tools and methodologies to provide actionable metrics.
- Proposing a roadmap for mitigation and transition to responsible AI uses.
- Supporting the client during the roadmap implementation.

BRIO has entered the real world, transforming blue-sky research into concrete results.

Giuseppe Primiero
LUCI Lab, University of Milan and MIRAI

Courses and Programmes

Courses

LAIS: Logic for the AI Spring 2, 2–6 September, Como, Italy.

Programmes

MA in Human Centered Artificial Intelligence: University of Milan, Italy.
MA in Reasoning, Analysis and Modelling: University of Milan, Italy.
APhil: MA/PhD in Analytic Philosophy, University of Barcelona.
Master Programme: MA in Pure and Applied Logic, University of Barcelona.
Doctoral Programme in Philosophy: Department of Philosophy, University of Milan, Italy.
LogiCS: Joint doctoral program on Logical Methods in Computer Science, TU Wien, TU Graz, and JKU Linz, Austria.
HPSM: MA in the History and Philosophy of Science and Medicine, Durham University.
LoPhiSC: Master in Logic, Philosophy of Science and Epistemology, Pantheon-Sorbonne University (Paris 1) and Paris-Sorbonne University (Paris 4).
Master Programme: in Artificial Intelligence, Radboud University Nijmegen, the Netherlands.
Master Programme: Philosophy and Economics, Institute of Philosophy, University of Bayreuth.
MA in Cognitive Science: School of Politics, International Studies and Philosophy, Queen's University Belfast.
MA in Logic and the Philosophy of Mathematics: Department of Philosophy, University of Bristol.
MA Programmes: in Philosophy of Science, University of Leeds.
MA in Logic and Philosophy of Science: Faculty of Philosophy, Philosophy of Science and Study of Religion, LMU Munich.
MA in Logic and Theory of Science: Department of Logic of the Eotvos Lorand University, Budapest, Hungary.
MA in Mind, Brain and Learning: Westminster Institute of Education, Oxford Brookes University.
MA in Philosophy of Biological and Cognitive Sciences: Department of Philosophy, University of Bristol.
MA programmes: in Philosophy of Language and Linguistics, and Philosophy of Mind and Psychology, University of Birmingham.
MRes in Methods and Practices of Philosophical Research: Northern Institute of Philosophy, University of Aberdeen.
MSc in Applied Statistics: Department of Economics, Mathematics and Statistics, Birkbeck, University of London.
MSc in Artificial Intelligence: Faculty of Engineering, University of Leeds.
MSc in Cognitive & Decision Sciences: Psychology, University College London.
MSc in Cognitive Systems: Language, Learning, and Reasoning, University of Potsdam.
MSc in Cognitive Science: University of Osnabrück, Germany.
MSc in Cognitive Psychology/Neuropsychology: School of Psychology, University of Kent.
MSc in Logic: Institute for Logic, Language and Computation, University of Amsterdam.

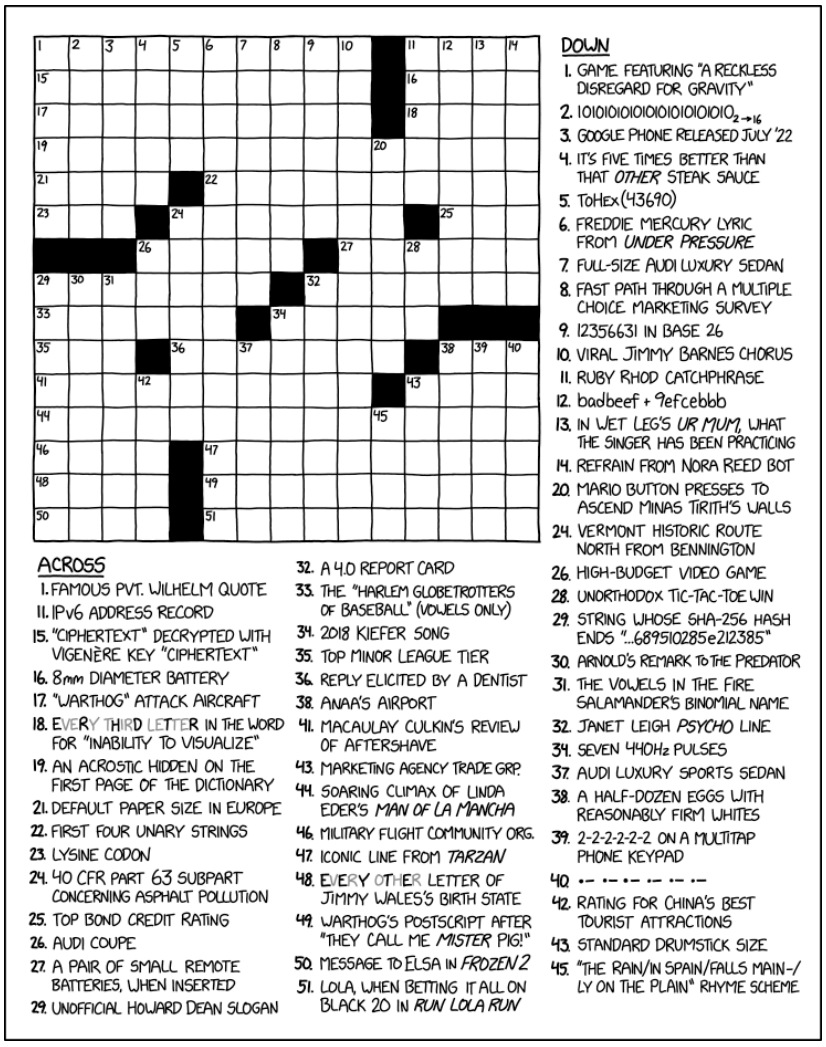
MSc in Mind, Language & Embodied Cognition: School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

MRes in Cognitive Science and Humanities: Language, Communication and Organization: Institute for Logic, Cognition, Language, and Information, University of the Basque Country (Donostia San Sebastián).

Research Master in Philosophy and Economics: Erasmus University Rotterdam, The Netherlands.

Doctoral Programme in Philosophy: Language, Mind and Practice, Department of Philosophy, University of Zurich, Switzerland.

MA in Philosophy: Dept. of Philosophy, California State University Long Beach.



ACROSS

1. FAMOUS PVT. WILHELM QUOTE
11. IPV6 ADDRESS RECORD
15. "CIPHERTEXT" DECRYPTED WITH VIGENÈRE KEY "CIPHERTEXT"
16. 8mm DIAMETER BATTERY
17. "WARTHOG" ATTACK AIRCRAFT
18. EVERY THIRD LETTER IN THE WORD FOR "INABILITY TO VISUALIZE"
19. AN ACROSTIC HIDDEN ON THE FIRST PAGE OF THE DICTIONARY
21. DEFAULT PAPER SIZE IN EUROPE
22. FIRST FOUR UNARY STRINGS
23. LYSINE CODON
24. 40 CFR PART 63 SUBPART CONCERNING ASPHALT POLLUTION
25. TOP BOND CREDIT RATING
26. AUDI COUPE
27. A PAIR OF SMALL REMOTE BATTERIES, WHEN INSERTED
29. UNOFFICIAL HOWARD DEAN SLOGAN
32. A 4.0 REPORT CARD
33. THE "HARLEM GLOBETROTTERS OF BASEBALL" (VOWELS ONLY)
34. 2018 KIEFER SONG
35. TOP MINOR LEAGUE TIER
36. REPLY ELICITED BY A DENTIST
38. ANAA'S AIRPORT
41. MACAULAY CULKIN'S REVIEW OF AFTERSHAVE
43. MARKETING AGENCY TRADE GRP.
44. SOARING CLIMAX OF LINDA EDER'S *MAN OF LA MANCHA*
46. MILITARY FLIGHT COMMUNITY ORG.
47. ICONIC LINE FROM *TARZAN*
48. EVERY OTHER LETTER OF JIMMY WALES'S BIRTH STATE
49. WARTHOG'S POSTSCRIPT AFTER "THEY CALL ME *MISTER* PIG!"
50. MESSAGE TO ELSA IN *FROZEN 2*
51. LOLA, WHEN BETTING IT ALL ON BLACK 20 IN *RUN LOLA RUN*

DOWN

1. GAME FEATURING "A RECKLESS DISREGARD FOR GRAVITY"
2. $10101010101010101010_{2\to16}$
3. GOOGLE PHONE RELEASED JULY '22
4. IT'S FIVE TIMES BETTER THAN THAT *OTHER* STEAK SAUCE
5. ToHex(43690)
6. FREDDIE MERCURY LYRIC FROM *UNDER PRESSURE*
7. FULL-SIZE AUDI LUXURY SEDAN
8. FAST PATH THROUGH A MULTIPLE CHOICE MARKETING SURVEY
9. 12356631 IN BASE 26
10. VIRAL JIMMY BARNES CHORUS
11. RUBY RHOD CATCHPHRASE
12. badbeef + 9efcebbb
13. IN WET LEG'S *UR MUM*, WHAT THE SINGER HAS BEEN PRACTICING
14. REFRAIN FROM NORA REED BOT
20. MARIO BUTTON PRESSES TO ASCEND MINAS TIRITH'S WALLS
24. VERMONT HISTORIC ROUTE NORTH FROM BENNINGTON
26. HIGH-BUDGET VIDEO GAME
28. UNORTHODOX TIC-TAC-TOE WIN
29. STRING WHOSE SHA-256 HASH ENDS "...689510285e212385"
30. ARNOLD'S REMARK TO THE PREDATOR
31. THE VOWELS IN THE FIRE SALAMANDER'S BINOMIAL NAME
32. JANET LEIGH *PSYCHO* LINE
34. SEVEN 440Hz PULSES
37. AUDI LUXURY SPORTS SEDAN
38. A HALF-DOZEN EGGS WITH REASONABLY FIRM WHITES
39. 2-2-2-2-2-2 ON A MULTITAP PHONE KEYPAD
40. •– •– •– •– •– •–
42. RATING FOR CHINA'S BEST TOURIST ATTRACTIONS
43. STANDARD DRUMSTICK SIZE
45. "THE RAIN/IN SPAIN/FALLS MAIN-/LY ON THE PLAIN" RHYME SCHEME