Fakultät für
Mathematik und Informatik

Matthias Thimm, Jürgen Landes und Kenneth Skiba (Eds.)

Proceedings of the First International Conference on Foundations, Applications, and Theory of Inductive Logic (FATIL2022)

Munich, Germany, October 12–14, 2022
Tagungsband

FernUniversität in Hagen

Proceedings of the First International Conference on Foundations, Applications, and Theory of Inductive Logic

(FATIL2022)

Matthias Thimm, Jürgen Landes, Kenneth Skiba (eds.)

Munich, Germany, October 12-14, 2022

Preface

Inductive reasoning is one of the most important reasoning techniques for humans and formalises the intuitive notion of "reasoning from experience". It has thus influenced both theoretical work on the formalisation of rational models of thought in Philosophy as well as practical applications in the areas of Artificial Intelligence and, in particular, Machine Learning.

The First International Conference on Foundations, Applications, and Theory of Inductive Logic (FATIL2022) brought together experts from all fields concerned with inductive reasoning. This includes in particular the following aspects:

Foundations of many of our best theories crucially depend on inductive logic and more widely induction. Uncertainty is ubiquitous in our lives and the philosophical problem arises to make sense of probabilities and to act sensibly in the face of uncertainties. General philosophy of science is much interested in (the reconstruction of) rational inference in general and in science, in particular, in cases with inconclusive evidence.

Theory of inductive inference can be developed within several traditions such as pure inductive logic or inductive logic based on the maximum entropy principle.

Applications have sprung from foundational thinking on induction in computer and data science. This includes aspects such as knowledge representation in multi-agent settings and machine learning approaches (such as inductive logic programming).

This conference welcomed contributions in all areas dealing with inductive reasoning. FATIL2022 was organised by the DFG Research Network "Foundations, Applications, and Theory of Inductive Logic"[1] and took place in Munich, Germany, 12-14th October 2022.

The conference received 15 submissions, which consisted of 4 full paper submissions and 11 extended abstracts. From these, 3 full papers and 9 extended abstracts were part of the final programme. In addition to that, the conference programme included keynote talks by Salem Benferhat, Luc De Raedt, Dov Gabbay, Henri Prade, Wolfgang Spohn, and Sandy Zabell, the abstracts of those keynote talks are also included in this volume.

October 2022 Matthias Thimm, Jürgen Landes and Kenneth Skiba (Chairs)

[1] `https://www.fatil.philosophie.uni-muenchen.de`

Organising Committee

General Chair
Matthias Thimm (University of Hagen)
Local Chair:
Jürgen Landes (Munich Center for Mathematical Philosophy)
Publicity Chair:
Kenneth Skiba (University of Hagen)

Programme Committee

Martin Adamcik (Assumption University of Thailand)
Christoph Beierle (FernUniversität in Hagen)
Salem Benferhat (Artois University)
Federico Cerutti (University of Brescia)
Claudia d'Amato (University of Bari)
Sébastien Destercke (CNRS, UMR Heudiasyc)
Andreas Herzig (CNRS, IRIT, University of Toulouse)
Hykel Hosni (University of Milan)
Simon Huttegger (University of California, Irvine)
Gabriele Kern-Isberner (TU Dortmund)
Kristian Kersting (TU Darmstadt)
Jens Lemanski (FernUniversität in Hagen)
Thomas Lukasiewicz (University of Oxford)
John D. Norton (University of Pittsburgh)
Nico Potyka (Imperial College London)
Henri Prade (IRIT - CNRS)
Marco Ragni (Technical University Chemnitz)
Tjitze Rienstra (Maastricht University)
Joe Roussos (Institute for Futures Studies, Stockholm)
Jan-Willem Romeijn (University of Groningen)
Ute Schmid (Universitaet Bamberg)
Sonja Smets (ILLC, University of Amsterdam)
Tom Sterkenburg (LMU Munich)
Frieder Stolzenburg (Harz University of Aplied Sciences)
Jon Williamson (University of Kent)
Francesca Zaffora Blando (Carnegie Mellon University)

Contents

Invited Talks

A story about the journey of reasoning with inconsistency under uncertainty

Sihem Belabbes[1] and Salem Benferhat[2]

[1] LIASD, IUT de Montreuil, Université Paris 8, Saint-Denis, France
belabbes@iut.univ-paris8.fr
[2] CRIL (Centre de Recherche en Informatique de Lens),
Université d'Artois and CNRS (UMR8188), Lens, France
benferhat@cril.fr

The sheer amount of the available information for reasoning and decision-making purposes makes its processing difficult to achieve in an efficient and reliable way. In particular, mining knowledge and answering queries from massive amounts of heterogeneous information brings additional challenges and raises new issues. Heterogeneity refers to data of different types, e.g. factual and structured data, unstructured textual data, images, maps, videos, sensitive personal data, and so on. Heterogeneity also refers to the various sorts of imperfections associated with the available data and knowledge. Indeed, information is often obtained from several (in)dependent sources which may have an incomplete, imperfect, erroneous and uncertain knowledge about the world. In addition, the sources may use their own models for representing the pieces of information as well as the relative preference and reliability levels thereof. Such scattered, fragmented, multi-source information needs to be merged and queried, in order to make data easily accessible. However, even if the information provided by each source is correct and consistent, concatenating multi-source information into a single knowledge base potentially causes inconsistency. Hence, standard inference mechanisms are trivialized because any conclusion can be drawn from a set of inconsistent pieces of information.

The problem of inconsistency management in knowledge bases has been a major focus at the confluence of several research areas, particularly in the domains of Databases, ontology engineering and logic-based knowledge representation, and which has met many application domains. Different research directions have been taken in identifying the sources of conflicts, dealing with inconsistent information and processing queries.

A well-accepted approach consists in constructing maximally consistent sub-bases (in terms of set inclusion) of the initial inconsistent knowledge base. Conclusions are often drawn by applying the standard inference mechanism of the representation language to those sub-bases (also called repairs) instead of to the initial inconsistent knowledge base. This approach is arguably the most natural. However, it is computationally difficult in the general case, and simply impractical in the case of massive datasets.

A tractable computational complexity may be achieved by replacing the inconsistent knowledge base by one (not necessarily maximal) of its consistent sub-bases through a unique choice for ignoring the problematic facts. The issue is to determine the most relevant facts that should be included in a sub-base and those that can be discarded. In order to avoid the random selection of a particular sub-base over others, often a single sub-base is obtained from the intersection of all the maximally consistent sub-bases of the inconsistent knowledge base.

Instead of ignoring the pieces of information that are involved in contradictions, another approach accommodates inconsistency by modifying instead the standard inference relation of the underlying representation language. This way, inconsistency is dealt with at the meta level and the conclusions that are drawn from the knowledge base are those that are supported by its most relevant elements. There is also an approach that restores the consistency of the knowledge base by rewriting some of its facts and adopting exception-tolerant reasoning.

Moreover, a significant body of work has focused on inconsistent knowledge bases that are equipped with a preference relation. Methods for handling inconsistency have been proposed for the case of total pre-orders and also for partial orders. In quantitative settings, inconsistency and uncertainty have been managed conjointly in various uncertainty frameworks such as possibility theory and ordinal conditional functions. However, tractability is obtained mostly for restricted languages (e.g. the lightweight ontology language DL-Lite).

This talk draws a broad picture and provides an analysis of the most prominent approaches for handling inconsistency in frameworks represented in Propositional Logic, Description Logics and Databases. We especially discuss the balance between the expressive power of the underlying representation language and the computational complexity of the reasoning task. We illustrate the problem of reasoning with inconsistent information on two case studies. In the first application, the dataset consists of videos of Southeast Asian traditional dances. The videos are annotated by several experts according to ontological knowledge, but the experts may have different opinions. Inconsistency arises when the same video receives contradictory annotations by different experts, and it needs to be handled in order to perform meaningful query answering. The second application is related to geographic information systems where the aim is to define new methods for representing, completing and querying heterogeneous and uncertain data concerning urban water and sanitation networks.

Acknowledgement

This research has received support from the ANR CROQUIS project (**C**ollecting, **R**epresenting, c**O**mpleting, merging, and **Q**uerying heterogeneous and **U**ncerta**I**n wa**S**tewater and stormwater network data), grant **ANR-21-CE23-0004** of the French research funding agency "Agence Nationale de la Recherche" (ANR).

From Probabilistic Logics to Neuro-Symbolic Artificial Intelligence

Luc De Raedt[1]

KU Leuven, Belgium

A central challenge to contemporary AI is to integrate learning and reasoning. The integration of learning and reasoning has been studied for decades already in the fields of statistical relational artificial intelligence and probabilistic programming. StarAI has focussed on unifying logic and probability, the two key frameworks for reasoning, and has extended this probabilistic logics machine learning principles. I will argue that StarAI and Probabilistic Logics form an ideal basis for developing neuro-symbolic artificial intelligence techniques. Thus neuro-symbolic computation = StarAI + Neural Networks. Many parallels will be drawn between these two fields and will be illustrated using the Deep Probabilistic Logic Programming language DeepProbLog.Speaker

Argumentation and Modal Public Announcement Logic

Dov Gabbay[1,2,3]

[1] University of Luxembourg, Luxenbourg
[2] King's College London, United Kingdom
[3] Bar Ilan University, Israel

Following the pioneering work of Davide Grossi, we translate argumentation into the modal logic K. There are two problems with such translations. First, some extensions such as the grounded extension and the preferred extensions require the use of second-order universal propositional quantifiers. Second, some semantics such as CF2 are originally defined by a recursive algorithm and therefore require translation of algorithms into modal logic. In this talk, we develop methodological methods of 1. translating/eliminating universal second-order propositional quantifiers and 2. we show how to translate algorithms into modal public announcement logic.

Transduction by logical analysis of data

Henri Prade

IRIT, CNRS & Université Paul Sabatier, Toulouse, France, prade@irit.fr

Abstract. Transduction is an inference operation which starting from a set of pieces of data makes a prediction for the missing value of another piece of data without relying on an induction step. The talk will discuss various ways of getting the best from data, with a special emphasis on the use of analogical proportions. Analogical proportions make parallel between two pairs of items on the basis of their differences and similarities.

Induction is a form of reasoning where general laws are inferred from a set of data. It is usually viewed as a matter of probability and statistics. Transduction directly infers a particular, factual conclusion from a set of data, without the need for an induction step followed by an instantiation of a general law. Without denying the clear interest of probabilities in this matter, this talk investigates another line of research based on the logical analysis of data.

A simple logical reading of examples considers them *one by one*; a bracketing of the logical description of a class can thus be obtained [8], or one may perform a Bayesian-like *possibilistic* classification (structurally similar to the standard Bayesian one) [7]. Nearest neighbors methods also handle examples one by one.

The basis of the analysis considered in this talk is the *comparison* of examples belonging to the same class as well as the comparison of examples belonging to different classes. This is closely related to the notion of ANALOGICAL PROPORTION (AP), namely statements linking 4 items (here the examples) of the form "a is to b as c is to d". An AP holds true if "a differs from b as c differs from d and b differs from a as d differs from c". AP's are also closely related to *taxonomic trees* [1,2] and *multi-valued dependencies* [10].

The presentation will include some historical references. It turns out that *Stuart Mill's rules of induction* for causal reasoning have an AP flavor [2]. Besides, his contemporary Gaspard Monge, in "An Elementary Treatise on Statics" was using numerically interpreted AP's extensively, for didactic purposes [13].

AP-based classifiers have a competitive accuracy on benchmarks [4], even if their complexity is basically cubic (some improvements are however possible for decreasing it). Theoretical results [5] [6] have shown that AP-based classifiers perform particularly well for linear functions in case of Boolean, or nominal attributes (but are not restricted to them). Moreover an improved analysis [3] of the comparisons between examples reveal regularities and differences in behavior that lead to completing the AP-based analysis by solving *Bongard problems* (problems that amount to find a property common to a set of items, but which is wrong for all the elements of another set). This helps to increase accuracy. The solving of Bongard problems is closely related to a special subclass of *logical proportions* [11] to which APs also belong.

Lastly, the potential of AP's for providing classifier-agnostic *explanations* will be emphasized [9]. Logic-based induction (as in version space learning) has the bad reputation of being sensitive to noisy data (still there may exist some possibilistic logic remedies [8]). The intended purpose of this talk is to show that there may exist neglected or ignored lines of research about induction / transduction, thus contributing to an enlarged understanding of this operation and perhaps opening the possibility of hybridizing logical and probabilistic methods.

Author's articles on which the talk is based

[1] N. Barbot, L. Miclet, H. Prade, G. Richard. A new perspective on analogical proportions. *Proc. 15th Europ. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'19)*, (G. Kern-Isberner, Z. Ognjanovic, eds.), Belgrade, Sept.18-20, LNCS 11726, Springer, 163-174, 2019.

[2] N. Barbot, L. Miclet, H. Prade, G. Richard. Analogical proportions and binary trees. In: *Logic in Question: Talks from the Annual Sorbonne Logic Workshop 2011-2019*, (J.-Y. Béziau, J.-P. Desclés, A. Moktefi, A. C. Pascu, eds.) Birkhäuser, 2022.

[3] M. Bounhas, H. Prade. Logical proportions-related classification methods beyond analogy. *Proc. 15th Int. Conf. Scalable Uncertainty Management (SUM'22)*, (F. Dupin de Saint-Cyr, M. Ozturk, N. Potyka, eds.) Paris, Oct.17-19, Springer, LNCS, 2022.

[4] M. Bounhas, H. Prade, G. Richard. Analogy-based classifiers for nominal or numerical data. *Int. J. Approx. Reason.*, 91: 36-55, 2017.

[5] M. Couceiro, N. Hug, H. Prade, G. Richard. Analogy-preserving functions: A way to extend Boolean samples. *Proc. 26th Int. Joint Conf. on Artificial Intelligence (IJCAI'17)*, (C. Sierra, ed.) Melbourne, Australia, Aug. 19-25, 1575-1581, 2017.

[6] M. Couceiro, E. Lehtonen, L. Miclet, H. Prade, G. Richard. When nominal analogical proportions do not fail. *Proc. 14th Int. Conf. on Scalable Uncertainty Management (SUM'20)*, (J. Davis, K. Tabia, eds.), virtual event / Bozen-Bolzano, Sept. 23-25, LNCS 12322, Springer, 68-83, 2020.

[7] D. Dubois, H. Prade. An overview of ordinal and numerical approaches to causal diagnostic problem solving. In: *Handbook of Defeasible Reasoning and Uncertainty Management Systems. Vol. 4. Abductive Reasoning and Learning*, (D. M. Gabbay, R. Kruse, eds.), Kluwer Acad. Publ., 231-280, 2000.

[8] D. Dubois, H. Prade. Towards a logic-based view of some approaches to classification tasks. *Proc. 18th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'20)*, M.-J. Lesot, S. M. Vieira, et al. , Lisbon, June 15-19, Part III., CCIS 1239, Springer, 697-711, 2020.

[9] S. Lim, H. Prade, G. Richard. Using analogical proportions for explanations. *Proc. 15th Int. Conf. on Scalable Uncertainty Management (SUM'22)*, (F. Dupin de Saint-Cyr, M. Ozturk, N. Potyka, eds.), Paris, Oct.17-19, Springer, LNCS, 2022.

[10] S. Link, H. Prade, G. Richard. Analogical proportions, multivalued dependencies and explanations. *Proc. 15th Int. Conf. on Scalable Uncertainty Management (SUM'22)*, (F. Dupin de Saint-Cyr, M. Ozturk, N. Potyka, eds.), Paris, Oct.17-19, Springer, LNCS, 2022.

[11] H. Prade and G. Richard. Homogenous and heterogeneous logical proportions. *IfCoLog J. of Logics and their Applications*, 1(1): 1-51, 2014.

[12] H. Prade, G. Richard. Analogical proportions: Why they are useful in AI. *Proc. 30th Int. Joint Conf. on Artificial Intelligence (IJCAI'21)*, (Z.-H. Zhou, ed.), virtual event / Montreal, Aug. 19-27, 4568-4576, 2021.

[13] H. Prade, G. Richard. Multiple analogical proportions. *AI Com.* 34: 211-228, 2021.

Some Strategic Considerations Concerning Inductive Logic

Wolfgang Spohn[1]

University of Konstanz, Germany

Abstract. The talk will reflect the situation of inductive logic rather than advance it. It will argue that non-deductive, inductive, or defeasible logic, reasoning, and argumentation is based on notions of conditional doxastic states and theories of doxastic change, and not the other way around. It will distinguish internal and external doxastic dynamics and focus on the external one rationally driven by evidence. It will argue that the external dynamics should be conceived as proceeding in steps (akin to a Markov process) and as starting from an initial state. This will require to distinguish an unrevisable and a defeasible apriori. This distinction finally points a way of dealing with conceptual changes (or at least expansions) in inductive logic. Each step of the talk is contested in the literature. So, it should have a critical as well as a constructive potential.

My talk will be devoted to a critical reflection of the current situation in inductive logic; it is not supposed to advance the field on a particular issue. I shall not discriminate between logic, reasoning, and argumentation. At the strategic level of my talk this may count as the same. Also, I shall use "inductive" in the wide sense of "non-deductive". Hence, I also do not distinguish between inductive logic (which may be said to be about degrees of confirmation) and non-monotonic reasoning (which attempts to arrive at defeasible yes-no conclusions). Throughout, I want to address inductive logic in this wide sense. We know that it is a variegated, ramified, contested, and little canonized field. The first point I want to argue is that there is a close relation between inductive logic and the dynamics of rational doxastic states. It may be disputed whether one side is the more basic one and, if so, which. I shall argue that the dynamics of doxastic states is more basic, and with it the notion of conditional doxastic states. The hope to ground rational doxastic dynamics in a (more objective?) account of good reasons or valid inferences or good arguments is futile. The second point is that we should distinguish between an internal and an external dynamic of doxastic states. The internal dynamic is about inference rules and reasoning processes, about elaborating the consequences of given premises in a given internal environment. The external dynamic is about what to conclude when the internal environment changes through external influence. Rationally, the external dynamic is driven only by new evidence. I want to argue that the external dynamic is the more basic one. This will include some clarifying remarks concerning the notion of evidence. The third point is to argue for the traditional picture of the external dynamic as consisting in the iteration of the basic triad: prior doxastic state + total evidence in between determines posterior doxastic state. This picture involves a regress to some initial doxastic state, sometimes also called the urprior. On the one hand, this will entail a criticism of both the view that total evidence alone can determine the posteriors and the view that the posterior state is determined by the urprior and the total evidence ab ovo (which would be the only changing item according to this view). On the other hand, the picture requires saying more about this urprior. In a Kantian spirit, one may equate it with a priori knowledge. However, it is important to distinguish the unrevisable a priori (which is the Kantian one) and the relativized defeasible a priori (which characterizes what I called the initial doxastic state; the urprior is presumably intended to be less realtivized). In the fourth and final part of the talk I shall discuss the attempts at rationally constraining the urprior or initial doxastic state. Here, Carnap's attempts at an inductive logic and Hintikka's amendment as well as objective Bayesianism find their place. On the whole there are not many offers. A crucial issue raised by the relativized nature of the defeasible a priori will be the issue of conceptual extension, which I find hardly treated in the literature. As a final dialectical step, I would like to defend the idea (not found in the literature) that new concepts come along with a lot of defeasible a priori baggage.

Forensic DNA identification evidence

Sandy Zabell[1]

Northwestern University, USA

The use of DNA identification evidence during the last several decades has revolutionized forensic science. But with the increasing complexity of the systems that are now being employed, a number of foundational challenges have come to the surface: the appropriate statistic to summarize the strength of the evidence, dealing with complex samples, searching data bases, safeguarding individual privacy, and effectively communicating results to a lay audience. In this survey talk, after describing the current system most commonly in use, I will present a typology of these issues, focusing on the recent advent of so-called probabilistic genotyping systems, which fit high-dimensional models, sometimes using Markov chain Monte Carlo estimation, and which can be technically challenging to explain to the trier of fact.

Papers

Full Papers

An Intuitive Generalisation of Information Geometry*

Martin Adamčík

Independent Researcher
maths38@gmail.com

Abstract. In this paper we recover some traditional results in geometry of probability distributions, and in particular the convergence of the alternating minimisation procedure, without actually referring to probability distributions. We will do this by discussing a new general concept of two types of points; admissible and agreeable, inspired by multi–agent uncertain reasoning. On one hand, this presents a unique opportunity to make traditional results accessible to a wider audience as no prior knowledge of the topic is required. On the other hand, it allows us to contemplate how a group of humans would seek an agreement without necessarily expressing it in terms of probability distributions, focusing instead on properties. Finally, we recover the traditional setting of probability distributions, including cross–entropy, in the appendix.

Keywords: Information geometry · Divergence · Uncertain reasoning · Cross–Entropy · Fixed point · Alternating minimisation

1 Intuition

"A point is that which has no part."

Euclid of Alexandria, [11]

Whenever we build a mathematical theory we need to consult our intuition. Should we not do it we may end up building a theory that little resembles the world we are living in, and which is equally inapplicable. In this section, we will start building an intuitive framework that deals with information. We will need to confer with our intuition in the form of our experience on how information is used and how conflicting statements are dealt with.

Our first notion will be indeed the *point*. As in the Euclidean definition that starts this section, it is a building block that is further indivisible. Our point is, however, introduced to represent information rather than the position in a three–dimensional world. We think of several different opinions on a particular matter; each different opinion can be represented as a point. We are not concerned with

* Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — 432308570.

what further constitutes the opinion and we disregard any knowledge concerning the origins of the opinion, it is simply an indivisible entity to us.

The points, which we have just introduced, can have any of the two following properties in this paper:

1. They can represent an admissible collective point of view of a given group of humans or collection of information sources, shortly called simply an *admissible point*,
2. and to represent an agreement of the group, shortly called an *agreeable point*.

Now, intuitively, an admissible point is meant only to represent the state of collective knowledge, individual members of the group could well disagree and there could be no, what we call, agreeable point. This collective framework of unceretain reasoning was pioneered by Wilmers [15], and we are directly extending it here. An illustration is in Figure 1, all figures can be found in the appendix.

Example. *To illustrate, one scientific study could suggest that the proportion of people that develop a particular disease is somewhere between* 10% *and* 30% *while the other study could indicate that this value is between* 20% *and* 50%. *One way of constructing a point is to specify an ordered pair of individually admissible proportions such as* $(25\%, 40\%)$, *where the first number is admissible according to the first study and the second number is admissible according to the second study. Agreeable admissible points in this particular representation will be the points* (x, x), $x \in [20\%, 30\%]$, *clearly representing the proportions on which the studies agree at the same time. There are other agreeable points that are not admissible, such as* $(35\%, 35\%)$, $(50\%, 50\%)$, $(0\%, 0\%)$ *and so on.*

The example above illustrates the kind of details we will need to go into before the intuitive concept that we develop here can be applied, but at this stage working out the details would only obstruct the general idea and the intuition behind it. We have therefore moved all technical examples and relevant references to the appendix. Here we only point out that our illustration fits Paris–Vencovská framework of uncertain reasoning as in [13].

The previous example was also straightforward enough in establishing agreeable points, but the following questions naturally arise:

1. What shall we do if admissible points contain no agreeable points?
2. How should we measure some kind of distance between an admissible point and an agreeable point in an effort to find a closest point of agreement?
3. Which intuitive principles such a notion of distance should satisfy?

We shall find the answers in this paper.

2 Information Divergence

In the previous section we saw the need for expressing some sort of information distance between two points, but we would not want to require much from this notion at this early stage. In particular, there is no apparent need for it to be a *metric*.

A metric is a symmetric distance between a pair of elements $\mathbf{x}, \mathbf{y}$ of a set. It assigns to each pair $(\mathbf{x}, \mathbf{y})$ a non–negative real number $d(\mathbf{x}, \mathbf{y})$, this number is independent on the order of elements, it is zero if and only if the elements are identical and it satisfies the triangular inequality; $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$.

Instead, we will consider a much weaker notion of *information divergence*, a mapping D that assigns an ordered pair of points a non–negative real number;

$$D(\mathbf{x}, \mathbf{y}) \geq 0.$$

We say that $D(\mathbf{x}, \mathbf{y})$ is the D *information divergence* from $\mathbf{x}$ to $\mathbf{y}$. Since the symmetry is not required, the D information divergence from $\mathbf{y}$ to $\mathbf{x}$ could be different and therefore we do not call it a distance but a divergence.

Now, let W be the set of all admissible points and V the set of agreeable points. Throughout the paper we will assume that they are both non–empty. Let $\Delta(W)$ be the set of all those agreeable points $\mathbf{v}$ that are such that $D(\mathbf{v}, \mathbf{w})$ is minimal subject to $\mathbf{v} \in V$ and $\mathbf{w} \in W$. In other words, we are looking here at all pairs $(\mathbf{v}, \mathbf{w})$, $\mathbf{v} \in V$ and $\mathbf{w} \in W$, establishing the minimal $D(\mathbf{v}, \mathbf{w})$ if it exits, and collecting all those $\mathbf{v}$ from V that give this minimal divergence into $\Delta(W)$. The purpose of the set $\Delta(W)$ is to determine those agreeable points that have the smallest D information divergence from them to admissible points and to use them as representatives of the set of all admissible points W. In other words, $\Delta(W) \subseteq V$ represents W; it is the agreement of a group of humans or collection of information sources. We will call the points in $\Delta(W)$ *representative points*. See Figure 2 for an illustration.

Intuitively, if $W \cap V \neq \emptyset$; i.e., there are agreeable admissible points, we expect the representation $\Delta(W)$ of W to be formed only by agreeable admissible points, although this ituition is not universally accepted [14]. The following property of D guarantees that this is the case:

Property 1 (Consistency). *Let* $\mathbf{v}$ *and* $\mathbf{w}$ *be any two points. Then*

$$D(\mathbf{v}, \mathbf{w}) = 0 \textit{ if and only if } \mathbf{v} = \mathbf{w}.$$

Observation 1. *Let D be such that it satisfies the consistency property. If there are agreeable admissible points then agreeable admissible points form all representative points;*

$$\textit{if } W \cap V \neq \emptyset \textit{ then } \Delta(W) = W \cap V.$$

Proof. First, if $\mathbf{v} \in W \cap V$ then by the consistency property $D(\mathbf{v}, \mathbf{v}) = 0$. We conclude that $\mathbf{v} \in \Delta(W)$ as $\mathbf{v}$ minimises $D(\mathbf{v}, \mathbf{w})$ subject to $\mathbf{v} \in V$ and $\mathbf{w} \in W$. (Note that $D(\mathbf{v}, \mathbf{v})$ cannot be smaller than zero by the definition.) Hence $\Delta(W) \supseteq W \cap V$.

Second, assume that $W \cap V \neq \emptyset$ and $\mathbf{v} \in \Delta(W) \subseteq V$ is such that $\mathbf{v} \notin W$. Then $D(\mathbf{v}, \mathbf{w}) = 0$ for some $\mathbf{w} \in W$, which by the consistency principle gives $\mathbf{v} = \mathbf{w}$. Hence $\Delta(W) \subseteq W \cap V$. □

The consistency property above is formulated more strongly than it is needed to prove Observation 1. Rather than considering any points $\mathbf{v}$ and $\mathbf{w}$, we could have required it only for $\mathbf{v} \in V$ and $\mathbf{w} \in W$. The reason for our choice is that we will need the stronger version later on.

In contrast, if $\mathbf{v} = \mathbf{w}$ implies $D(\mathbf{v}, \mathbf{w}) = 0$ but there are $\mathbf{v} \neq \mathbf{w}$ such that $D(\mathbf{v}, \mathbf{w}) = 0$, it could be possible to have $W \cap V \neq \emptyset$ and $\Delta(W) \supsetneqq W \cap V$, so further weakening of the consistency property would be undesirable.

3 Projections

Our notion of an information divergence is too general to have further useful properties on its own; in particular, if there are no agreeable admissible points we cannot even say that the set of all representative points is always non–empty. We will keep adding assumptions concerning both D and sets of agreeable and admissible points V and W based on what appears rational to us in the context of information geometry. At some point, however, we will need to show that the list of our assumptions is consistent; we will need to find a particular information divergence, and sets W and V that satisfy all those assumptions.

In this section we will require D to have the following properties:

Property 2 (Projection). *Assume that* $\mathbf{v}$ *is an agreeable point. Then there is a unique admissible point* $\mathbf{w}$ *such that* $D(\mathbf{v}, \mathbf{w})$ *is minimal subject to* $\mathbf{w} \in W$.

The unique point $\mathbf{w}$ from the previous property will be denoted $\pi_W(\mathbf{v})$; it is the *D–projection* of $\mathbf{v}$ into W. An illustration is in Figure 3.

Property 3 (Conjugated Projection). *Assume that* $\mathbf{w}$ *is an admissible point. Then there is a unique agreeable point* $\mathbf{v}$ *such that* $D(\mathbf{v}, \mathbf{w})$ *is minimal subject to* $\mathbf{v} \in V$.

The unique point $\mathbf{w}$ from the previous property will be denoted $\widehat{\pi}_V(\mathbf{w})$; it is the *conjugated D–projection* of $\mathbf{w}$ into V. An illustration is in Figure 4.

Intuitively, if we present a group of humans with a point of agreement, we expect them to find a single point among those they consider admissible as their personal opinion in view of the presented agreement. On the other hand, we should be able to establish agreement regardless on which specific admissible point the group presents to us.

Taking this further, the following process taken from [1] and inspired by an earlier version of [15] could resemble a real life agreement seeking:

Example. *Consider a group of humans with their set of admissible points* W. *The group elects a committee whose task is to find a single agreeable point from the set* V. *Naturally, the committee presents the group with their personal opinion or any other provisional starting point* $\mathbf{v}_0$ *that they see appropriate. The group then decides which point from those they consider admissible must have been the case to reach the conclusion suggested by the committee; they project* *the committee's point to their set W. At this stage, being present with a single admissible point, it is now possible for the committee to determine the conjugated projection of that admissible point to the set V; finding the corresponding agreeable point $\mathbf{v}_1$ of the group. Now, it is not at all necessary that $\mathbf{v}_1 = \mathbf{v}_0$. Nevertheless, the committee would be compelled to iterate the whole process until the above process stabilises on a single agreeable point.*

The points of interest from the previous example, although at this stage it is not clear if they even exist, will be called *fixed points*. More explicitly, an agreeable point $\mathbf{v} \in V$ is a fixed point if

$$\widehat{\pi}_V(\pi_W(\mathbf{v})) = \mathbf{v}.$$

The set of all fixed points will be denoted $\Theta(W)$. See Figure 5 for an illustration.

The example above is of course only one possible way of finding an agreement, although we argue that it is a rational one. An interesting question is how does this way relate to previously suggested information divergence D minimisation, which yields the set $\Delta(W)$. There could be something:

Observation 2. *Let D be such that it satisfies the projection and conjugated projection properties. Then representative points are also fixed points;*

$$\Delta(W) \subseteq \Theta(W).$$

Proof. Let $\mathbf{v} \in \Delta(W)$ and let d be the smallest D information divergence $D(\mathbf{v}, \mathbf{w})$ subject to $\mathbf{v} \in V$ and $\mathbf{w} \in W$. Such a real number exists by the definition of $\Delta(W)$ and note that in this paper we always assume that both V and W are non–empty.

Clearly, $D(\mathbf{v}, \pi_W(\mathbf{v})) \geq d$. Now assume that $D(\mathbf{v}, \pi_W(\mathbf{v})) > d$ so there must be $\mathbf{w} \in W$ such that $D(\mathbf{v}, \pi_W(\mathbf{v})) > D(\mathbf{v}, \mathbf{w})$. But this contradicts the definition of $\pi_W(\mathbf{v})$. So it must be that

$$D(\mathbf{v}, \pi_W(\mathbf{v})) = d.$$

Now, assume that $\widehat{\pi}_V(\pi_W(\mathbf{v})) \neq \mathbf{v}$. Nevertheless,

$$D(\widehat{\pi}_V(\pi_W(\mathbf{v})), \pi_W(\mathbf{v})) = D(\mathbf{v}, \pi_W(\mathbf{v})) = d,$$

otherwise we would contradict the definition of $\widehat{\pi}_V(\pi_W(\mathbf{v}))$. Finally, the equation above implies that both $\mathbf{v}$ and $\widehat{\pi}_V(\pi_W(\mathbf{v}))$ minimise $D(\mathbf{v}, \pi_W(\mathbf{v}))$ subject to $\mathbf{v} \in V$. Such a minimiser is, however, by the conjugated projection property required to be unique, thus

$$\widehat{\pi}_V(\pi_W(\mathbf{v})) = \mathbf{v}.$$

□

It seems that after concluding this section we have more questions than answers:

1. What properties should we require from an information divergence D and sets W, V so that $\Delta(W) = \Theta(W)$? Is it even possible?
2. If we iterate the process from the example above; i.e., create a sequence $\{\mathbf{v}_i\}_{i=0}^{\infty}$, where $\mathbf{v}_{i+1} = \widehat{\pi}_V(\pi_W(\mathbf{v}_i))$, what properties should we require from the information divergence D and sets W and V so that we find an agreement in that way?

We shall find answers in the following sections.

4 Pythagorean Properties

The following property informally says that a group might establish the divergence of their agreement to an admissible point by adding their divergence to the projection of that point to the set of agreeable points and the divergence of the projection to the point concerned.

Property 4 (Pythagorean for Agreeable Points). *Let $\mathbf{v} \in V$ be an agreeable point and $\mathbf{w} \in W$ be an admissible point. Then*

$$D(\mathbf{v}, \widehat{\pi}_V(\mathbf{w})) + D(\widehat{\pi}_V(\mathbf{w}), \mathbf{w}) = D(\mathbf{v}, \mathbf{w}).$$

This property is counter–intuitive from the point of view of the classical Euclidean distance. Although it does not violate the triangular inequality, it is certainly not a property of a distance we are used to. On the other hand, it quite closely resembles how squares taken over the sides of a right–angled triangle behave in the Euclidean geometry (hence the name), see Figures 7 and 8 for an illustration.

Intuitively, using an analogy from the Euclidean geometry, we expect the set of agreeable points in respect to the conjugated D–projection to behave as a flat space into which we projects admissible points. This is quite a strong requirement, we would not want to be so harsh on the set of admissible points. The following property will make admissible points to behave as a convex set.

Property 5 (Pythagorean for Admissible Points). *Let $\mathbf{v} \in V$ be an agreeable point and $\mathbf{w} \in W$ be and admissible point. Then*

$$D(\mathbf{v}, \pi_W(\mathbf{v})) + D(\pi_W(\mathbf{v}), \mathbf{w}) \leq D(\mathbf{v}, \mathbf{w}).$$

This property is similar to the Pythagorean property for agreeable points but it is weaker. And if the inequality from the statement actually holds in some case for a particular D then this information divergence D is not a metric. See Figures 9 and 10 for an illustration.

The following observation gives us something that also follows from the consistency property on Page , but without assuming it.

Observation 3. *Let D be such that it satisfies the projection and conjugated projection properties, and the Pythagorean properties for agreeable and admissible points. If $\mathbf{v} \in V$ is a fixed point then $D(\mathbf{v}, \mathbf{v}) = 0$ and $D(\pi_W(\mathbf{v}), \pi_W(\mathbf{v})) = 0$.*

Proof. By the Pythagorean property for agreeable points

$$D(\mathbf{v}, \widehat{\pi}_V(\pi_W(\mathbf{v}))) + D(\widehat{\pi}_V(\pi_W(\mathbf{v})), \pi_W(\mathbf{v})) = D(\mathbf{v}, \pi_W(\mathbf{v})).$$

But since $\mathbf{v}$ is fixed the above is equivalent to

$$D(\mathbf{v}, \mathbf{v}) + D(\mathbf{v}, \pi_W(\mathbf{v})) = D(\mathbf{v}, \pi_W(\mathbf{v})),$$

which is possible only if $D(\mathbf{v}, \mathbf{v}) = 0$.

Similarly, by the Pythagorean property for admissible points

$$D(\mathbf{v}, \pi_W(\mathbf{v})) + D(\pi_W(\mathbf{v}), \pi_W(\mathbf{v})) \leq D(\mathbf{v}, \pi_W(\mathbf{v})),$$

which is possible, due to non–negativity of information divergence, only if

$$D(\pi_W(\mathbf{v}), \pi_W(\mathbf{v})) = 0.$$

□

5 Fixed Points are Representative Points

The following natural property says that the D information divergence from one admissible point to another admissible point should not be smaller than the D information divergence from and to the corresponding agreeable points. Intuitively, seeking an agreement should not take us further apart, see Figure 11.

Property 6 (Convexity). *Let* $\mathbf{w}, \mathbf{u} \in W$. *Then*

$$D(\mathbf{w}, \mathbf{u}) \geq D(\widehat{\pi}_V(\mathbf{w}), \widehat{\pi}_V(\mathbf{u})).$$

We have now all the tools sufficient to prove that fixed points are also representative points, if there is actually a representative point.

Theorem 1 (Characterisation). *Let D be such that it satisfies the projection and conjugated projection properties, the Pythagorean properties for both admissible and agreeable points, and the convexity property. If a representative point exists then the set of fixed points and the set of representative points are equal;*

$$\varDelta(W) = \varTheta(W).$$

Proof. By Observation 2 on Page we already know that $\varDelta(W) \subseteq \varTheta(W)$ so it is sufficient to show that $\varDelta(W) \supseteq \varTheta(W)$.

Because we assumed that a representative point exists and we already know that every representative point is also a fixed point, we may assume that $\widehat{\pi}_V(\mathbf{w}) \in \varDelta(W)$, for some $\mathbf{w} \in W$. To make the argument, we now also assume that $\mathbf{v} \in \varTheta(W)$ and show that $\mathbf{v} \in \varDelta(W)$ in what follows.

The Pythagorean property for representative points

$$D(\mathbf{v}, \widehat{\pi}_V(\mathbf{w})) + D(\widehat{\pi}_V(\mathbf{w}), \mathbf{w}) = D(\mathbf{v}, \mathbf{w})$$

and the Pythagorean property for admissible points

$$D(\mathbf{v},\mathbf{w}) \geq D(\mathbf{v},\pi_W(\mathbf{v})) + D(\pi_W(\mathbf{v}),\mathbf{w})$$

give

$$D(\mathbf{v},\widehat{\pi}_V(\mathbf{w})) + D(\widehat{\pi}_V(\mathbf{w}),\mathbf{w}) \geq D(\mathbf{v},\pi_W(\mathbf{v})) + D(\pi_W(\mathbf{v}),\mathbf{w}), \qquad (1)$$

see Figure 12 for an illustration.

Since $\mathbf{v}$ is a fixed point and hence $\mathbf{v} = \widehat{\pi}_V(\pi_W(\mathbf{v}))$, by the convexity property

$$D(\mathbf{v},\widehat{\pi}_V(\mathbf{w})) \leq D(\pi_W(\mathbf{v}),\mathbf{w}). \qquad (2)$$

Now, Equations 1 and 2 give

$$D(\widehat{\pi}_V(\mathbf{w}),\mathbf{w}) \geq D(\mathbf{v},\pi_W(\mathbf{v})).$$

Since $\mathbf{w} \in \varDelta(W)$ the above must hold with equality and therefore $\mathbf{v} \in \varDelta(W)$.

□

The proof above was based on ideas from [2].

Observation 4. *Let D be such that it satisfies the projection and conjugated projection properties, the Pythagorean properties for both admissible and agreeable points, and the convexity property. Let $\mathbf{v},\mathbf{u} \in \varDelta(W) = \varTheta(W)$. Then*

$$D(\mathbf{v},\mathbf{u}) = D(\pi_W(\mathbf{v}),\pi_W(\mathbf{u})).$$

Proof. Looking at (1) in the previous proof, which employed the identical assumptions, and taking $\mathbf{u} = \widehat{\pi}_V(\mathbf{w})$, we obtain

$$D(\mathbf{v},\mathbf{u}) + D(\mathbf{u},\mathbf{w}) \geq D(\mathbf{v},\pi_W(\mathbf{v})) + D(\pi_W(\mathbf{v}),\mathbf{w}).$$

Since $\mathbf{v},\mathbf{u} \in \varDelta(W)$ we have $D(\mathbf{u},\mathbf{w}) = D(\mathbf{v},\pi_W(\mathbf{v}))$ and the above becomes

$$D(\mathbf{v},\mathbf{u}) \geq D(\pi_W(\mathbf{v}),\mathbf{w}).$$

Finally, by the convexity property, the above is possible only with the equality.

□

6 Enter Metric Topology

> "Every reasonable non–pathological space in topology will turn out to be a metric space. On the other hand, developments (. . .) showed there was a need to study a more general class of spaces than merely Euclidean spaces."

Donal W. Kahn, [12]

Thus far we have avoided the need to introduce any topological structure on the set of all points, but this is going to change in this section. First, let us finally introduce a symbol for the set of points here considered as X. Then, let us equip the set of points X with a metric $d(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are any points. Recall that the notion of metric was discussed on Page .

We say that a sequence $\{\mathbf{v}_i\}_{i=1}^{\infty}$ of points *converges* to a point $\mathbf{v}$ if for any real number $\epsilon > 0$ there is j such that $d(\mathbf{v}_i, \mathbf{v}) < \epsilon$ for all $i > j$. We call such a $\mathbf{v}$ a *limit point*.

What we need to establish now is a connection between the metric d and the divergence D, which is a mapping from a Cartesian product $X \times X$ to $\mathbb{R}$;

$$D : X \times X \to \mathbb{R}.$$

Therefore, we need to have a product metric

$$d_p((\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2)) = \Big([d(\mathbf{x}_1, \mathbf{x}_2)]^p + [d(\mathbf{y}_1, \mathbf{y}_2)]^p\Big)^{\frac{1}{p}},$$

where $1 \leq p < \infty$, in place. Then we can define that a mapping $f : X \times X \to \mathbb{R}$ is *continuous*, if for any real number $\epsilon > 0$ there is $\delta > 0$ such that whenever $d_p((\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2)) < \delta$ we have $|f(\mathbf{x}_1, \mathbf{y}_1) - f(\mathbf{x}_2, \mathbf{y}_2)| < \epsilon$. The last expression is just the standard metric on $\mathbb{R}$, and our definition follows the usual definition of continuity of a mapping between metric spaces. The connection we were looking for is then the following.

Property 7 (Continuity). *D is continuous.*

Intuitively, the property above says that if two pairs of points are close to each other in the product metric, then D does not rip them apart in $\mathbb{R}$.

The following is a straightforward and intuitive consequence of D being continuous, and it is how we will employ continuity to obtain future the results.

Observation 5. *Let D satisfy the continuity property. Assume that a sequence of agreeable points $\{\mathbf{v}_i\}_{i=1}^{\infty}$ converges to $\mathbf{v}$ and a sequence of admissible points $\{\mathbf{w}_i\}_{i=1}^{\infty}$ converges to $\mathbf{w}$. Then the sequence*

$$\{D(\mathbf{v}_i, \mathbf{w}_i)\}_{i=1}^{\infty}$$

converges to $D(\mathbf{v}, \mathbf{w})$.

Proof. For any $\epsilon > 0$ we are tasked with finding j such that $|D(\mathbf{v}_i, \mathbf{w}_i) - D(\mathbf{v}, \mathbf{w})| < \epsilon$ for all $i > j$. Since D is continuous, for any $\epsilon > 0$ there is $\delta > 0$ such that whenever

$$\Big([d(\mathbf{v}_i, \mathbf{v})]^p + [d(\mathbf{w}_i, \mathbf{w})]^p\Big)^{\frac{1}{p}} < \delta$$

we have $|D(\mathbf{v}_i, \mathbf{w}_i) - D(\mathbf{v}, \mathbf{w})| < \epsilon$. Now we simply select j so that $[d(\mathbf{v}_i, \mathbf{v})]^p + [d(\mathbf{w}_i, \mathbf{w})]^p < \delta^p$ for all $i > j$. This is always possible since $\{\mathbf{v}_i\}_{i=1}^{\infty}$ converges to $\mathbf{v}$ and $\{\mathbf{w}_i\}_{i=1}^{\infty}$ converges to $\mathbf{w}$. $\square$

Since we operate in a metric space, we can define that a subset of points X is *compact* if every sequence that can be constructed from its elements has a convergent subsequence and the limit point of this convergent subsequence lies in this subset. In other words, it has Bolzano–Weierstrass property, which in metric spaces is equivalent to compactness.

Observation 6. *If V and W are compact, and D satisfies the continuity property then a representative point exists.*

Proof. Consider the set of all real numbers $D(\mathbf{v},\mathbf{w})$ such that $\mathbf{v}\in V$ and $\mathbf{w}\in W$. This set is bounded from below so it has also the greatest lower bound (a basic property of real numbers). Let us denote it b.

Now, for every $\epsilon>0$ there are $\mathbf{v}\in V$ and $\mathbf{w}\in W$ such that

$$\epsilon+b>D(\mathbf{v},\mathbf{w})\geq b,$$

otherwise b would not be the greatest lower bound. Therefore, we can construct a sequence $\{D(\mathbf{v}_i,\mathbf{w}_i)\}_{i=1}^{\infty}$ that converges to b. Now, due to the compactness of V the sequence $\{\mathbf{v}_i\}_{i=1}^{\infty}$ has a convergent subsequence, say $\{\mathbf{v}_{i_j}\}_{j=1}^{\infty}$. Let $\{\mathbf{w}_{i_j}\}_{j=1}^{\infty}$ be the corresponding sequence in W, which is also compact, so it has also a convergent subsequence. Let $\mathbf{w}\in W$ be its limit point and let $\mathbf{v}\in V$ be the limit point of $\{\mathbf{v}_{i_j}\}_{j=1}^{\infty}$. Then due to Observation 5

$$D(\mathbf{v},\mathbf{w})=b$$

so $\mathbf{v}$ must be a representative point. □

Looking now at the statement of Theorem 1 on Page we can replace the requirement for existence of a representative point by requiring compactness of V and W, and asking D to satisfy the continuity property.

7 Convergence

The following property will be needed to prove that a representative point can be reached by an iterative process.

Property 8 (Four Points). *Let $\mathbf{w},\mathbf{u}\in W$ and $\mathbf{v}\in V$. Then*

$$D(\widehat{\pi}_V(\mathbf{w}),\mathbf{u})\leq D(\mathbf{w},\mathbf{u})+D(\mathbf{v},\mathbf{u}).$$

The four–point property is illustrated in Figure 13.

Thus far we had one property that linked the concept of divergence D and the metric topology given by d; it was the continuity property. Here we provide another one, which somewhat goes in the opposite direction.

Property 9 (Connectivity). *If $\{D(\mathbf{v}_i,\mathbf{v})\}_{i=1}^{\infty}$ converges to zero then so does $\{d(\mathbf{v}_i,\mathbf{v})\}_{i=1}^{\infty}$.*

Naturally, the property above implies that if $\{D(\mathbf{v}_i, \mathbf{v})\}_{i=1}^{\infty}$ converges to zero then $\mathbf{v}$ is the limit point of $\{\mathbf{v}_i\}_{i=1}^{\infty}$. We will use this in the following theorem.

Theorem 2 (Convergence). *Let D be such that it satisfies the projection and conjugated projection properties, the Pythagorean properties for both admissible and agreeable points, and the consistency, convexity, continuity, four–point and connectivity properties. Let $\mathbf{v}_0 \in V$. Define a sequence $\{\mathbf{v}_i\}_{i=0}^{\infty}$ recursively by $\mathbf{v}_{i+1} = \widehat{\pi}_V(\pi_W(\mathbf{v}_i))$. If V and W are compact then the sequence $\{\mathbf{v}_i\}_{i=0}^{\infty}$ converges to a fixed point.*

Proof. First notice that

$$D(\mathbf{v}_i, \pi_W(\mathbf{v}_i)) \geq D(\pi_W(\mathbf{v}_i), \widehat{\pi}_V(\pi_W(\mathbf{v}_i))) \geq$$

$$\geq D(\widehat{\pi}_V(\pi_W(\mathbf{v}_i)), \pi_W(\widehat{\pi}_V(\pi_W(\mathbf{v}_i))))$$

so the sequence of non–negative real numbers $D(\mathbf{v}_i, \pi_W(\mathbf{v}_i))_{i=0}^{\infty}$ converges and its limit point exists (the closed interval $[0, D(\mathbf{v}_0, \pi_W(\mathbf{v}_0))]$ is compact in $\mathbb{R}$ equipped with the standard metric). We will denote this limit information divergence λ.

Furthermore, due to the compactness of V and W the sequences $\{\mathbf{v}_i\}_{i=0}^{\infty}$ and $\{\pi_W(\mathbf{v}_i)\}_{i=0}^{\infty}$ have both a convergent subsequence with a corresponding limit point, we denote these limits points $\mathbf{v} \in V$ and $\mathbf{w} \in W$ respectively. Therefore, by Observation 5,

$$D(\mathbf{v}, \mathbf{w}) = \lambda.$$

What we need to prove at this stage is that the whole sequence $\{\mathbf{v}_i\}_{i=0}^{\infty}$, not just its subsequence, converges to $\mathbf{v}$. We will do this considering Figure 14.

By the four–point property

$$D(\mathbf{v}_i, \mathbf{w}) \leq D(\pi_W(\mathbf{v}_{i-1}), \mathbf{w}) + D(\mathbf{v}, \mathbf{w}).$$

and by the Pythagorean property for admissible points

$$D(\pi_W(\mathbf{v}_i), \mathbf{w}) + D(\mathbf{v}_i, \pi_W(\mathbf{v}_i)) \leq D(\mathbf{v}_i, \mathbf{w}).$$

Since

$$D(\mathbf{v}_i, \pi_W(\mathbf{v}_i)) \geq D(\mathbf{v}, \mathbf{w})$$

it follows that

$$D(\pi_W(\mathbf{v}_i), \mathbf{w}) \leq D(\pi_W(\mathbf{v}_{i-1}), \mathbf{w}).$$

However, we already know that a subsequence of $\{\pi_W(\mathbf{v}_i)\}_{i=0}^{\infty}$ converges to $\mathbf{w}$, so this means that $\{D(\pi_W(\mathbf{v}_i), \mathbf{w})\}_{i=0}^{\infty}$ converges, by Observation 5, to $D(\mathbf{w}, \mathbf{w})$, which is by the consistency property 0. Finally, using the connectivity property, the whole sequence $\{\pi_W(\mathbf{v}_i)\}_{i=0}^{\infty}$ must converge to $\mathbf{w}$.

By the convexity property $D(\pi_W(\mathbf{v}_i), \mathbf{w}) \geq D(\widehat{\pi}_V(\pi_W(\mathbf{v}_i)), \mathbf{v})$ for all i so also

$$\{D(\mathbf{v}_i, \mathbf{v})\}_{i=1}^{\infty}$$

converges to zero which in turn means, making the same argument as above, that $\{\mathbf{v}_i\}_{i=0}^{\infty}$ converges to $\mathbf{v}$ as desired.

However, in order to apply the convexity property above, we need to first establish that $\widehat{\pi}_V(\mathbf{w}) = \mathbf{v}$. For a contradiction let us assume that $\mathbf{v} \neq \widehat{\pi}_V(\mathbf{w})$. By the Pythagorean property for agreeable points

$$D(\widehat{\pi}_V(\mathbf{w}), \widehat{\pi}_V(\mathbf{w}_i)) + D(\widehat{\pi}_V(\mathbf{w}_i), \mathbf{w}_i) = D(\widehat{\pi}_V(\mathbf{w}), \mathbf{w}_i),$$

for all i. Since $\{\mathbf{w}_i\}_{i=1}^{\infty}$ converges to $\mathbf{w}$, and $\{\mathbf{v}_i\}_{i=1}^{\infty}$ has a subsequence converging to $\mathbf{v}$ (so we focus only on it), and by Observation 5, we can also write

$$D(\widehat{\pi}_V(\mathbf{w}), \mathbf{v}) + D(\mathbf{v}, \mathbf{w}) = D(\widehat{\pi}_V(\mathbf{w}), \mathbf{w}).$$

By the assumption and the consistency property $D(\widehat{\pi}_V(\mathbf{w}), \mathbf{v}) > 0$, so we have that $D(\mathbf{v}, \mathbf{w}) < D(\widehat{\pi}_V(\mathbf{w}), \mathbf{w})$. But this is not possible, a contradiction.

Similarly we can establish a contradiction with the uniqueness of the D–projection should $\mathbf{w} \neq \pi_W(\mathbf{v})$ utilising the Pythagorean property for admissible points. Therefore $\mathbf{v} = \widehat{\pi}_V(\pi_W(\mathbf{v}))$ and $\mathbf{v}$ is a fixed point. □

Finally, considering Theorem 1 we may claim that the fixed point from the theorem above is also a representative point.

The algorithm (and in fact the idea of the proof presented above) is due to Csiszár and Tusnády [8], who developed it for a particular information divergence and it is known as an *alternating minimisation procedure*. The algorithm was then generalised many times in the literature, see e.g. [6], and the version above can be considered as a further step. Nevertheless, it is still the same idea developed in 1984.

8 Conclusion

We have now achieved the goal as initially stated; we have introduced information geometry without actually specifying the exact nature of admissible and agreeable points we worked with. However, the paper is far from finished. First, in the appendix the classical setting will be formally established and staples of inductive logic; discrete probability distributions and cross–entropy, will be discussed.

Second, as this aspired to be an actual generalisation of information geometry, the future development should be aimed to find a non–trivial and different formalisation of the intuitive concept than the one from the appendix, which is one usually used in inductive logic. This is exciting as one could hope to recover information geometry on mathematical objects originally meant to capture something else, leading to entirely new connections and applications.

Finally, we should also admit that the results presented here were somewhat easy; we placed in enough properties so that the proofs of the desired results went through. The hard job is the opposite: What properties are necessary? This question remains open for now.

References

1. Adamčík, M.: Collective Reasoning under Uncertainty and Inconsistency. Phd thesis, University of Manchester (2014)
2. Adamčík, M.: The information geometry of Bregman divergences and some applications in multi–expert reasoning. Entropy **16**, 6338–6381 (2014)
3. Adamčík, M.: On the applicability of the 'number of possible states' argument in multi–expert reasoning. Journal of Applied Logic **19**, 20–49 (2016)
4. Adamčík, M.: A logician's approach to meta–analysis with unexplained heterogeneity. Journal of Biomedical Informatics **71**, 110–129 (2017)
5. Adamčík, M.: A note on how Rényi entropy can create a spectrum of probabilistic merging operators. Kybernetika **55**, 605–617 (2019)
6. Bauschke, H.H., Combettes, P.L., Noll, D.: Joint minimization with alternating Bregman proximity operators. Pacific Journal of Optimization **2**, 401–524 (2006)
7. Bregman, L.M.: The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. Math. Phys. **1**, 200–217 (1967)
8. Csiszár, I., Tusnády, G.: Informational geometry and alternating minimization procedures. Statistic and Decisions **1**, 205–237 (1984)
9. Genest, C., Wagner, C.G.: Further evidence against independence preservation in expert judgement synthesis. Aequationes Mathematicae **32**, 74–86 (1987)
10. Jaynes, E.T.: Where do we stand on maximum entropy? In: Levine, R.D., Tribus, M. (eds.) The Maximum Entropy Formalism. pp. 15–118. M.I.T. Press (1979)
11. Joyce, D.E.: Online eddition of Euclid's Elements. http://aleph0.clarku.edu/ djoyce/java/elements (1998), [Online; accessed 12–December–2016]
12. Kahn, D.W.: Topology. Dover Publications, New York (1995)
13. Paris, J.B.: The uncertain reasoner companion. Cambridge University Press, Cambridge (1994)
14. Williamson, J.: Deliberation, judgement and the nature of evidence. Economics and Philosophy **31**, 27–65 (2015)
15. Wilmers, G.M.: A foundational approach to generalising the maximum entropy inference process to the multi–agent context. Entropy **17**, 594–645 (2015)

Appendix

"One could not see the forest for the trees."

A Common Proverb

In this paper we have accumulated a large number of properties that we require from an information divergence D and from the sets of agreeable and admissible points. Naturally we should ask the following question: Is it actually possible to satisfy them all? In this section we show particular examples that satisfy all the properties, but we will need some additional notions to define them.

Obdurate Committee

> "The point is that we are not ignoring the dynamics, and we are not getting something from nothing, (...) for these all circumstances that are not under the experimenter's control must, of necessity, be irrelevant. (...) Solution by the Maximum Entropy Principle is so unbelievably simple just because it eliminates those irrelevant details right at the beginning of the calculation by averaging over them."

Edwin T. Jaynes, [10]

Here we consider an obdurate committee who stubbornly refuses to iterate the process $\mathbf{v}_1 = \widehat{\pi}_V(\pi_W(\mathbf{v}_0))$. This will help us to further illustrate the setting, toy with it, but foremost illustrate some singular points of information geometry.

First, we postulate existence of the *most uninformative point* $\mathbf{u}$ in the set of agreeable points V. Second, the committee finds $\pi_W(\mathbf{u})$, a unique agreeable point that has the smallest D–divergence from $\mathbf{u}$. If we wanted to represent W by a single point, this is the most natural option as, in respect to D, it has the least added 'information' to it among the agreeable points.

This generalises the concept of the famous *most entropic point*; we recover the usual concept if we choose a specific information divergence D and a specific concept of the point. We will elaborate the details later in this appendix. We only mention that the committee is not ignoring the dynamics of the set of admissible points W by selecting that single point there as well as the experimenter is not doing so in the citation above. If the dynamics were laboriously worked out, we would have obtained this solution anyway.

Let us denote $\pi_W(\mathbf{u})$ by $\mathbf{ME}_D(W)$, and call it the most entropic point in W (in respect to D). Now, the committee wishes to find the agreeable point (if it is not already and agreeable admissible point). To that end, $\widehat{\pi}_V(\mathbf{ME}_D(W))$ is picked, and we denote $O(W) = \{\widehat{\pi}_V(\mathbf{ME}_D(W))\}$ the singleton containing it. An illustration is in Figure 6.

This *obdurate point* need not be a fixed point; and even less a representative point, considering Observation 2. It would indeed be a stubborn committee not to iterate the process further, but be content with it. The committee would argue that the advantage of O is that it contains a single point. We would point out that $O(W) \neq W \cap V = \Delta(W)$, if $W \cap V \neq \emptyset$ and $W \cap V$ has at least two elements (given D satisfies the consistency property), as shown in Observation 1. Nevertheless, starting the whole iteration process from the most uninformative point appears a well justified idea that indeed lead to a unique point as investigated in Section 7.

Finally, let us point out the following obvious statements.

Observation 7. *If W is a singleton, then*

$$O(W) = \Delta(W).$$

Observation 8. *If $W \subseteq V$, then*

$$O(W) \subseteq \Delta(W).$$

The prior follows from Property 3, while the latter from Observation 1.

Points

To provide an actual example of what was discussed in the paper, we start with the J–dimensional *Euclidean space*, which is a set of all ordered J–tuples

$$\mathbf{v} = (v_1, \ldots, v_J),$$

where every v_j is a real number. In other words, $\mathbf{v} \in \mathbb{R}^J$. A $(J-1)$–dimensional *probabilistic simplex* $\mathbb{D}^J$, $J \geq 2$, is a subspace of the J–dimensional Euclidean space defined as those $\mathbf{v} \in \mathbb{R}^J$ that satisfy

$$\sum_{j=1}^{J} v_j = 1.$$

We will confine ourselves to the case when $v_j > 0$, for all $1 \leq j \leq J$, to avoid any pathological cases, which makes $\mathbb{D}^J$ an open set. Such a defined *discrete probability distribution* $\mathbf{v}$ could perhaps represent a probabilistic opinion that an individual may have about the world, and thus it plays a central role in inductive logic and uncertain reasoning [13, 15]. More recently, in [4] they have been used to represent results of individual medical studies.

We say that a subset W of points in $\mathbb{R}^I$ is *convex* if for any two $\mathbf{v}, \mathbf{w} \in W$ we have that also

$$(\lambda \cdot v_1 + (1-\lambda) \cdot w_1, \ldots, \lambda \cdot v_I + (1-\lambda) \cdot w_I) \in W,$$

for all $\lambda \in [0,1]$. We say that a subset W of points in $\mathbb{R}^I$ is *closed* if the limit point of every convergent sequence constructed from the elements of W has its limit inside W, in respect to the standard Euclidean metric.

Now, let us consider a closed convex set of points

$$W \subseteq \underbrace{\mathbb{D}^J \times \ldots \times \mathbb{D}^J}_{n}.$$

Note that $I = Jn$ (in the definition of convexity above) and $\mathbf{w} \in W$ is of the form $\mathbf{w} = (\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(n)})$, where each $\mathbf{v}^{(i)} \in \mathbb{D}^J$ is a probability distribution admissible by the member i of a group of n individuals. This set W will be an example of a set of admissible points discussed earlier in the paper.

Finally, let

$$V \subseteq \underbrace{\mathbb{D}^J \times \ldots \times \mathbb{D}^J}_{n}$$

be such that in each $\mathbf{v} \in V$ all members are in agreement; $\mathbf{v} = (\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(n)})$, where $\mathbf{v}^{(1)} = \ldots = \mathbf{v}^{(n)}$. This set V is not closed (because $\mathbb{D}^J$ is not), but we can fix a sufficiently small $\epsilon > 0$ and ask every $v_j^{(i)} > \epsilon$, $1 \leq i \leq n$, $1 \leq j \leq J$. A suitable ϵ exists (in a sense that $W \subseteq V$ must be possible), since W is assumed closed. Such a set V will be an example of a set of agreeable points discussed earlier in the paper.

Clearly, it could be that there are some agreeable points in W, but V and W could be as well disjoint. In any case, W is assumed non–empty, while V is non–empty by definition. Both W and V are defined closed and bounded, and hence they are both compact. Note that compactness was required in Observation 6 and Theorem 2.

Divergences

After we have introduced the points, let us now define a divergence from one point to another. In [5], the following divergence from $\mathbf{v} \in V$ to $\mathbf{w} \in W$ based on the Rényi entropy was defined:

$$D_r(\mathbf{v}, \mathbf{w}) = \frac{1}{n} \sum_{j=1}^{Jn} [(w_j)^r - (v_j)^r - r(w_j - v_j)(v_j)^{r-1}],$$

where $2 \geq r > 1$. For $r = 2$ this divergence becomes the well known *squared Euclidean distance*

$$E(\mathbf{v}, \mathbf{w}) = \frac{1}{n} \sum_{j=1}^{Jn} (v_j - w_j)^2,$$

exceptionally a symmetric divergence. The proof that the set of representative points $\Delta^{D_r}(W)$ based on the Rényi entropy is well defined is in [1].

Another way to define the divergence D from $\mathbf{v} \in V$ to $\mathbf{w} \in W$ is to take the *Kullback–Leibler divergence* (also known as cross–entropy)

$$\mathrm{KL}(\mathbf{v}, \mathbf{w}) = \frac{1}{n} \sum_{j=1}^{Jn} w_j^{(i)} \log \frac{w_j^{(i)}}{v_j}.$$

A limit theorem relating the set of representative points $\Delta^{D_r}(W)$ based on the Rényi entropy to the set of representative points $\Delta^{\mathrm{KL}}(W)$ based on the Kullback–Leibler divergence has been proven in [5];

$$\emptyset \neq \lim_{r \searrow 1} \Delta^{D_r}(W) \subseteq \Delta^{\mathrm{KL}}(W).$$

Whether or not the above holds with equality is an open problem.

The proofs that the divergences defined above satisfy all Properties 1 to 9 discussed in this paper are scattered in [1] and [2], and they are all special cases of a general convex Bregmann divergence [7].

What we discussed in this paper now gives us

$$\Delta^{D_r}(W) = \Theta^{D_r}(W), \text{ and } \Delta^{\mathrm{KL}}(W) = \Theta^{\mathrm{KL}}(W),$$

the representative and fixed points are the same points, and we can get a representative point by iterating projections and conjugated projections.

Discussion

The technical nature of this appendix obscured how natural and simple these examples actually are. We will mention some singular points in what follows.

Regardless on which of the above mentioned divergences is taken for D, the conjugated D–projection of an admissible point $\mathbf{w} = (\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}) \in W$ to the set of agreeable points $(\underbrace{\mathbf{v}, \ldots, \mathbf{v}}_{n}) \in V$ in fact gives

$$\mathbf{v} = \Big(\frac{1}{n}\sum_{i=1}^{n} w_1^{(i)}, \ldots, \frac{1}{n}\sum_{i=1}^{n} w_J^{(i)}\Big).$$

So we have here only the ordinary arithmetic mean applied to J coordinates respectively. In the literature this operator is known as the *linear pooling operator*, see [9]. It is a common choice of representing different opinions $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)} \in \mathbb{D}^J$ of n individuals as a single point in $\mathbb{D}^J$, a natural agreeable point.

Defining the most uninformative point $\mathbf{u} = (\underbrace{\mathbf{v}, \ldots, \mathbf{v}}_{n}) \in V$ using the *uniform probability distribution*

$$\mathbf{v} = \Big(\frac{1}{J}, \ldots, \frac{1}{J}\Big) \in \mathbb{D}^J,$$

$\mathbf{ME}_{\mathrm{KL}}(W)$, defined as the KL–projection of $\mathbf{u}$ into W, is the usual most entropic point in W. It is defined as that $\mathbf{w}$ that maximises the Shannon entropy

$$-\sum_{j=1}^{Jn} w_j \log w_j.$$

An obdurate committee would then take this most entropic point and find the conjugated KL–projection in the set of agreeable points V, which we now know to be equivalent to applying a linear pooling operator, and be content with it.

We suggest that a rational committee would iterate the whole process endlessly until a representative point in $\varDelta^{\mathrm{KL}}(W) = \varTheta^{\mathrm{KL}}(W) \subseteq V$ is reached. A combinatorial argument in favour of using $\varDelta^{\mathrm{KL}}(W)$ in a specific context was presented in [3].

Should there be only one individual, $n = 1$, then $W = \varDelta^{\mathrm{KL}}(W) = \varTheta^{\mathrm{KL}}(W) \subseteq V$, so there would be no need to iterate the process as $\mathbf{ME}_{\mathrm{KL}}(W)$ would be trivially, see Observation 8, a fixed point. This would correspond to the classical most entropic solution when there are no conflicting sources of information.

Figures

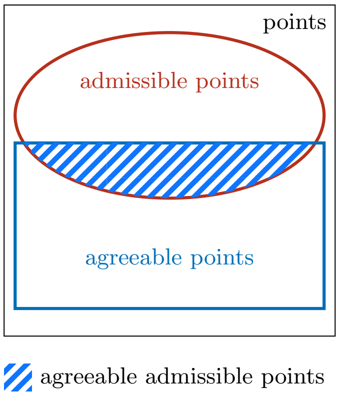


Fig. 1. An illustration of the set of all points.

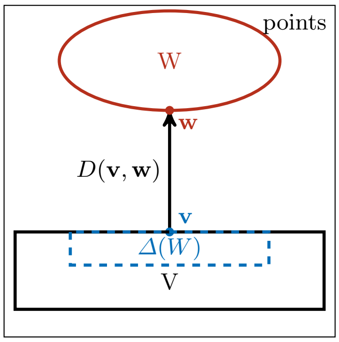


Fig. 2. An illustration of the representative points.

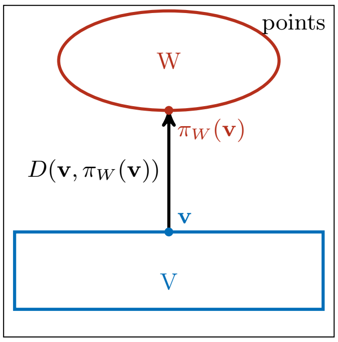


Fig. 3. An illustration of the D–projection.

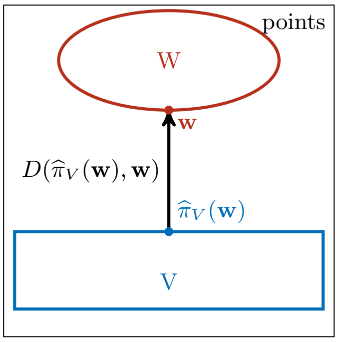


Fig. 4. An illustration of the conjugated D–projection.

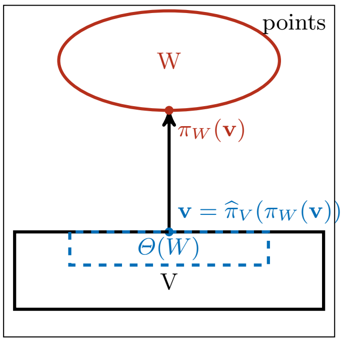


Fig. 5. An illustration of the fixed points.

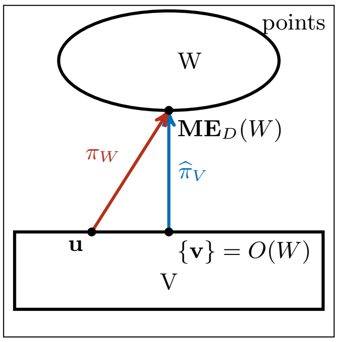
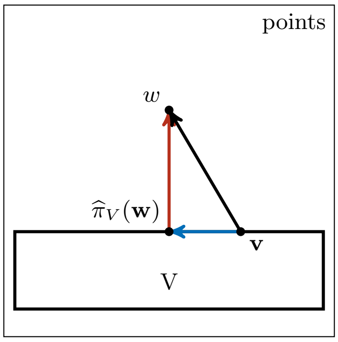
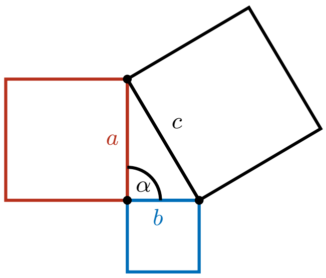


Fig. 6. An illustration of an obdurate committee.



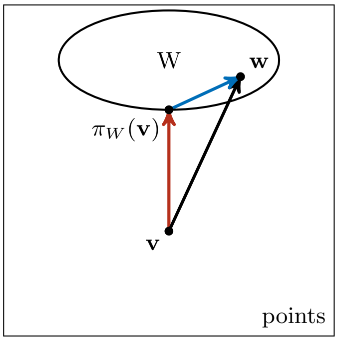
$$D(\mathbf{v}, \widehat{\pi}_V(\mathbf{w})) + D(\widehat{\pi}_V(\mathbf{w}), \mathbf{w}) = D(\mathbf{v}, \mathbf{w})$$

Fig. 7. An illustration of the Pythagorean property for agreeable points.



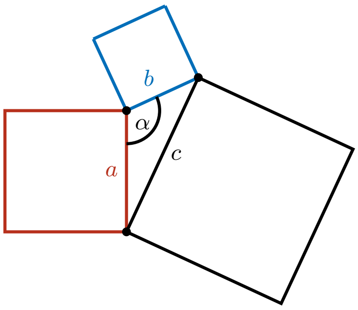
if $\alpha = 90°$ then $a^2 + b^2 = c^2$

Fig. 8. How squares behave in the Euclidean geometry.



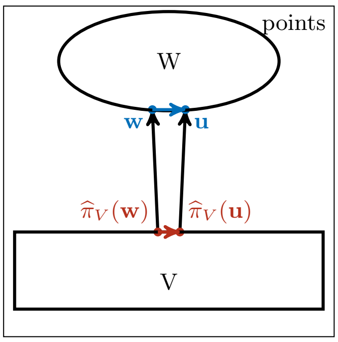
$$D(\mathbf{v}, \pi_W(\mathbf{v})) + D(\pi_W(\mathbf{v}), \mathbf{w}) \leq D(\mathbf{v}, \mathbf{w})$$

Fig. 9. An illustration of the Pythagorean property for admissible points.



if $90^\circ \leq \alpha \leq 180^\circ$ then $a^2 + b^2 \leq c^2$

Fig. 10. How squares behave in the Euclidean geometry.



$$D(\mathbf{w}, \mathbf{u}) \geq D(\widehat{\pi}_V(\mathbf{w}), \widehat{\pi}_V(\mathbf{u}))$$

Fig. 11. An illustration of the convexity property.

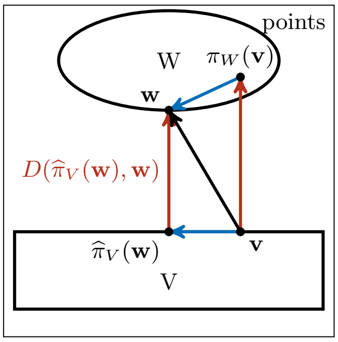
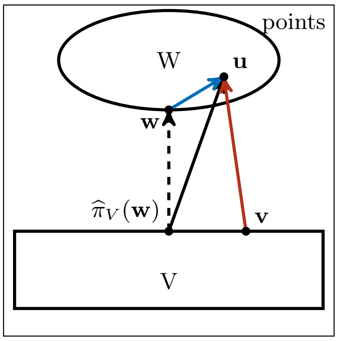


Fig. 12. An illustration of the proof for Theorem 1.



$D(\widehat{\pi}_V(\mathbf{w}), \mathbf{u}) \leq D(\mathbf{w}, \mathbf{u}) + D(\mathbf{v}, \mathbf{u})$

Fig. 13. An illustration of the four points property.

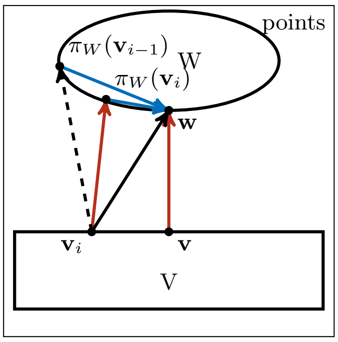


Fig. 14. An illustration of the proof of Theorem 2.

Inductive reasoning, conditionals, and belief revision

Gabriele Kern-Isberner[0000-0001-8689-5391]

Dept. of Computer Science, TU Dortmund, 44221 Dortmund, Germany

Abstract. This paper presents a broad view on inductive reasoning by embedding it in theories of epistemic states, conditionals, and belief revision. More precisely, we consider inductive reasoning as a specific case of belief revision on epistemic states where three-valued conditionals are a basic means for representing beliefs. We present a general framework for inductive reasoning from conditional belief bases that also allows for taking background beliefs into account, and illustrate this by probabilistic reasoning based on optimum entropy.

Keywords: inductive reasoning · conditionals · belief revision · reasoning at optimum entropy

1 Introduction

In its original sense, inductive reasoning means deriving generic knowledge from given examples in a way that completes the example-based information concisely to make it applicable to other situations. In this paper, we take a bit broader view on inductive reasoning: we pursue the idea that inductive reasoning should be able to "generate" new beliefs from given beliefs and ideally, complete the beliefs of a human being as far as possible. This is a very common and basic problem in the area of knowledge representation in artificial intelligence. Here, it is usually assumed that knowledge and beliefs of a human being, or an agent, respectively, can be represented by a knowledge base, i.e., a finite set of formulas in a suitable logic, and that more knowledge and beliefs can be inferred from this base. In artificial intelligence, the distinction between knowledge and beliefs is vague, because its main goal is to model knowledge and behaviour of agents, so knowledge often means subjective knowledge, which is very close to beliefs. We do not want to enter this fundamental discussion here but will use the term beliefs throughout this paper as a synonym for subjective knowledge.

So, inductive reasoning should be able to extend the beliefs of a belief base in a non-trivial, principled way. Of course, the logic framework in which beliefs are represented plays a crucial role here. In the simple case of propositional logic, deduction, or more generally, a Tarski consequence operator would satisfy the general requirements of an inductive reasoning operator, and similarly for first-order predicate logic. Beyond classical logics, non-monotonic logics using so-called default rules, or rules with exceptions, provide more powerful inference operators, prominent approaches here are Reiter's default logic and answer set programming. Both are symbolic and able to infer formulas from belief bases of facts and rules. In quantitative logical settings, probability theory offers a rich semantic framework for nonmonotonic reasoning, and the principle of maximum entropy (*MaxEnt principle*) [7, 13] yields a most powerful inductive inference operator from probabilistic belief bases. There are also popular approaches using qualitative structures like (total) preorders, or semi-quantative methodologies based on Spohn's ordinal conditional functions, also called ranking functions [17], like system Z [6] that allow for reasoning from conditional belief bases.

This paper aims at describing inductive reasoning in a broad context where we elaborate on connections to conditionals and belief change theory, and where we are able to distinguish clearly between background, or generic, beliefs and evidential, or contextual, information, a feature that is listed in [3] as one of three basic requirements a *plausible exception-tolerant inference system* has to meet. We build upon previous works, in particular [9, 11], and elaborate a general vision of inductive reasoning in the context of belief revision. While is has been well known that nonmonotonic reasoning and belief revision are "two sides of the same coin" [5], the focus here is on inductive reasoning as a concept that merges techniques from both areas to bring forth a methodology in which reasoning and revision can interact in various ways to realise inductive reasoning from different background beliefs and under different contextual information. A core concept in this methodology are epistemic states which are equipped with meta-structures supporting reasoning and revision, and beliefs are expressed by conditionals in the first place. Note that, of course, also propositional beliefs are covered in our approach by identifying a conditional $(A|\top)$, where $\top$ is a tautology, with the plausible belief A. Interestingly, total preorders on possible worlds are meta-structures that provide a solid foundation for reasoning, revision, and conditionals, and indeed, they are a basic requirement for AGM revision [8]. So, we build upon AGM revision but go far beyond by addressing iterative revision and conditional revision.

The outline of the paper is as follows: We recall basic definitions and notations in Section 2 and discuss the nature of epistemic states in Section 3, also pointing out their connections to inductive reasoning, conditionals, and belief revision. We also exemplify inductive reasoning and belief revision in probabilistics via the principles of optimum entropy. Finally, we compare inductive reasoning/revision to applying beliefs to specific situations, which we call focusing in Section 4, and conclude in Section 5.

2 Basics and notations

The propositional language $\mathcal{L}$ with formulas A, B is defined in the usual way by virtue of a finite signature Σ with atoms $a, b, \ldots$ and junctors $\wedge, \vee$, and $\neg$ for conjunction, disjunction, and negation, respectively. The $\wedge$-junctor is mostly omitted, so that AB stands for $A \wedge B$, and negation is usually indicated by overlining the corresponding proposition, i.e. $\overline{A}$ means $\neg A$. Literals are positive or negated atoms. The set of all propositional interpretations over Σ is denoted by Ω_Σ. As the signature will be fixed throughout the paper, we will usually omit the subscript and simply write Ω. Possible worlds are understood as a synonym for interpretations, and are usually represented by a complete conjunction of the corresponding literals, i.e., a conjunction mentioning all atoms of the signature such that exactly those atoms are negated that are evaluated to *false*. Also the satisfaction relation $\models$ between worlds and formulas is defined in the usual way: $\omega \models A$ iff ω evaluates A to *true*. In this case, we say ω is a model of A. The set of all models of A is denoted by $Mod(A)$. Then, $A \models B$ for two formulas $A, B \in \mathcal{L}$ if $Mod(A) \subseteq Mod(B)$.

$\mathcal{L}$ is extended to a conditional language $(\mathcal{L} \mid \mathcal{L})$ by introducing a conditional operator $|$: $(\mathcal{L} \mid \mathcal{L}) = \{(B|A) \mid A, B \in \mathcal{L}\}$. $(\mathcal{L} \mid \mathcal{L})$ is a flat conditional language, no nesting of conditionals is allowed. Conditionals $(B|A)$ with *antecedent* (or *premise*) A and *consequent* B are basically considered as three-valued entities in the sense of de Finetti [2] which can be verified ($\omega \models AB$), falsified ($\omega \models A\overline{B}$), or simply not applicable ($\omega \models \overline{A}$) in a possible world ω. So, they have to be interpreted within richer semantic structures such as *epistemic states* like probability distributions, or ranking functions [17]. In this paper, we choose both of these semantic frameworks to exemplify our approach.

Probability distributions in a logical environment can be identified with probability functions $P : \Omega \to [0,1]$ with $\sum_{\omega\in\Omega} P(\omega) = 1$. The probability of a formula $A \in \mathcal{L}$ is given by $P(A) = \sum_{\omega\models A} P(\omega)$. Since $\mathcal{L}$ is finite, Ω is finite, too, and we only need additivity instead of σ-additivity. Conditionals are interpreted via conditional probabilities, so that $P(B|A) = \frac{P(AB)}{P(A)}$ for $P(A) > 0$, and $P \models (B|A)\,[x]$ iff $P(A) > 0$ and $P(B|A) = x$ ($x \in [0,1]$).

Ordinal conditional functions (OCFs), (also called *ranking functions*) $\kappa : \Omega \to \mathbb{N} \cup \{\infty\}$ with $\kappa^{-1}(0) \neq \emptyset$, were introduced first by Spohn [17]. They express degrees of plausibility of propositional formulas A by specifying degrees of disbeliefs of their negations $\overline{A}$. More formally, we have $\kappa(A) := \min\{\kappa(\omega) \mid \omega \models A\}$, so that $\kappa(A \vee B) = \min\{\kappa(A), \kappa(B)\}$. A conditional $(B|A)$ is accepted in the epistemic state represented by κ, written as $\kappa \models (B|A)$, iff $\kappa(AB) < \kappa(A\overline{B})$, i.e. iff AB is more plausible than $A\overline{B}$.

In general, let Ψ be any epistemic state, specified by some structure that is found appropriate to express conditional beliefs from a suitable conditional language $(\mathcal{L} \mid \mathcal{L})^*$, in which conditionals may be equipped with quantitative degrees of belief, according to the chosen framework. For instance, for probability functions, $(\mathcal{L} \mid \mathcal{L})^* = (\mathcal{L} \mid \mathcal{L})^{prob} = \{(B|A)[x] \mid A, B \in \mathcal{L}, x \in [0,1]\}$, and in qualitative environments, $(\mathcal{L} \mid \mathcal{L})^* = (\mathcal{L} \mid \mathcal{L})$. Moreover, an entailment relation $\models$ is given between epistemic states and conditionals; basically, $\Psi \models (B|A)^*$ means that $(B|A)^*$ is accepted in Ψ, where acceptance is defined suitably. Let $\mathcal{E}^* = \mathcal{E}^*_\Sigma$ denote the set of all such epistemic states using $(\mathcal{L} \mid \mathcal{L})^*$ for representation of (conditional) beliefs. Moreover, epistemic states are considered as (epistemic) models of sets of conditionals $\Delta \subseteq (\mathcal{L} \mid \mathcal{L})^*$: $Mod^*(\Delta) = \{\Psi \in \mathcal{E}^* \mid \Psi \models \Delta\}$. As usual, $\Delta \subseteq (\mathcal{L} \mid \mathcal{L})^*$ is *consistent* iff $Mod^*(\Delta) \neq \emptyset$, i.e. iff there is an epistemic state which is a model of Δ.

3 Inductive reasoning based on epistemic states and belief revision

In this section, we develop our general approach to inductive reasoning as a special case of epistemic belief revision. Epistemic states serve as a mediator between reasoning and revision by providing both an epistemic background for reasoning and an ideal outcome of induction from and revision by (conditional) belief bases. First, we discuss the semantic structures of epistemic states that are required for this purpose; in particular, we emphasize the crucial role of conditionals in this context. Then we present the technical realisations of our approach on an abstract level. Finally, we elaborate on the different types of belief and information that our approach can handle.

3.1 Epistemic states and conditionals

In this paper, in the context of inductive reasoning and belief revision, we take a pragmatic view on epistemic states. We expect (the representation of) epistemic states to be equipped with some meta-structures which in suitable logical frameworks allow for performing reasoning and belief revision, and to be complete in the sense that answers to all possible queries (in the respective) framework can be generated, to the best of the human's beliefs. Note that we use the term "revision" here in a general sense, as a synonym for integrating new information to one's current beliefs, i.e., as a super-concept also including update [8] or focusing [3]. When the specific change operator called revision in the AGM theory is meant, we speak of "AGM revision", or specify this explicitly.

As a crucial feature to go beyond classical logic towards modelling of human's beliefs, we presuppose that epistemic states can evaluate conditionals to be accepted or not accepted. We avoid saying that a conditional is *true* or not in an epistemic state because, on the one hand, conditionals are not binary but three-valued, and, on the other hand, the understanding of conditionals in commonsense reasoning is not truth-functional at all. To accept a conditional, humans would expect a meaningful connection between antecedent and consequent. This is crucial for our approach to inductive reasoning because this connection can be used for reasoning in a way that captures human-like thinking. The basic idea is simple: A conditional $(B|A)$ is accepted if its verification AB is deemed to be more plausible, or probable, than its falsification $A\overline{B}$. The inherent connection between antecedent and consequent is taken into regard by considering A and B resp. A and $\overline{B}$ jointly when assessing plausibility, or probability. Beyond plain comparison, also degrees of plausibility, or probability, can be assigned to verification and falsification so as to measure the strength of a conditional, if the respective semantic framework allows for that.

The fundamental connection between epistemic states, conditionals, plausibility, (inductive) reasoning, and belief revision on which this paper relies can be roughly expressed by the following equivalences:

$$\Psi \models (B|A) \quad \text{iff} \quad AB \prec_\Psi A\overline{B} \quad \text{iff} \quad A \mid\!\sim_\Psi B \quad \text{iff} \quad \Psi * A \models B, \tag{1}$$

where Ψ is an epistemic state in $\mathcal{E}^*$, $\preceq_\Psi$ is a suitable relation expressing plausibility (or probability)[1], $\mid\!\sim_\Psi$ is an inference relation based on Ψ, and $*$ is an epistemic (or iterative) revision operator that takes an epistemic state and a proposition and returns again an epistemic state (in the sense of [1]). More generally, we assume that $*$ can also deal with much more complex beliefs given by sets of conditionals Δ such that $\Psi * \Delta \in \mathcal{E}^*$, and we also adopt the success postulate of AGM theory here, i.e., we presuppose that $\Psi * \Delta \models \Delta$. This also includes the case of revising by a (plausible) proposition A via identifying A with $(A|\top)$. Equation (1) reveals that both epistemic states and conditionals are also carriers of strategic information that become effective for reasoning and revision. Our focus here is on the inference relation $\mid\!\sim_\Psi$, and basing it on an epistemic state Ψ helps clarifying formally what is understood by induction. Before we go into more details here, we need to make explicit more clearly what we expect from the meta-structures associated with an epistemic state.

Indeed, a purely qualitative preorder might be a suitable meta-structure that is associated with an epistemic state. Of course, there are more sophisticated representation frameworks, such as possibility theory, ranking functions, and probability functions. But also modal logical frameworks seem to be good candidates for representing epistemic states, or heterogeneous structures consisting of different components (with reasonable interactions between them) might prove useful. This is not necessarily a question of numerical or symbolic representation, both types of frameworks can be fine.

But when it comes to numbers it should be clear that the crucial point here is not that they may provide a richer semantics, but they definitely provide richer structures that calculations for information processing might follow. And this makes them quite distinguished candidates for epistemic states in the context of reasoning and belief change. It is not by accident that probability theory with its two independent arithmetic operators (addition and multiplication) has been playing a major role here. Although AGM might have marked the beginning of symbolic belief revision and of devising rational postulates for belief change, performing practical belief change has been done for a much longer time in the probabilistic framework. Presumably the first belief change operator ever is probabilistic conditioning, and Jeffrey's rule [14] shows a possible way of incorporating even uncertain evidence. So, it is not for the numbers that we should care about probability theory but for the rich arithmetic structure that provides a powerful apparatus to express and process information (cf. also [14]). Via the multiplication operator, (conditional) independencies (and hence monotonic inference behaviour) can be expressed, and its inverse operator, division, allows to easily transform one distribution into another at the occurrence of new information via conditioning. Furthermore, the addition operator takes care of disjunctive propositional information, e.g., to allow for reasoning by cases. Having once adopted such basic techniques, information processing becomes easy.

[1] Note that $A \prec_\Psi B$ iff $A \preceq_\Psi B$ and not $B \preceq_\Psi A$

3.2 Inductive reasoning and belief revision

If we understand inductive reasoning as completing partial beliefs (as specified in a belief base Δ) as best as possible, then its result should be an epistemic state Ψ_Δ:

$$\Psi_\Delta = ind(\Delta), \tag{2}$$

where *ind* is some inductive reasoning mechanism; we also say that Δ is *inductively represented* by Ψ via *ind*, or that Δ *inductively generates* Ψ. For instance, Δ may be a set of conditionals, and *ind* might be specified by system Z [6], or c-representations [10], associating to each consistent set of conditionals a ranking function [17]. Inductive reasoning from Δ is then implemented by reasoning from $\Psi = ind(\Delta)$ via the conditionals being accepted in Ψ. That is, *ind* realises *model-based inductive reasoning.*

But this cannot be the end of the story. The mind of a human being is always evolving and changing by learning, or receiving new information $\mathcal{I}$ in general, where $\mathcal{I}$ can just be a fact, more complex contextual information also including conditionals (e.g., when we enter a new country, different compliance rules apply), or even trigger some deeper learning processes. Starting a new inductive reasoning process each time when we receive new information would make our beliefs incoherent, $\Psi = ind(\Delta)$ and $\Psi' = ind(\mathcal{I})$ might be completely unrelated (except for that they have been built up by the same inductive reasoning formalism). Integrating new information $\mathcal{I}$ into existing beliefs represented by an epistemic state Ψ is exactly the task of (epistemic or iterated) belief revision [1], returning a new epistemic state Ψ' after revising Ψ by $\mathcal{I}$:

$$\Psi_\Delta * \mathcal{I} = ind(\Delta) * \mathcal{I} = \Psi' \tag{3}$$

Note that we use $*$ here in a generic sense as a placeholder for a suitable change operator. Regarding that $\Psi_\Delta = ind(\Delta)$ has been built up inductively from a belief base Δ, and that also $\mathcal{I}$ will also be only partial information on some current context usually, the following questions naturally arise immediately: How do *ind* and $*$ interact? What (maybe completely different) roles do Ψ_Δ, Δ and $\mathcal{I}$ play in this scenario?

We first discuss the second question by analysing different qualities of beliefs with respect to the roles they play in the reasoning process. Roughly, we can distinguish between background, or generic, and evidential, or contextual knowledge, as well as between explicit and implicit beliefs. From background or generic knowledge, the agent takes beliefs which hold in general and of which she can make use of in different situations. For instance, the current beliefs of an agent getting up on a usual Monday morning might be different from those on a usual Sunday, but presumably his generic background has not changed much. The evidential resp. contextual information $\mathcal{I}$ she receives might include that it is Monday and raining, and that due to new construction areas she has to take some detours when going to work. We prefer the attribute "contextual" to "evidential" in the following, since this information may relate not only to a specific situation and can be much more complex than some evidential facts. For instance, the temporal scope of context may be one hour or one week, the scope may refer to a specific house or to a whole country, or it may contain information on abstract contexts, such as holidays or working environments. Assuming that $\Psi_\Delta = ind(\Delta)$ expresses background beliefs, incorporating contextual information cannot be done simply via the "union" of Ψ_Δ and $\mathcal{I}$ (whatever this might be), or by the union of Δ and $\mathcal{I}$ because this would ignore the different natures of background beliefs and contextual information. The agent's new epistemic state should rather arise from the adaptation of Ψ_Δ to contextual information. This is expressed by (3), but only as a base case when we start reasoning from a belief base including our core background beliefs. However, this process must be iterative, i.e., $\Psi = \Psi_\Delta$ may more generally be the result of such a revision $\Psi = \Psi_{prior} * \mathcal{I}_{prior}$, or new information $\mathcal{I}'$ arrives that triggers a new change process $(\Psi_\Delta * \mathcal{I}) * \mathcal{I}'$, so that (3) evolves to the iterative change problem

$$(\Psi_\Delta * \mathcal{I}) * \mathcal{I}' = (ind(\Delta) * \mathcal{I}) * \mathcal{I}'. \tag{4}$$

And here, three essentially different reasoning resp. revision scenarios are possible (note that the $*$-operators are just placeholders to be specified adequately):

- First, the context to which $\mathcal{I}$ refers has evolved, and $\mathcal{I}'$ is information on this new context for which, however, $\mathcal{I}$ is still relevant. This scenario is often referred to as *updating*. Then the two $*$-operators in (4) would be of the same type, and $\Psi_\Delta * \mathcal{I}$ would be changed to $(\Psi_\Delta * \mathcal{I}) * \mathcal{I}'$. A modification of this scenario applies if the contexts to which $\mathcal{I}$ and $\mathcal{I}'$ refer are completely unrelated, but the agent uses the same background beliefs Ψ_Δ for reasoning, then we would end up with $\Psi_\Delta * \mathcal{I}'$.
- Second, $\mathcal{I}'$ refers to the same context as $\mathcal{I}$. In this case, $\mathcal{I}$ and $\mathcal{I}'$ should be considered to be on the same level, and we would obtain $\Psi_\Delta * (\mathcal{I} \cup \mathcal{I}')$. This is a typical case of *belief revision* in the AGM-sense.
- Third, $\mathcal{I}'$ enriches or modifies background beliefs, i.e., it affects the basis from which reasoning with the information $\mathcal{I}$ is performed. This is what happens when *learning*. In the first case, if $\mathcal{I}'$ is fully compatible with Δ, $ind(\Delta \cup \mathcal{I}') * \mathcal{I}$ would be a proper solution. If $\mathcal{I}'$ contradicts (parts of) Δ, then $\Psi_\Delta * \mathcal{I}' = ind(\Delta) * \mathcal{I}'$ would provide suitable background beliefs, and $(ind(\Delta) * \mathcal{I}') * \mathcal{I}$ would be the result of the revision problem.

Therefore, we argue that the distinction between revision and update [8], and also the relation between belief change and learning is not just a technical issue, but has to be made on a conceptual and modelling level. The involved revision operators $*$ might respect such differences, but from the discussion above it becomes clear that also differences can be made by different ways of applying one and the same revision operator $*$ in different scenarios, also involving inductive reasoning. While (3) claims that involving belief revision is necessary for a coherent perspective of inductive reasoning, the third of the cases elaborated above shows how inductive reasoning can affect belief revision: Changing $ind(\Delta)$ to $ind(\Delta \cup \mathcal{I}')$ makes the revision of background beliefs possible. For more formal investigations of the differences between AGM-like revision and update, and for a reconciliation with AGM theory, please see [11].

Elaborating further on this intimate connection between inductive reasoning and belief revision, we might even envisage inductive reasoning involving background beliefs expressed by an epistemic state Ψ_{bk}, i.e., $\Psi = ind_{\Psi_{bk}}(\Delta)$, and then inductive reasoning from Δ might be realised by revision:

$$\Psi = ind_{\Psi_{bk}}(\Delta) = \Psi_{bk} * \Delta. \tag{5}$$

And when no background beliefs are available or relevant, we assume some uniform epistemic state Ψ_u as a starting point:

$$ind = ind_{\Psi_u}. \tag{6}$$

This implements inductive reasoning from epistemic states thoroughly via epistemic belief revision because this approach yields

$$\Psi_\Delta = ind(\Delta) = \Psi_u * \Delta. \tag{7}$$

This means that each epistemic revision operator that is able to handle complex information Δ induces an inductive inference operator. This makes inductive reasoning coherent, as explained above, and allows us to embed inductive reasoning in a richer methodology.

This embedding has two further important advantages: First, revision methodologies may yield immediately mechanisms of inductive reasoning and suitable quality criteria. Second, splitting up inductive reasoning clearly into its inductive mechanism, its involved background beliefs, and context-based beliefs makes formalisms more explicit and more broadly (and flexibly) applicable. However, only very few approaches to epistemic revision with sets of conditionals exist; in Section 3.4, we briefly present the principle of minimum cross-entropy for probabilities as a suitable methodology on the base of which inductive reasoning in the respective semantic frameworks can be realised in a straightforward way.

3.3 Different types of beliefs

Our approach to inductive reasoning via belief revision sketched above also distinguishes between explicit beliefs in a belief base, and implicit beliefs derivable in an epistemic state. The necessity of such a distinction is quite obvious in a belief change scenario, since implicit resp. derived beliefs are more easily changed than explicit beliefs. Having to give up explicit beliefs not only needs more effort, but it is quite a different thing. Formally, if $\Psi_\Delta = ind(\Delta)$, and the new information $\mathcal{I}$ is in conflict with Δ, e.g., $\Delta \cup \mathcal{I}$ is inconsistent, then we are still able to perform revision in the sense of updating via $\Psi_\Delta * \mathcal{I} = ind(\Delta) * \mathcal{I}$, whereas revision as genuine revision in the AGM sense via $ind(\Delta \cup \mathcal{I})$ would not be possible. If the agent comes to know that an explicit belief is (presumably) false, she might react more reluctant to incorporate it, trying perhaps to collect more evidence etc. If finally, she is ready to believe the new information, there are three possibilities: In the first case, the new information $\mathcal{I}$ might contradict the derived beliefs in Ψ_Δ but is nevertheless consistent with Δ, AGM revision $ind(\Delta \cup \mathcal{I})$

would be a suitable option. In the second case, the agent acknowledges that her previous explicit beliefs were erroneous before, in which case she has to perform a proper belief base change by applying merging techniques which are able to resolve conflicts[2]. In the third case, the agent admits that the current context has changed, and she has to adapt her beliefs to these changes, in which case one would find some updating process appropriate. Summarizing, our approach to inductive reasoning is able to deal with (and properly distinguish between) generic, background and contextual beliefs, on the one hand side, and explicit and implicit beliefs, on the other. This is possible by considering inductive reasoning within belief revision frameworks, and provides perfect grounds for a rich methodology that ensures coherence over different reasoning scenarios.

Furthermore, we mention an axiom for iterated revision that is particularly suitable to express coherence in the above sense, but which has been considered only in very few of the current belief revision frameworks and has been introduced under the name *Coherence* in [9] where it plays a crucial role for characterizing the principle of minimum cross entropy, but actually goes back to [16]:

(Coherence) $\Psi * (\Delta_1 \cup \Delta_2) = (\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$.

(Coherence) demands that adjusting any intermediate epistemic state $\Psi * \Delta_1$ to the full information $\Delta_1 \cup \Delta_2$ should result in the same epistemic state as adjusting Ψ by $\Delta_1 \cup \Delta_2$ in one step. The rationale behind this axiom is that if the new information drops in in parts, changing any intermediate state of belief by the full information should result unambigously in a final belief state. So, it guarantees the change process to be *logically coherent*.

Note that (Coherence) does not claim that $(\Psi*\Delta_1)*\Delta_2$ and $(\Psi*\Delta_1)*(\Delta_1\cup\Delta_2)$ are the same, just to the contrary – these two revised epistemic states will be expected to differ in general, because the first is not supposed to maintain prior contextual information, Δ_1, whereas the second should do so, according to success. However, (Coherence) can help ensuring independence of parts of the history that serves as background beliefs for inductive reasoning. In the situation described by (5) where we reason inductively from Δ with background beliefs Ψ_{bk}, imagine that we still are aware of the last conditional information Δ_0 that shaped Ψ_{bk}, i.e., $\Psi_{bk} = \Psi_1 * \Delta_0$, which would be mandatory to be able to distinguish among the different scenarios sketched above. But in general, it will be the case that Ψ_{bk} and Δ_0 do not determine Ψ_1 uniquely, so that there may be a different Ψ_2 satifying also $\Psi_{bk} = \Psi_1 * \Delta_0 = \Psi_2 * \Delta_0$. For updating Ψ_{bk}, this is irrelevant because only Ψ_{bk} matters. However, for AGM-like revision, we would like to compute $\Psi_{bk} * \Delta = \Psi_1 * (\Delta_0 \cup \Delta)$, but also $\Psi_2 * (\Delta_0 \cup \Delta)$ would be a suitable candidate. Here (Coherence) guarantees that the resulting epistemic state would be the same:

$$\Psi_1 * (\Delta_0 \cup \Delta) = (\Psi_1 * \Delta_0) * (\Delta_0 \cup \Delta) = (\Psi_2 * \Delta_0) * (\Delta_0 \cup \Delta) = \Psi_2 * (\Delta_0 \cup \Delta).$$

[2] Note that this would also be possible in our general framework, however, we leave this for future work here to not distract from the main focus of this paper

This makes clear that in our conceptual framework of inductive reasoning in the context of belief revision, integrating background beliefs and different pieces of information can be done in different but coherent ways. This means, having to deal with different pieces of information, the crucial question is not whether one information is more recent than others, but which pieces of information should be considered to be on the same level, i.e., belonging to the same type of belief (background vs. contextual), or referring to the same context (which may, but is not restricted to be, of temporal type). Basically, pieces of information on the same level are assumed to be compatible with one another, so simple set union will return a consistent set of formulas (please also see footnote 2). Pieces of information on different levels do not have to be consistent, here latter, or more reliable ones may override those on previous levels.

3.4 Reasoning on optimum entropy and with OCFs

We briefly illustrate the concepts presented in this section by inductive reasoning and revision with probabilities and ranking functions.

The principles of maximum entropy and minimum cross-entropy are powerful methodologies for inductive reasoning and belief revision in probabilistics. Due to lack of space, we cannot recall them fully here but refer in particular to [13, 9, 10]. For a (consistent) set of probabilistic conditionals $\varDelta$, the principle of maximum entropy selects the unique probability distribution $ME(\varDelta)$ with maximum entropy, and if prior information P is given, then the principle of minimum cross-entropy selects (under mild consistency conditions) a unique probability distribution $P *_{ME} \varDelta$ that is a model of $\varDelta$ and has minimal information distance to P, thus realizing probabilistic belief revision. The crucial equation for understanding and analyzing ME-revision is given by

$$P *_{ME} \varDelta(\omega) = \alpha_0 P(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i B_i}} \alpha_i^{1-x_i} \prod_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i \overline{B_i}}} \alpha_i^{-x_i}, \tag{8}$$

with the α_i's being exponentials of the Lagrange multipliers, one for each conditional in $\varDelta$, and have to be chosen properly to ensure that $P *_{ME} \varDelta$ satisfies all conditionals in $\varDelta$ with the associated probabilities. α_0 is simply a normalizing factor. For a complete axiomatization of the principle of minimum cross-entropy within the scope of probabilistic revision by conditional-logical postulates, see [9]. If P_u is a suitable uniform distribution, both ME-principles are related via $ME(\varDelta) = P_u *_{ME} \varDelta$. This means that ME is an inductive reasoning mechanism derived from a belief revision operator in the sense of (7), and $*_{ME}$ realises inductive reasoning from general background beliefs P in the sense of (5). Let us further note that ME-revision also satisfies (Coherence) [16]. Hence the ME-methodology is quite a perfect example to illustrate all concepts and relationships presented in this paper in a probabilistic framework.

Transferring the basic ideas underlying the ME-principles to the framework of ranking functions brings us to *c-revisions* and *c-representations* [10]. Formally, the c-revision methodology provides approaches to revision of ranking functions

κ by consistent sets Δ of conditionals, and inductive reasoning from conditional belief bases (also by taking background beliefs into account) according to (7) and (5) via the following schema:

$$\kappa *_c \Delta(\omega) = \kappa_0 + \kappa(\omega) + \sum_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i \overline{B_i}}} \kappa_i^- \tag{9}$$

where the parameters κ_i^- have to be chosen suitably to ensure that $\kappa *_c \Delta \models \Delta$, and κ_0 is a normalization factor. A c-representation of Δ is obtained from that by choosing the uniform prior $\kappa_u(\omega) = 0$ for all $\omega \in \Omega$. C-revisions satisfy (Coherence), but only when considered as a family of revisions (for more technical details, please see [12]).

4 Focusing and Conditioning

Focusing means applying generic knowledge to a reference class appropriate to describe the context of interest (cf. [3]). As this reference classe is assumed to be specified by factual information and indicates a shift in context (to that reference class), focusing should be performed by updating the current epistemic state to *factual* information which is certain, i.e. with probability 1. It can easily be shown that for ME-change, updating with such information results in conditioning the prior epistemic state, and indeed, conditioning is usually considered to be the proper operation for focusing.

However, in a probabilistic setting, conditioning has been used for revision, too [4, 3]. So revision and focusing are often supposed to coincide in the framework of Bayesian probabilities though they differ conceptually: revision is not only *applying knowledge*, but means incorporating a new constraint so as to *change knowledge*. Due to this conceptual mismatch, paradoxes have been observed. Gärdenfors investigated *imaging* as another proper probabilistic change operation [4]. Dubois and Prade argued that the assumption of having a uniquely determined probability distribution to represent the available knowledge at best is responsible for that flaw [3]).

However, we will show that in our framework, it is easily possible to treat revision as different from focusing without giving up the assumption of having a single, distinguished epistemic state as a base for inferences. The following proposition reveals the difference between revision by a certain information A, and focusing to A by conditioning; the proofs are straightforward using (8) but tedious.

Proposition 1. *Let P be a distribution, $\Delta \subseteq (\mathcal{L} \mid \mathcal{L})^{prob}$ a (P-consistent[3]) set of probabilistic conditionals, and suppose $A[1]$ to be a certain probabilistic fact.*

(i) Focussing on A, i.e., updating P with $A[1]$ is done by ME-revision and yields $P *_{ME} \{A[1]\} = P(\cdot|A)$; in particular, $(P *_{ME} \Delta) *_{ME} A[1] = (P *_{ME} \Delta)(\cdot|A)$.

[3] Δ is P-consistent if there is a distribution Q with $Q \models \Delta$ and $Q(\omega) = 0$ whenever $P(\omega) = 0$.

(ii) AGM-revising $P *_{ME} \Delta$ *with* $A[1]$ *yields* $P *_{ME} (\Delta \cup \{\mathrm{A}[1]\}) = P(\cdot|\mathrm{A}) *_{ME} \Delta$.

We illustrate that the correct usage of focusing and revision in the probabilistic framework helps resolving well-known paradoxes by considering an example that motivated the application of alternative approaches to uncertain reasoning like Dempster-Shafer theory [15].

Example 1. In a well-known example, Peter, Paul, and Mary are killers one of whom has been hired by Big Boss to commit a murder. Police Inspector Smith knows that Big Boss has first tossed a coin to decide whether it should be a male (Peter or Paul), or a female (Mary), but he does not know about the outcome of the tossing. So, initially, the explicit beliefs of Smith are given by $\Delta_1 = \{(Peter \vee Paul)[0.5], Mary[0.5]\}$, and his initial epistemic state can be calculated via the principle of maximum entropy: $P_1 = ME(\Delta_1)$. It is straightforward to see that $P_1(Mary) = 0.5, P_1(Paul) = P_1(Peter) = 0.25$.

Now Smith comes to know that Peter has been arrested right before the murder, so he could not have committed the crime. This piece of information can be encoded by $R_2 = \{\neg Peter[1]\}$. When incorporating Δ_2 by an update operation (which amounts to a conditioning here), the new epistemic state would be $P_2 = P_1(\cdot|\neg Peter)$, and hence the new beliefs concerning Paul and Mary would be $P_2(Mary) = \frac{2}{3}$, and $P_2(Paul) = \frac{1}{3}$. This seems to be unintuitive, as it gives undue precedence to Mary. However, this flaw is neither an argument against maximum entropy, nor against probability theory in general, but caused by the confusion between focusing and revision. The correct change operation here is revision as simultaneous change, which amounts to computing $P_3 = ME(\Delta_1 \cup \Delta_2)$. Now, in fact, we obtain $P_3(Mary) = P_3(Paul) = 0.5$, as expected.

A statement analogical to Proposition 1 holds for focussing and AGM-revision, in particular c-revision, for OCFs.

5 Conclusion

The aim of this paper is to describe inductive reasoning from conditional belief bases in a rich epistemic framework that takes epistemic states and conditionals as basic encodings of information. Allowing inductive reasoning from background beliefs (in the form of belief bases or epistemic states) leads us naturally to consider also belief revision. More boldly, our main claim here is that inductive reasoning can be considered as a special case of epistemic belief revision. In this way, a coherent and homogeneous approach to inductive reasoning is possible that allows us to realize different forms of inductive reasoning via AGM-like revision, updating, and focusing. We presented a general, abstract framework based on epistemic states and conditionals, and illustrated our ideas both for ordinal and probabilistic environments. We also showed how commonly known paradoxes can be avoided in our framework.

References

1. A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89:1–29, 1997.
2. B. de Finetti. La prévision, ses lois logiques et ses sources subjectives. In *Ann. Inst. H. Poincaré*, volume 7. 1937. English translation in *Studies in Subjective Probability*, ed. H. Kyburg and H.E. Smokler, 1964, 93-158. New York: Wiley.
3. D. Dubois and H. Prade. Non-standard theories of uncertainty in plausible reasoning. In G. Brewka, editor, *Principles of Knowledge Representation*. CSLI Publications, 1996.
4. P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, Mass., 1988.
5. P. Gärdenfors. Belief revision and nonmonotonic logic: Two sides of the same coin? In *Proceedings European Conference on Artificial Intelligence, ECAI'92*, pages 768–773. Pitman Publishing, 1992.
6. M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84:57–112, 1996.
7. E. Jaynes. *Papers on Probability, Statistics and Statistical Physics*. D. Reidel Publishing Company, Dordrecht, Holland, 1983.
8. H. Katsuno and A. Mendelzon. On the difference between updating a knowledge base and revising it. In *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning, KR'91*, pages 387–394, San Mateo, Ca., 1991. Morgan Kaufmann.
9. G. Kern-Isberner. Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence*, 98:169–208, 1998.
10. G. Kern-Isberner. A thorough axiomatization of a principle of conditional preservation in belief revision. *Annals of Mathematics and Artificial Intelligence*, 40(1-2):127–164, 2004.
11. G. Kern-Isberner. Linking iterated belief change operations to nonmonotonic reasoning. In G. Brewka and J. Lang, editors, *Proceedings 11th International Conference on Knowledge Representation and Reasoning, KR'2008*, pages 166–176, Menlo Park, CA, 2008. AAAI Press.
12. G. Kern-Isberner and D. Huvermann. What kind of independence do we need for multiple iterated belief change? *J. Applied Logic*, 22:91–119, 2017.
13. J. Paris. *The uncertain reasoner's companion – A mathematical perspective*. Cambridge University Press, 1994.
14. J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, Ca., 1988.
15. G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, NJ, 1976.
16. J. Shore and R. Johnson. Properties of cross-entropy minimization. *IEEE Transactions on Information Theory*, IT-27:472–482, 1981.
17. W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In W. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics, II*, pages 105–134. Kluwer Academic Publishers, 1988.

Rationally, the Universe is Infinite – Maybe*.

Jürgen Landes[1 [0000−0003−3105−6624]]

Department of Philosophy “Piero Martinetti”, University of Milan, Italy
juergen_landes@yahoo.de
https://jlandes.wordpress.com/

> Two things are infinite: the universe and human stupidity; and I'm not sure about the universe.

Abstract. This note shows that an explication of rationality within Pure Inductive Logic requires us to believe that the universe has infinitely many elements with probability one. A weaker explication of rationality within Pure Inductive Logic leaves open the possibility and even the necessity of a finite universe.

Keywords: Pure Inductive Logic · Uncertain Inference · Rationality· Induction · Universe.

1 Introduction

“Two things are infinite: the universe and human stupidity; and I'm not sure about the universe.” This quote, sometimes attributed to Albert Einstein [2, p. 478], captures (among other things) our desires to learn about the universe and the arising difficulties. Unlike Einstein, who developed physical theories that have been tested empirically, this short paper instead seeks to inform our beliefs about the size of the universe from the armchair relying on our rational faculties.

The Copernican Principle postulating that our place in the world is not special [1]; it has been used to reason about the expected physical size of an alien (most species are expected to exceed 300 kg in body mass) and the size of their home planet(s) [16]. However, the principle and its applications have a less than stellar standing according to some critics. The framework of Pure Inductive Logic instead is a rigorous approach to uncertain inference in the absence of background information [5,14,15].

The rest of this short note is organised as follows: I next introduce the formal framework (§ 2.1), explicate rationality in this framework (§ 2.2) and capture

* I gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) 405961989 and 432308570

what it means to rationally hold that the universe is infinite. I go on to show that depending on the explication of rationality the universe is rationally infinite (Theorem 2) or not (Theorem 3) and conclude (§ 3). The appendix contains some technicalities.

2 Formal Analysis

There are a number of approaches for drawing uncertain inferences. In this paper, I shall employ Pure Inductive Logic (PIL) [14]. This approach has three main ingredients: firstly, a fixed language of first order logic is chosen to represent the propositions we want to reason about; secondly, all sentences of this language are assigned probabilities representing our rational degrees of belief, our credences. Finally, principles of rationality constrain the choice of probabilities.

2.1 The Framework

Since I'm here interested only in sizes of domains and not in properties of the elements of these domains, I pick the language without relation symbols and without function symbols.

The Language $\mathbb{L}$ of first order logic contains countably many constant symbols $a_1, a_2, \ldots$ and countably many variables $x, y, z, \ldots$. The language $\mathbb{L}$ does not contain relation symbols nor function symbols but does contain a symbol for equality, $\equiv$. From these symbols the sentences of the language, $S\mathbb{L}$, and the quantifier free sentences of the language, $QFS\mathbb{L}$, can be constructed in the usual way [14, P. 9].

Probability functions w are maps $w : S\mathbb{L} \to [0, 1]$ satisfying three axioms [14, P. 11]:

P1 If $\tau \in S\mathbb{L}$ is a tautology, $\models \tau$, then $w(\tau) = 1$.
P2 If $\tau, \theta \in S\mathbb{L}$ are mutually exclusive, $\tau \models \neg\theta$, then $w(\tau \vee \theta) = w(\tau) + w(\theta)$.
P3 $w(\exists x \theta(x)) = \lim_{n\to\infty} w(\bigvee_{i=1}^{n} \theta(a_i))$.

P1 expresses the thought that sure events have probability 1. P2 is the usual condition concerning the additivity of mutually exclusive events. P3 means that for every element in an underlying domain (or universe) there is at least one constant symbol representing it. We can think of the constants as names for the elements of the underlying domain. Every name is associated with one and only one element; however, elements may have more than one name.

Three axioms of equality are imposed to ensure that the logic behaves as properly with respect to equality. The following types of sentences are added to the tautologies of the logic:

1. $a_i \equiv a_i$ for all constants a_i. (Reflexivity)
2. $a_i \equiv a_k \to a_k \equiv a_i$ for all constants a_i, a_k. (Symmetry)
3. $(a_i \equiv a_k \wedge a_k \equiv a_s) \to a_i \equiv a_s$ for all constants a_i, a_k, a_s. (Transitivity)

In the following, logical consequence is restricted to structures which satisfy these axioms of equality.

The equality symbol needs to behave in the expected way also with respect to assigned probabilities. It is hence required that every probability function w satisfies three further axioms of equality [14, Chapter 37] (probabilities need to respect logical equivalence):

1. $w(a_i \equiv a_i) = 1$ for all constants a_i. (Reflexivity)
2. $w(a_i \equiv a_k \to a_k \equiv a_i) = 1$ for all constants a_i, a_k. (Symmetry)
3. $w((a_i \equiv a_k \wedge a_k \equiv a_s) \to a_i \equiv a_s) = 1$ for all constants a_i, a_k, a_s. (Transitivity)

Gaifman's Theorem states that probability functions are determined by the probabilities they assign to quantifier free sentences [4]. Probability functions are hence uniquely determined by assigning numbers to all quantifier free sentences such that the assignment satisfies P1 and P2.

A table τ (on p) is a complete description of the world containing only $a_1, \ldots, a_p$ [14, P. 276]:

$$\tau = \bigwedge_{i,k=1}^{p} a_i \equiv^{\epsilon_{i,k}} a_k$$

where $\epsilon_{i,k} \in \{0,1\}$ with $\equiv^1 := \equiv$ and $\equiv^0 := \not\equiv$. We denote by T_p the set of all tables on p that are reflexive, symmetry and transitive.

A simple application of the Disjunctive Normal Form Theorem shows that for all quantifier free sentences, $\varphi \in QFS\mathbb{L}$, such that no constant with a greater index than p is mentioned in φ, it holds that

$$\models \varphi \leftrightarrow \bigvee_{\substack{\tau \in T_p \\ \tau \models \varphi}} \tau \ .$$

Since probability functions assign logically equivalent sentences the same probability [14, Proposition 3.1], it follows that every probability function is completely determined by the probabilities it assigns to tables. Gaifman's Theorem and this observation hence allow us to quickly and simply define probability functions $w : S\mathbb{L} \to [0,1]$.

2.2 Rationality

Rationality can then be explicated by principles of rationality constraining the choice of a probability function. Traditionally, principles of rationality in Pure Inductive Logic are often rooted in intuitions we have about symmetries, (ir-)relevance and/or induction. I next introduce two principles of rationality.

Constant Exchangeability (CX) is the most widely accepted principle of rationality in PIL. It brands it as irrational to treat constants differently in the absence of all information. Formally, if $a_k \notin \varphi \in S\mathbb{L}$, then $w(\varphi) = w(\varphi(\frac{a_k}{a_i}))$

where $\varphi(\frac{a_k}{a_i})$ is obtained from $\varphi \in S\mathbb{L}$ by replacing all occurrences of the constant a_i in φ by the constant a_k.

Johnson's Sufficientness Principle (JSP) is one of the oldest principle of rationality [7], it even predates [3] by more than a decade. JSP says in the present context that the probability of an unobserved constant, a_{p+1}, being indiscernible from an observed constant, $w(a_{p+1} \equiv a_i)$, only depends on the number of observations, p, and the number of constants which are indiscernible from a_{p+1}. That is, $w(\tau_{p+1}|\tau_p)$ for $\tau_p \in T_p, \tau_{p+1} \in T_{p+1}$ with $\tau_{p+1} \vDash \tau_p$ only depends on p and on $|\{1 \leq k \leq p \ : \ \tau_{p+1} \vDash a_k \equiv a_{p+1}\}|$. In other words, $w(\tau_{p+1}|\tau_p)$ is given by some function $f(p, |\{1 \leq k \leq p \ : \ \tau_{p+1} \vDash a_k \equiv a_{p+1}\}|)$.

The set of probability functions satisfying CX and JSP is given by a 1-parameter family w_λ parametrised by $\lambda \in (0, \infty)$ and two probability functions at the endpoints, $\lambda \in \{0, \infty\}$, see [8, Theorem 21] and [14, Theorem 38.1].

$\lambda = 0$ gives rise to the unique 1-heterogeneous probability function according to which all constants are equal to each other with probability one (Carnap's c_0), $w_0(a_i \equiv a_{i+1}) = 1$ for all i. All names refer to the same element. That is, the universe contains only one element with probability one.

For $\lambda = \infty$ one obtains the completely independent probability function according to which all two pairwise different constants are different with probability one. For every element in the universe there is only one name, $w_\infty(a_i \equiv a_k) = 0$ for all $i \neq k$, and thus the universe contains infinitely many elements with probability one.

In the context of PIL, these two probability functions are peculiar functions which are often considered as unsuitable for explicating inductive intuitions. The first probability function is unsuitably opinionated in the absence of background information. The second probability function does not capture inductive entailment [18, P. 58]. There are however further probability functions satisfying CX and JSP.

Theorem 1 (Theorem 21 of [8]). *All other probability functions w defined on $\mathbb{L}$ satisfying CX and JSP are given by*

$$w_\lambda(\tau) = \lambda^{t-1} \cdot \left(\prod_{i=1}^{p-1} \frac{1}{i + \lambda} \right) \cdot \left(\prod_{j=1}^{t} (x_j - 1)! \right),$$

where $\tau \in T_p$, t is the number of discernible constant symbols in τ (the number of equivalence classes) and the x_j are the size of equivalence classes according to τ.[1]

[1] There is an unfortunate slip-up in [8, Theorem 21] where the exponent of the first factor is t rather than $t - 1$. This factor arises from every constant symbol which is different from all the previous ones, with the exception of the very first constant a_1 for which $w(a_1 \equiv a_1) = 1$ needs to hold by the axioms of equality.

In the terminology of PIL, the x_j are the *spectrum* of τ.[2]

To simplify notation let T_p^t and $T_p^{\leq t}$ denote the sets of tables on p which have exactly t equivalence classes, respectively, less or equal than t equivalence classes.

2.3 Rationally, the Universe contains infinitely many Elements

The number of elements in the universe can be measured by how many constant symbols are pairwise indiscernible. However, a table τ on p can only tell us about the first p constant symbols. The number of elements in the universe becomes accessible only when p is sent to infinity. This allows us to state that the universe contains exactly t elements with non-zero probability by:

$$\lim_{p\to\infty} w(T_p^t) := \lim_{p\to\infty} \sum_{\tau\in T_p^t} w(\tau) > 0$$

according to some probability function w.

That the universe contains infinitely many elements with probability zero can then be formalised by increasing the number of indiscernible elements:

$$\lim_{p\to\infty} w(T_p^{\leq t}) := \lim_{p\to\infty} \sum_{\tau\in T_p^{\leq t}} w(\tau) = 0 \quad \text{for all fixed numbers } t \in \mathbb{N} \ .$$

The rational probability that the universe contains only finitely many elements is thus equal to the following expression

$$\lim_{t\to\infty} \lim_{p\to\infty} w_\lambda(T_p^{\leq t}) \ .$$

I'm obliged to Alena Vencovská and Jeff Paris for pointing out the following result proved in the Appendix.

Theorem 2. *For all $\lambda > 0$ it holds that*

$$\lim_{t\to\infty} \lim_{p\to\infty} w_\lambda(T_p^{\leq t}) = 0 \ .$$

[2] Tangentially, we can now also formalise and answer the question asked in many bars "Don't I know you from somewhere?" Formally, the probability of having met is

$$\begin{aligned} w_\lambda\Big(\bigvee_{i=1}^{p} a_{p+1} \equiv a_i | \tau\Big) &= 1 - \frac{w_\lambda(\tau \wedge \bigwedge_{i=1}^{p} a_{p+1} \not\equiv a_i)}{w_\lambda(\tau)} \\ &= 1 - \frac{\lambda^{t+1}}{\lambda^t} \cdot \frac{1}{p+\lambda} \cdot \frac{1! \cdot \prod_{j=1}^{t}(x_j-1)!}{\prod_{j=1}^{t}(x_j-1)!} = \frac{p}{p+\lambda}, \end{aligned}$$

which converges to one for ever greater sample sizes, $\tau \in T_p$ with growing p, for $0 < \lambda < \infty$. As we get older (increasing p), we become more and more sure to have already met the person at the bar. Note that $w_\lambda(\tau) > 0$ for all $p \geq 1$ and all $\tau \in T_p$; the conditional probability considered is hence well-defined. $w_0(\bigvee_{i=1}^{p} a_{p+1} \equiv a_i|\tau) = 1$ and $w_1(\bigvee_{i=1}^{p} a_{p+1} \equiv a_i|\tau) = 0$.

Thus, CX and JSP jointly entail that either there is only one single thing in the universe ($w = w_0$) or the universe contains infinitely many elements ($w = w_\lambda$ with $\lambda > 0$). Having rejected the former answer, we conclude with probability one that there are infinitely many elements in the universe.

2.4 A Weaker Notion of Rationality

The axiom JSP is arguably too strong of a demand to explicate rationality. Could the conditional probability $w(\tau_{p+1}|\tau_p)$ not also depend on, say, the number of equivalence classes in τ_p? The question arises what we ought to believe about the make up of the universe under a weaker construal of rationality, i.e., how much pressure can Theorem 2 bear before breaking down?

Alternatively, one may think that our inductive assumptions ought not to deductively entail that the universe is infinite with probability one. Instead, a finite universe should remain an open possibility.[3] Within the PIL framework one may then be willing to give up on CX, JSP or both these axioms. Since the axiom CX is more widely accepted (and studied) than JSP, I will here drop JSP and keep CX.

Just assuming the axiom CX, [14, Corollary 37.2] demonstrates that for every probability function w satisfying Ex on the language containing only the equality symbol there has to exist an associated probability function w' defined on the language containing only one binary relation symbol but not the equality symbol. Furthermore, w' satisfies the axiom of Spectrum Exchangeability (Sx) [6,9,10,11,12,13] and [14, Chapters 27-38]. This axiom roughly states that probabilities of worlds only depend on how the world distinguishes constants but not on the actual properties of the constants at this world.

Probability functions on this predicate language L (and all other predicate languages with at least one non-unary relation symbol) come in two different shapes. They either satisfy Li with Sx or they don't (Li stands for *L*anguage *I*nvariance). Li with Sx harkens back to Carnap [17, P. 974].

Li with Sx requires that there exists a mutually consistent family of probability functions $w^{\mathcal{L}}$ that all satisfy Sx defined on all predicate languages such that for all predicate languages L, L' it holds that w^L and $w^{L'}$ agree on all sentences φ that are sentences of L and L'. In words, whenever w^L is a probability function on L satisfying Sx and L' is a language containing all the relation symbols of L and further relation symbols, then there exists a probability function on L' that satisfies Sx, $w^{L'}$; $w^{L'}$ agrees with w^L on all the sentences of L.

All probability functions w **satisfying Li with Sx** have a particular form [14, Theorem 37.1]:

$$w^L = \int_{\vec{p}\in\mathbb{B}} u^{\vec{p},L}\, d\mu(\vec{p})$$

[3] Many thanks to Jon Williamson for pointing this out to me.

for some measure μ, the *de Finetti prior*, where the $u^{\vec{p},L}$ are particular probability functions defined on L that satisfy Sx. Technical details do not concern us here; some can be found in the appendix.

Applying [14, Corollaries 37.2 and 37.3] we find that all probability functions w (defined on the language with equality but without relation symbols $\mathbb{L}$) satisfying Ex can be written as:

$$w = \int_{\vec{p}\in\mathbb{B}} u^{\vec{p}}_{\mathrm{Eq}}\, d\mu(\vec{p}) \ .$$

Again, many details shall not be relevant here, see [14, § 37] for details and the appendix for main points.

Relevantly, since w is a "convex mixture" of $u^{\vec{p}}_{\mathrm{Eq}}$ (w is an integral), the following are logically equivalent:

1. $\lim_{t\to\infty}\lim_{p\to\infty} w(T_p^{\leq t}) = 0$ and
2. the measure $\mu(\vec{p})$ puts all mass on probability functions which satisfy $\lim_{t\to\infty}\lim_{p\to\infty} u^{\vec{p}}_{\mathrm{Eq}}(T_p^{\leq t}) = 0$.

The latter condition is known to depend only on the infinite sequence of numbers $\vec{p} = \langle p_0, p_1, p_2, \ldots\rangle$. By construction this is a sequence of, not necessarily strictly, positive numbers summing to 1 such that $p_1 \geq p_2 \geq p_3 \ldots$.[4] The following two conditions are equivalent [14, Chapters 29 and 30]:

1. $\lim_{t\to\infty}\lim_{p\to\infty} u^{\vec{p}}_{\mathrm{Eq}}(T_p^{\leq t}) = 0$ and
2. $p_0 > 0$ or there are infinitely many $p_i > 0$.

Letting

$$\mathbb{B}_\infty := \{\vec{p}\in\mathbb{B} \mid p_0 > 0 \text{ or there are infinitely many } p_i > 0\} \ ,$$

I can now compactly state main result of this section.

Theorem 3. *For all probability functions w on $S\mathbb{L}$ satisfying Ex that are associated with a family of probability functions defined predicate languages satisfying Li with Sx it holds that*

$$w = \int_{\vec{p}\in\mathbb{B}} u^{\vec{p}}_{Eq}\, d\mu(\vec{p})$$

for some measure $\mu(\vec{p})$. Furthermore, the following three conditions are equivalent

1. $\lim_{t\to\infty}\lim_{p\to\infty} w(T_p^{\leq t}) = 0$,
2. $\mu(\mathbb{B}_\infty) = 1$,
3. $w = \int_{\vec{p}\in\mathbb{B}_\infty} u^{\vec{p}}_{Eq} d\mu(\vec{p})$.

[4] Note that there is no requirement that $p_0 \geq p_1$.

As a result, the probability that the universe contains infinitely many elements is equal to the measure the de Finetti prior μ assigns to B_∞, $\int_{p \in B_\infty} d\mu(\vec{p})$. In other words, Theorem 2 can take some heat but breaks down when stretched widely.

In particular, if $\int_{p \in B_\infty} d\mu(\vec{p}) = 0$, then the universe contains infinitely many elements with probability zero. In that case, we are rational to fully believe that the universe contains only finitely many elements.

3 Conclusions

We have seen that the framework of PIL allows us to reason about the (size of the) universe. Depending on how exactly rationality is explicated, PIL gives different answers as to how many different elements we should rationally expect to see in the universe – in the absence of all evidence. Assuming that elements are of equal size (or at least that there is a strictly positive lower bound on their size), believing that there are infinitely many different elements entails a belief in an infinitely large universe.

It is tempting to capture our uncertainty about the size of the universe by the philosophical puzzle of adopting *the* correct de Finetti prior, which has posed a formidable challenge to many great thinkers. A convincing solution has yet to be found.

Reversing our train of thought, this analysis can also be read as a way to inform our thinking about rational probabilities via science, cf. [19] for whether one should in general read $A \to B$ as "A entails B" or as "$\neg B$ entails $\neg A$". Knowing that the universe is infinite and reasonably homogeneous entails that we know of the existence of infinitely many elements. Subsequently, assuming Li with Sx rules out certain de Finetti priors [(science + Li with Sx) constrains the set of rational priors]. However, until science has determined the size of the universe, I suggest to read my argument in the here presented direction and worry about the reverse when that day comes.

Acknowledgements Many thanks to Alena Vencovská and Jeff Paris for pointing out Theorem 2 and further useful comments. Thanks also to Jon Williamson for the suggestion in § 2.4 and further helpful comments.

References

1. Bondi, H.: Cosmology. Cambridge University Press, Cambridge (1952)
2. Calaprice, A. (ed.): The ultimate quotable Einstein. Princeton University Press, Princeton, 2 edn. (2011)
3. Carnap, R.: On Inductive Logic. Philosophy of Science **12**(2), 72–97 (1945), https://doi.org/10.1086/286851
4. Gaifman, H.: Concerning measures in first order calculi. Israel Journal of Mathematics **2**(1), 1–18 (1964), https://doi.org/10.1007/BF02759729

5. Hill, A., Paris, J.B.: An Analogy Principle in Inductive Logic. Annals of Pure and Applied Logic **164**(12), 1293–1321 (2013), https://doi.org/10.1016/j.apal.2013.06.013
6. Howarth, E., Paris, J.B.: The Theory of Spectrum Exchangeability. The Review of Symbolic Logic **8**, 108–130 (3 2015), https://doi.org/10.1017/S1755020314000331
7. Johnson, W.E.: Probability: The Deductive and Inductive Problems. Mind **41**(164), 409–423 (1932)
8. Landes, J.: The Principle of Spectrum Exchangeability within Inductive Logic. Ph.D. thesis, Manchester Institute for Mathematical Sciences (2009), https://jlandes.files.wordpress.com/2015/10/phdthesis.pdf
9. Landes, J., Paris, J.B., Vencovská, A.: Language Invariance and Spectrum Exchangeability in Inductive Logic. In: Mellouli, K. (ed.) Proceedings of ECSQARU. LNAI, vol. 4724, pp. 151–160. Springer (2007), https://doi.org/10.1007/978-3-540-75256-1_16
10. Landes, J., Paris, J.B., Vencovská, A.: Representation Theorems for probability functions satisfying Spectrum Exchangeability in Inductive Logic. International Journal of Approximate Reasoning **51**(1), 35–55 (2009), https://doi.org/10.1016/j.ijar.2009.07.001
11. Landes, J., Paris, J.B., Vencovská, A.: A characterization of the Language Invariant families satisfying Spectrum Exchangeability in Polyadic Inductive Logic. Annals of Pure and Applied Logic - Proceedings of IPM 2007 **161**(6), 800–811 (2010), https://doi.org/10.1016/j.apal.2009.06.010
12. Landes, J., Paris, J.B., Vencovská, A.: A survey of some recent results on Spectrum Exchangeability in Polyadic Inductive Logic. Synthese **181**, 19–47 (2011), https://doi.org/10.1007/s11229-009-9711-9
13. Nix, C.J., Paris, J.B.: A Note on Binary Inductive Logic. Journal of Philosophical Logic **36**(6), 735–771 (2007), https://doi.org/10.1007/s10992-007-9066-y
14. Paris, J.B., Vencovská, A.: Pure Inductive Logic. Cambridge University Press, Cambridge (2015)
15. Paris, J.B., Vencovská, A.: Six Problems in Pure Inductive Logic. Journal of Philosophical Logic **48**, 731–747 (2019), https://doi.org/10.1007/s10992-018-9492-z
16. Ruhmkorff, S., Jiang, T.: Copernican Reasoning About Intelligent Extraterrestrials: A Reply to Simpson. Journal for General Philosophy of Science **50**(4), 561–571 (2019), https://doi.org/10.1007/s10838-019-09465-7
17. Schlipp, P.A. (ed.): The Philosophy of Rudolf Carnap. Open Court, La Salle (1963)
18. Williamson, J.: Lectures on Inductive Logic. Oxford University Press, Oxford (2017), https://doi.org/10.1093/acprof:oso/9780199666478.001.0001
19. Williamson, J.: One philosopher's modus ponens is another's modus tollens: Pantomemes and nisowir. Metaphilosophy (2022), https://doi.org/10.1111/meta.12546

Appendix

Proof of Theorem 2

Theorem 2. *For all $\lambda > 0$ it holds that*

$$\lim_{t\to\infty}\lim_{p\to\infty} w_\lambda(T_p^{\leq t}) = 0 \ .$$

Proof. $\lambda = \infty$ is a trivial case.

Now consider $0 < \lambda < \infty$. Let us first note that the probability that any given table $\tau \in T_p$ with any number of equality classes, the conditional probability (given τ) that no extension by n more constants introduces a new class is

$$g(p,n) := \frac{p}{p+\lambda} \cdot \frac{p+1}{p+1+\lambda} \cdot \ldots \cdot \frac{p+n-1}{p+n-1+\lambda} \quad .$$

Since $\lambda > 0$, there exists some $M \in \mathbb{N}$ such that $\lambda \geq \frac{1}{M}$. We thus note

$$g(p,n) \leq \frac{p}{p+\frac{1}{M}} \cdot \frac{p+1}{p+1+\frac{1}{M}} \cdot \ldots \cdot \frac{p+n-1}{p+n-1+\frac{1}{M}} \quad .$$

We next observe that for all the factors on the right ($0 \leq i \leq n-1$) it holds that

$$\begin{aligned}\frac{p+i}{p+i+\frac{i}{M}} = \frac{M}{M} \cdot \frac{p+i}{p+i+\frac{1}{M}} &= \frac{pM+iM}{pM+iM+1} \\ &\leq \sqrt[M]{\frac{pM+iM}{pM+iM+1} \cdot \frac{pM+iM+1}{pM+iM+2} \cdot \ldots \cdot \frac{pM+iM+M-1}{pM+iM+M}} \quad .\end{aligned}$$

Inserting these inequalities back in, we obtain

$$g(p,n) \leq \sqrt[M]{\frac{pM}{pM+nM}} = \sqrt[M]{\frac{p}{p+n}} \quad . \tag{1}$$

Applying this observation to $\tau = a_1 \equiv a_1$ (the only table in T_1), we see that the probability of the first $n+1$ constants are all the same is

$$\frac{1}{1+\lambda} \cdot \frac{2}{2+\lambda} \cdot \ldots \cdot \frac{n}{n+\lambda} \leq \sqrt[M]{\frac{1}{1+n}} < \sqrt[M]{\frac{1}{n}} \quad .$$

Hence, the probability of the first $n+1$ constants falling into at least two classes is greater or equal to $1 - \sqrt[M]{\frac{1}{n}}$.

Similarly, for every table $\tau \in T_{n+1}$ the conditional probability that the next n^2 constants do not introduce a new class is

$$\begin{aligned}\frac{n+1}{n+1+\lambda} \cdot \frac{n+2}{n+2+\lambda} \cdot \ldots \cdot \frac{n^2+n}{n^2+n+\lambda} &\overset{(1)}{\leq} \sqrt[M]{\frac{n+1}{n^2+n+1}} \\ &= \sqrt[M]{\frac{n}{n} \cdot \frac{1+\frac{1}{n}}{n+1+\frac{1}{n}}} \\ &\leq \sqrt[M]{\frac{2}{n}} \quad .\end{aligned}$$

With ever greater n this expression vanishes, too. Thus, overall it holds that $\lim_{p\to\infty} w_\lambda(T_p^{\leq 1}) = 0$.

Continuing in this manner adding p^t constants to a table $\tau \in T_{p^{t-1}+1}$ we have

$$\begin{aligned} g(p^{t-1}+1, p^t) &= \frac{p^{t-1}+1}{p^{t-1}+1+\lambda} \cdot \ldots \cdot \frac{p^{t-1}+p^t}{p^{t-1}+p^t+\lambda} \\ &\overset{(1)}{\leq} \sqrt[M]{\frac{p^{t-1}+1}{p^{t-1}+p^t+1}} \\ &= \sqrt[M]{\frac{1+\frac{1}{p^{t-1}}}{1+p+\frac{1}{p^{t-1}}}} \\ &\leq \sqrt[M]{\frac{2}{p}} \ . \end{aligned}$$

We hence find that for all large enough p greater some fixed $t \in \mathbb{N}$ it holds that $\lim_{p\to\infty} w_\lambda(T_p^{\leq t}) = 0$.

We thus conclude that for all $\lambda > 0$ it holds that

$$\lim_{t\to\infty} \lim_{p\to\infty} w_\lambda(T_p^{\leq t}) = 0 \ .$$

Some further Technicalities

The probability functions $u^{\vec{p},L}$ and $u_{\mathrm{Eq}}^{\vec{p}}$ are generated by drawing balls from an urn. Consider an urn containing a black ball and at most countably-many balls with unique colours (black is not a colour here). We draw balls from the urn with the following probabilities: $p_0 \geq 0$ for the black ball and $p_i \geq 0$ for the colours. We hence need to have that $\sum_{i=0}^{\infty} p_i = 1$. Let $\mathbb{B}$ be the set of all such urns.

Draws are with replacement (balls are put back into the urn after pulling them out) and independent from each other. We assume without loss of generality that the colours are ordered such that $p_i \geq p_{i+1}$ for all $i \geq 1$ (again note the special role of 'black', p_0). Let $c_1, \ldots, c_p \in \mathbb{N}^p$ (to simplify notation $0 \in \mathbb{N}$ is assumed) be a sequence of drawn balls, then the probability of this draw is $\prod_{i=1}^{p} p_{c_i}$.

Given a fixed urn (a non-negative vector $\vec{p}$ of countable length, non-increasing from the second element onwards and summing to one) and a draw of p balls, $c_1, \ldots, c_p$, we then associate a unique table $\tau \in T_p$ with this draw as follows:

$$a_i \equiv a_i \text{ for all } 1 \leq i \leq p$$

and for all $i \neq k$

$$\begin{aligned} a_i &\equiv a_k, \quad \text{if } c_i = c_k \neq 0 \\ a_i &\not\equiv a_k, \quad \text{if } c_i = c_k = 0 \text{ or } c_i \neq c_k \ . \end{aligned}$$

Intuitively, if two balls have the same colour different from black, then the corresponding constants are equal. If the colours are different or at least one of them is black, then the corresponding constants are different.

For a given $\tau \in T_p$ let $C(\tau)$ be the set of draws of p balls that are associated with this table. Then the probability function $u^{\vec{p}}_{\mathrm{Eq}}$ is defined by the values it gives to all tables:

$$u^{\vec{p}}_{\mathrm{Eq}}(\tau) = \sum_{\vec{c} \in C(\tau)} \prod_{i=1}^{p} p_{c_i} \quad .$$

So, $u^{\vec{p}}_{\mathrm{Eq}}(\tau)$ is the joint probability of all draws associated with the table τ.

The probability functions $u^{\vec{p},L}$ are also defined in terms of (i.i.d.) sampling urns from balls and constructing possible worlds associated with draws of balls. The definition is however much more messy, see [14, Chapter 29] for full details.

One key difference is that the drawing of a previously unseen colour or a black ball does not entail that the constant is necessarily distinguishable from the previous constants in the associated worlds. Instead, upon drawing a new colour or a black ball all ways the new constant can behave[5] are equally likely. On the other hand, if a colour is not drawn for the first time, then there is only one way that constant can behave; its has to behave in the exact same way as all constants associated with the same colour.

One important connection between the $u^{\vec{p},L}$ and the $u^{\vec{p}}_{\mathrm{Eq}}$ is the fact that for all fixed $\vec{p}$ the latter is – in some sense – equal to the limit of $u^{\vec{p},L}$ where the limit is over ever larger languages containing finitely many relation symbols, at least one of them is not unary, but no symbol for equality [14, Corollary 37.3], $u^{\vec{p}}_{\mathrm{Eq}}$ "$=$" $\lim_{|L| \to \infty} u^{\vec{p},L}$.

[5] There is one condition here that is best explained by means of an example. Suppose that $c_2 = c_4$ and ($c_6 = 0$ or $c_6 \notin \{c_1, c_2, c_3, c_4, c_5\}$). Then a_6 cannot be such that a_2 and a_4 can be told apart. That is, $Ra_2a_6 \wedge \neg Ra_4a_6$ cannot hold in associated worlds. The condition thus is that indistinguishability due to colours (those that are not due to chance) must be preserved. By contrast, indistinguishability due to chance is broken for ever longer draws of balls with limit probability one, if the urn contains a black ball ($p_0 > 0$) or infinitely many colours. In words, indistinguishability between constants obtains in the limit, if and only if they are associated with the same non-black colour – if the urn contains a black ball ($p_0 > 0$) or if the urn contains infinitely many colours. Compare this with $u^{\vec{p}}_{\mathrm{Eq}}$ according to which constants are equal, if and only if they are associated with the same non-black colour.

Extended Abstracts

Logic-based Tractable Approximations of Probability *

Paolo Baldi[1] and Hykel Hosni[1]

University of Milan,
{paolo.baldi,hykel.hosni}@unimi.it

Probabilities can be formulated as functions over sets or, equivalently, in logical terms, over formulas of classical logic [5]. The logical formulation helps in particular in highlighting two strong idealizations which result from the combination of classical logic and the axioms of probability. First, classical probability functions are unable to distinguish between "probabilistic knowledge" and "probabilistic ignorance", since $P(\neg\theta) = 1 - P(\theta)$ for any formula θ. This amounts to saying that they have no direct way of representing a very common situation: the agent doesn't known anything about θ. Another important problem comes from the monotonicity of probability functions with respect to $\vdash$: i.e.

$$\theta \vdash \varphi \text{ implies } P(\theta) \leq P(\varphi), \tag{1}$$

This property is mathematically convenient, but puts on an agent a very heavy burden which owes to the intractability of $\vdash$. Indeed, a logical deduction from θ to φ might be highly nontrivial and hard to find, making the application of (1) a constraint of rationality that realistic agents may not be in a position to comply with.

In our recent work [2] we tackle both problems by replacing classical logic $\vdash$ with the family of Depth-bounded Boolean logic (DBBL) investigated in [4, 3].

A characteristic feature of DBBLs is its informational nature. Whereas connectives in classical logic are defined in terms of truth-values, DBBL provide an informational view of logical consequence which, as a by-product, also provides a tractable approximation of it.

The central idea behind the (semantic) approach to DBBL is to distinguish two kinds of information. The first is information which the agent possesses explicitly or can trivially infer from it. For definiteness, it is the kind of information that an agent holds when she holds the information that a conjunction is true, i.e. that both conjuncts are true.

The second is information that the agent does not actually possess, but temporarily assumes in the course of *hypotethical reasoning*. This is, for instance, the information used in a proof by cases of a mathematical theorem.

The hierarchy of Depth-bounded Boolean logic arises by bounding the number k of allowed nested iterations of inferences using hypothetical information. If $k = 0$, then only information actually held by the agent can be used as the

* Funded by the Italian Ministry of University (MUR) as part of the the PRIN 2017 project n. 20173YP4N 3"

premises of a logical deduction, yielding the 0-depth consequence relation denoted by $\vdash_0$.

For $0 \leq k < m$ it can be shown that $\vdash_k \subset \vdash_m$, whereas $\lim_{k\to\infty} \vdash_k = \vdash$. This justifies interpreting the hierarchy of DBBLs as an approximation to classical logic, which is indeed attained when the agent is allowed an unbounded use of hypothetical information.

In analogy to DBBLs, we introduce a hierarchy of *Depth-bounded Belief functions* which approximate probability functions and asymptotically coincide with them. Our construction is inspired by the theory of Dempster-Shafer Belief functions [6], as suggested by our choice of terminology. As in DS-theory, none of the belief functions in our hierarchy is constrained by additivity, except for the one attained in the limit, which is a probability function.

Let us recall that Belief functions allow for $B(\varphi \vee \psi) \geq B(\varphi) + B(\psi)$, with $\varphi, \psi \vdash \bot$, since they reflect the information that an agent possesses. Information for a disjunction may indeed be held in the absence of any information about the disjuncts, as in the famous Ellsberg-like scenarios.

Our framework connects higher logical abilities of an agent (as captured by the index of the relation $\vdash_k$) with the ability to obtain increasingly tighter approximations of $B_k(\varphi)+B_k(\psi)$ by $B_k(\varphi\vee\psi)$. This puts forward a seemingly novel approach, a logic-based one, to Ellsberg-like scenarios. As a welcome byproduct, reasoning tasks, such as variants of PSAT, based on members B_k of our hierarchy, will turn out to be computationally tractable.

Our main results read as follows. We show that each probability function can be approximated by a hierarchy of Depth-bounded Belief functions, and, conversely, we single out the conditions under which our Depth-bounded Belief functions actually determine a probability in the limit. Finally, we prove that under rather palatable restrictions, the depth-bounded functions introduced here are an adequate tool to tackle the well-known unfeasibility of logic-based uncertain reasoning [5].

The framework and results presented here are based on [2]. If time allows, we will sketch ongoing work, first reported in [1], which provide an approximation of probabilities using *qualitative* depth-bounded belief functions.

References

1. Paolo Baldi and Hykel Hosni. Logical approximations of qualitative probability. In *ISIPTA*, volume 147 of *Proceedings of Machine Learning Research*, pages 12–21. PMLR, 2021.
2. Paolo Baldi and Hykel Hosni. A logic-based tractable approximation of probability. *Journal of Logic and Computation*, 05 2022.
3. Marcello D'Agostino. An informational view of classical logic. *Theoretical Computer Science*, 606:79–97, nov 2015.
4. Marcello D'Agostino, Marcelo Finger, and Dov Gabbay. Semantics and proof-theory of depth bounded Boolean logics. *Theoretical Computer Science*, 480:43–68, apr 2013.
5. J. B. Paris. *The Uncertain Reasoner's Companion: A Mathematical Perspective*. Cambridge University Press, 1994.

6. G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.

Towards Learning Argumentation Frameworks from Labelings

Lars Bengel[0000−0003−0360−0485]

Artificial Intelligence Group, University of Hagen, Germany
lars.bengel@fernuni-hagen.de

1 Introduction

An abstract argumentation framework (AF) due to Dung [1] is defined as a graph $F = (\mathsf{Arg}, \mathsf{R})$, where the nodes are *arguments* and the edges represent *attacks* between these arguments. An attack from an argument A to another argument B means that, if we consider the argument A to be acceptable, then we have to reject B, since A contradicts B. The goal of this approach is to model human argumentation in a formal manner and to use this model for reasoning.

A central notion in abstract argumentation is that of an *(argumentation) semantics*. A semantics σ characterizes sets of arguments (called *extensions*) that are jointly acceptable. In particular, they usually require that extensions are *conflict-free*, i. e., that there is no conflict between arguments of the extension, and that it defends itself, i.e., it counterattacks all its attackers (the latter property is called *admissibility*). Based on these notions, we can define different semantics, like e. g., complete semantics, where all arguments defended by a set E have to be contained in E.

We are especially interested in labeling-based semantics [2]. Instead of just providing sets of acceptable arguments, they determine *labelings* that assign to each argument one of three labels. If an argument is considered acceptable it is labeled in, if the argument is attacked by some acceptable argument it gets the label out and, otherwise, it receives the label undec.

In general, semantics allows us to perform *inference*, by determining semantical information (extensions or labelings) from given syntactical information (in the form of an AF). Using, e. g., the labelings of an AF we can then draw conclusions from the framework by considering the accepted arguments. We are however interested in the reverse direction, i. e., the process of *inductively learning* a syntactic structure from semantical information. This process can be considered as a form of inductive reasoning, where we generalize from observations (in the form of a given set of labelings) to a suitable AF that explains this input. This AF then also enables further reasoning possibilities, e. g., computing additional compatible labelings.

Assume a scenario, where we can discuss with a person about their beliefs on a specific subject. From this discussion, we can obtain knowledge about their beliefs, in the form of labelings. Now, to gain a better understanding of their internal reasoning, we want to learn AFs that are compatible with these labelings.

Knowing the graph structure that is consistent with a labeling is helpful in making the labeling explainable. This may also allow us to construct better counterarguments in order to persuade them to change their beliefs.

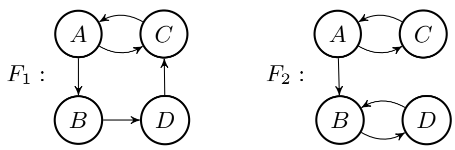


Fig. 1. Some AFs that are compatible with the complete labeling ℓ.

Example 1. Consider the complete labeling $\ell = \{\mathsf{in} : \{A, D\}, \mathsf{out} : \{B, C\}, \mathsf{undec} : \emptyset\}$. Both AFs in Figure 1 are compatible and can be used to explain ℓ. For example, F_1 would tell us that D may be accepted despite the attack $B \to D$, because A defends D against B. That also means, constructing an argument that refutes A would be very effective in challenging a person that beliefs in ℓ. This information is not immediately apparent from the labeling ℓ alone.

2 Approach

There exist few approaches that address the problem of learning AFs from labelings or similar problems [3, 4]. However, neither of them fully addresses the scenario outlined above. The main issue with these approaches is, they only compute a single solution for an input, while in reality there might be multiple compatible AFs, e. g., the AFs in Figure 1 are both compatible with ℓ. Therefore, we propose a new algorithm that, given a set of labelings and associated semantics, computes so-called *attack constraints* for each argument. For example, given an argument in a complete input labeling ℓ, the constraints are computed according to the following method.

$$AttCon_{co}(a, \ell) = \begin{cases} \bigwedge_{b \in \mathsf{Arg} \setminus \mathsf{out}(\ell)} \neg r_{ba} & \text{if } a \in \mathsf{in}(\ell) \\ \bigvee_{b \in \mathsf{in}(\ell)} r_{ba} & \text{if } a \in \mathsf{out}(\ell) \\ \bigwedge_{b \in \mathsf{in}(\ell)} \neg r_{ba} \wedge \left(\bigvee_{c \in \mathsf{undec}(\ell)} r_{ca} \right) & \text{if } a \in \mathsf{undec}(\ell) \end{cases} \tag{1}$$

The atoms of these formulas directly correspond to attacks in AFs and thus allow for an efficient representation of the set of all AFs that are compatible with the input. Moreover, with this approach, we can easily incorporate additional labelings and refine our result. Current work includes elaborating this idea and conducting an experimental study of its feasibility.

Acknowledgements The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant 375588274).

References

1. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial intelligence **77**(2), 321–357 (1995)
2. Caminada, M.W., Gabbay, D.M.: A logical account of formal argumentation. Studia Logica **93**(2), 109–145 (2009)
3. Riveret, R., Governatori, G.: On learning attacks in probabilistic abstract argumentation. In: Proceedings of the AAMAS 2016. pp. 653–661 (2016)
4. Niskanen, A., Wallner, J., Järvisalo, M.: Synthesizing argumentation frameworks from examples. Journal of Artificial Intelligence Research **66**, 503–554 (2019)

Inductive Inferences in *CL* Diagrams

Reetu Bhattacharjee[1] and Jens Lemanski[2][0000−0003−3661−4752]

[1] Classe di Lettere e Filosofia (Faculty of Humanities), Scuola Normale Superiore, Pisa, Italy reetu.bhattacharjee@sns.it
[2] Institute of Philosophy, University of Münster, Münster, Germany and Institute of Philosophy, FernUniversität in Hagen, Hagen, Germany
jens.lemanski@fernuni-hagen.de

CL diagrams – the abbreviation of *C*ubus *L*ogicus – are inspired by J.C. Lange's logic machine from 1714 [2]. In recent times, Lange's diagrams have been used for extended syllogistics, bitstring semantics, analogical reasoning and many more [3]. Recently it has been proved that *CL* diagrams can also form a logical system that is sound and complete [1].

A typical *CL* diagram consists of structural elements and content elements. Structural elements are solid boxes nested like a binary tree and dotted boxes as auxiliary lines that indicate the number of boxes still subsumed. The content elements such as arrows depict logical operations over the solid boxes. The solid boxes in Fig. 1 show 3 classes, T, H, NH, and 4 individuals, X, Y, Z, W. Detailed definitions can be found in [1].

Here we will demonstrate inductive reasoning in *CL* with a simple example of an inductive syllogism:

75% of all humans are taller than 2 feet.
Gareth is a human.

Therefore, Gareth is taller than 2 feet. $\therefore$

In Fig. 1, 'T' represents set of the heights of all the living being, 'H' the set of humans, 'NH' non-humans. The solid boxes in Fig. 1 can be divided into smaller boxes using blue dotted lines (see Fig. 2). Each blue box in Fig. 2 represents 25% of the entire set. For the convention purpose, here we take any solid box

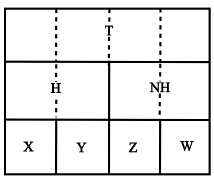


Fig. 1.

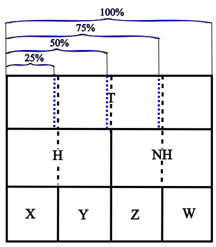


Fig. 2.

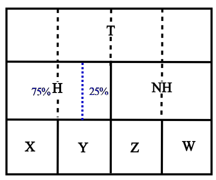


Fig. 3.

as the complete 100% and if it is divided in 4 parts then each of the part is 25%. This is not a strict convention. We can divide the boxes in any way. The only thing that we have to keep in mind is that the addition of all the divided parts of a solid box is exactly 100%. For example, in Fig. 3, the solid box H is divided into two parts.

Now let us assume that the maximum height of the living being is 16ft. So, each blue box represent the intervals as shown in Fig. 4. These blue boxes can further be divided as shown in Fig. 5. Fig. 6 represents the boxes of all heights greater than 2 feet.

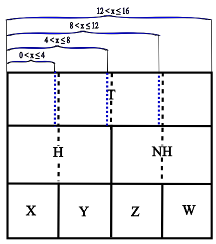


Fig. 4.

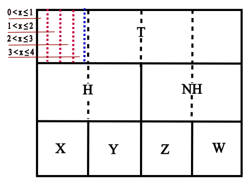


Fig. 5.

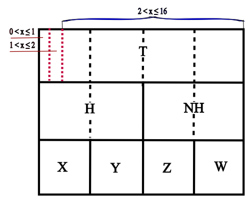


Fig. 6.

In inductive reasoning we use arrows and coloured boxes as content elements to represent propositions. The broken arrow in Fig. 7 goes from the 75% box in H (arrow shaft) to the $2 \leqslant 16$ box in T (arrowhead), indicating that 75% of all humans are taller than 2 feet. The proposition "Gareth is a human" is represented by including 'Gareth' directly under the box for humans, e.g. Fig. 8. Finally, we use the following inference rule to obtain Fig. 9 which represent that Gareth (arrow shaft) is taller than 2 feet (arrowhead).

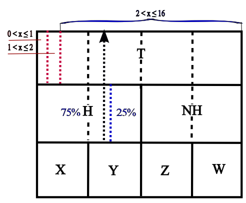


Fig. 7.

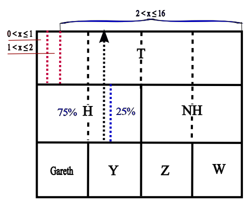


Fig. 8.

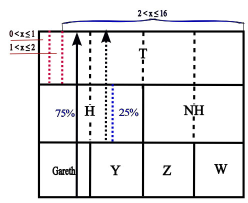


Fig. 9.

Inference Rule: If a broken arrow connects the maximum collection of smallest subboxes of a box, say A, to the maximum collection of smallest subboxes of another box B then a solid arrow can be drawn from the individual box that is included in A to the maximum collection of smallest subboxes of B.

'Broken arrow' is a representation of 'observation', whereas a 'solid arrow' is a representation of the 'conclusion' that has been infered from our observation. In inductive reasoning we arrive at some conclusion through observation. This is similar in our example: Like in the 'inference rule', we first observe the position of the broken arrow then we arrive at our conclusion that Gareth's height is more than 2 feet.

Acknowledgements This research benefited from the research grant "Quod erat videndum : heuristic, logical and cognitive aspects of diagrams in mathematical reasoning" as part of the research project MIUR - "Departments of Excellence", call 2017 - Faculty of Humanities at Scuola Normale Superiore, Pisa, Italy received by Reetu Bhattacharjee in 2022. The research of the second author was supported by the project 'History of Logic Diagrams in Kantiansm' (AZ.10.22.1.037PH) from the Fritz Thyssen-Stiftung.

References

1. Lemanski, J., Jansen, L.: Calculus *CL* as a formal system. In: Pietarinen, A.V., Chapman, P., Smet, L.B., Giardino, V., Corter, J., Linker, S. (eds.) Diagrammatic Representation and Inference. Diagrams 2020. LNCS 12169, pp. 445–460 (2020)
2. Prade, H., Marquis, P., Papini, O.: Elements for a history of artificial intelligence. In: Marquis, P., Papini, O., Prade, H. (eds.) Guided Tour of Artificial Intelligence Research. Bd. 1: Knowledge Representation, Reasoning and Learning, pp. 1–45. Springer (2020)
3. Schang, F., Lemanski, J.: A bitstring semantics for calculus *CL*. In: Vandoulakis, I. (ed.) The Exoteric Square of Opposition, pp. 163–174. Springer (2022)

Partial Information Decomposition for the Analysis of Inductive Inferences with Multiple Premises

Aaron J. Gutknecht[1][0000−0002−2704−6944], Michael Wibral[1][0000−0001−8010−5862], and Abdullah Makkeh[1][0000−0002−3581−8262]

Campus Institute for Dynamics of Biological Networks, Georg-August University Goettingen

Abstract. The theory of partial information decomposition (PID) is an extension of classical information theory that allows to describe how information about some target variable is distributed over a range of source variables. In this way PID distinguishes fundamental types of uncertainty reduction. In the simplest case of two information sources the total reduction in our uncertainty about the target may be partly redundant to both information sources, partly unique to one of them, and partly synergistic, i.e. some aspect of our uncertainty about the target may only be reduced if we have access to both information sources at the same time. The problem of disentangling these components necessarily arises in any inductive inference with more than one "premise" (information source, observation, etc.). We argue, therefore, that PID theory is of great relevance to the foundations of inductive inference in particular as it pertains to the allocation of epistemic value over multiple information sources. Our formulation of the theory draws from principles of mereology and formal logic thereby placing PID on the foundation of two of the most elementary concepts of human thought: the part-whole relationship and the relation of logical implication.

Keywords: Information Theory · Mereology · Logic · Inductive Inference.

The quest for a decomposition of information in terms of unique, redundant, and synergistic components goes back to the 1950s, but it was only in the seminal 2010 paper by Williams and Beer [6] that the mathematical structure of the problem was described in its entirety and the first full solution was proposed. Since then there has been a flurry of alternative approaches aiming to overcome some of the shortcomings of the original solution (for instance [1, 4, 2]). However, no consensus could be reached thus far. Doing so would be of wide-ranging practical significance since problems of the PID type are ubiquitous in virtually all fields of quantitative research. In neuroscience, for instance, PID pertains to the question of how stimuli are encoded in a network of multiple neurons. Do the neurons encode the stimulus synergistically? Or are differerent aspects of the stimulus encoded uniquely by different neurons? How redundant is the encoding? In machine learning one may ask how information about the output is allocated over different potential features. This analysis can then be used for efficient feature selection [7].

It should be noted that the problem soon becomes very complex as a larger number of information sources are considered because ever more complicated types of information arise (think for example of information shared by some variables yet at the same time synergistic with respect to some other variables). In order to deal with this complexity, our own work focuses, on the one hand, on the mathematical structure and conceptual foundations of PID [3], and on the other, on concrete numerical measures of the different PID components [5]. In the former line of research we show how partial information decomposition can be fully explained in terms of parthood relationships between information contributions of the source variables about the target variable. In doing so, we arrive at a unique solution to the hierarchical organization of information contributions that must be respected if the information decomposition is to embody the intuitive notion of one entity being part of another one. In the second line of research we establish a connection between PID theory and formal logic. In particular, we describe how the different PID components can be measured by the local information provided by a special class of statements about the observed values of the information sources: the class of statements with monotonic truth-tables. The defining feature of such statements is that changing the truth value of an atomic statement from false to true, cannot make the statement false if it was previously true.

Here, we would like to suggest that PID could play an important role in the analysis of inductive inferences in the sense of empirical / statistical inference from data to hypothesis. Especially from a Bayesian perspective, one may consider the target variable as representing some parameter Θ of interest and the information sources as multiple observations or experimental outcomes $X_1, \ldots, X_n$ (the "premises" of the inductive inference). Our goal is to use these observations to reduce our uncertainty about the parameter. In the most simple case of two observations, PID aims to decompose the joint mutual information into the four basic components of redundancy, uniqueness, and synergy:

$$I(X_1, X_2 : \Theta) = Red(X_1, X_2 : \Theta) + U(X_1 : \Theta) + U(X_2 : \Theta) + Syn(X_1, X_2 : \Theta) \quad (1)$$

Viewed in this light, PID distinguishes between different types of uncertainty reduction and, importantly, allows us to describe these types in a quantitative way. Is our reduction in uncertainty about Θ uniquely due to a particular observation? To what degree does it result from synergistically combining information obtained from multiple observations? Perhaps, our uncertainty reduction is even purely synergistic so that we cannot learn anything about the parameter by looking at an individual observation in isolation. Or is some part of our uncertainty reduction redundant to multiple observations? By answering all possible questions of this nature, PID provides a detailed picture of the epistemic structure of inductive inferences with multiple premises. It tells us which observations, or collections of observations, matter to what degree when it comes to reducing our uncertainty about the parameter.

References

1. Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., Ay, N.: Quantifying Unique Information. Entropy **16**(4), 2161–2183 (apr 2014). https://doi.org/10.3390/e16042161, http://www.mdpi.com/1099-4300/16/4/2161
2. Finn, C., Lizier, J.T.: Pointwise partial information decomposition using the specificity and ambiguity lattices. Entropy **20**(4), 297 (2018)
3. Gutknecht, A.J., Wibral, M., Makkeh, A.: Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic. Proceedings of the Royal Society A **477**(2251), 20210110 (2021)
4. Ince, R.A.A.: Measuring multivariate redundant information with pointwise common change in surprisal. Entropy **19**(7), 318 (2017)
5. Makkeh, A., Gutknecht, A.J., Wibral, M.: Introducing a differentiable measure of pointwise shared information. Physical Review E **103**(3), 032149 (2021)
6. Williams, P.L., Beer, R.D.: Nonnegative decomposition of multivariate information. arXiv preprint arXiv:1004.2515 (2010)
7. Wollstadt, P., Schmitt, S., Wibral, M.: A rigorous information-theoretic definition of redundancy and relevancy in feature selection based on (partial) information decomposition. arXiv preprint arXiv:2105.04187 (2021)

On Bridging the Gap Between Machine Learning and Knowledge Representation and Reasoning: The Case of Abstract Argumentation

Isabelle Kuhlmann

Artificial Intelligence Group, University of Hagen, Germany
isabelle.kuhlmann@fernuni-hagen.de

Motivation Two major fields in Artificial Intelligence are *Machine Learning* (ML) and *Knowledge Representation and Reasoning* (KRR). In ML, algorithms typically have the advantage that, once they are trained, they can yield solutions rather fast, but the disadvantage that the results are not guaranteed to be correct, and that solutions usually have no rationale or justification comprehensible to humans. By contrast, in KRR, algorithms are typically sound and complete. Hence, they yield correct, and usually "explainable" results. Consequently, a combination of ML and KRR approaches is highly promising for a plethora of problems originating from both fields. On the one hand, KRR methods can be used, e.g., to make ML approaches explainable; on the other hand, ML algorithms can be used to speed up KRR approaches. For instance, ML can be used for algorithm selection or for approximation in KRR. In the following, we will illustrate an example of the latter in order to outline both opportunities and challenges of combining ML and KRR.

Deciding Acceptability in Abstract Argumentation Using Deep Learning An *abstract argumentation framework* (AF) [2] is a tuple $F = (\mathsf{Args}, R)$, with Args being a set of *arguments* and $R \subseteq \mathsf{Args} \times \mathsf{Args}$ representing an *attack relation* between such arguments. AFs can be viewed as directed graphs, where the nodes represent Args and the edges R. In Abstract Argumentation (AA), one is usually interested in identifying *extensions*, i.e., sets of arguments that are jointly acceptable. Which constraints a set has to satisfy in order to be considered an extension, depends on the *semantics* used in the reasoning process. Typical problems in AA are deciding whether an argument is credulously accepted (DC) or skeptically accepted (DS), i.e., whether it is contained in at least one, or in all extensions of a given semantics, respectively. Most algorithms for solving such problems are sound and complete (see [6] for a recent overview). However, these problems exhibit a high complexity—e.g., deciding DS for *preferred* semantics is Π_2^P-complete [3]—which hinders a scalable behavior. Therefore, some authors suggest to use deep learning approaches [5, 1, 7] to compute solutions which are only approximate on the one hand, but fast to obtain on the other hand.

The objective of an ML system is to "learn" from given data (*training set*) and to use the acquired "knowledge" to make predictions about previously unknown data. In a feasibility study, Kuhlmann and Thimm [5] trained a *Graph*

Convolutional Network (GCN) [4] on a set of AFs. Each argument in the training set had a label marking it as either "accepted" (a) or "not accepted" (na). The authors demonstrated that a GCN is able to correctly classify arguments as a or na to a certain degree, but also pointed out some issues. The classes are often unevenly distributed, as there are usually more unaccepted than accepted arguments in an AF. This issue is generally not new in ML; there exist, e.g., a number of *augmentation* techniques which counteract this problem. However, augmentation methods for other graph data (e.g., citation or social networks), cannot simply be adopted for AFs, as they mostly consist of adding/deleting arguments or entire sub-graphs, or of manipulating node features. Since arguments do not possess any features, and modifying the graph topology could change the arguments' acceptability status, we cannot use such methods. Malmqvist et al. [7] address this problem by introducing a scheme to dynamically balance the training data, as well as a randomized training regime. Further, Craandijk and Bex [1] propose an *Argumentation Graph Neural Network* (AGNN) which learns a message-passing algorithm.

Even though some of the results (in particular those in [1]) are quite promising, the existing works on this topic are difficult to compare, since they all use different datasets—which is most likely due to the lack of a standard dataset for such purposes. Hence, data selection poses an additional challenge in combining ML and argumentation. Another problem of applying existing techniques, such as GCNs, to AA, lies in the fact that they are based on the assumption that closely connected data points are similar and are thus likely to belong to the same class. Although this is true for most graph-structured data (again, such as citation or social networks), it is explicitly not true for AFs: If one argument attacks another, they are direct neighbors, but they cannot both be accepted.

Conclusion and Future Work In the scope of this work, we discussed an example of combining ML (specifically, neural networks) with KRR (specifically, AFs). We explained that there are already quite promising results regarding this problem, but we also identified challenges. Examples of future work include the need for an appropriate (standard) dataset, as well as some learning approaches which are specifically designed with the "adversarial" nature of AFs in mind.

References

1. Craandijk, D., Bex, F.: Deep learning for abstract argumentation semantics. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. p. 1667–1673 (2020)
2. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77**(2), 321–358 (1995)
3. Dvorák, W., Dunne, P.E.: Computational problems in formal argumentation and their complexity. Handbook of formal argumentation pp. 631–687 (2018)
4. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)

5. Kuhlmann, I., Thimm, M.: Using graph convolutional networks for approximate reasoning with abstract argumentation frameworks: A feasibility study. In: International Conference on Scalable Uncertainty Management. pp. 24–37. Springer (2019)
6. Lagniez, J.M., Lonca, E., Mailly, J.G., Rossit, J.: Design and results of ICCMA 2021. arXiv preprint arXiv:2109.08884 (2021)
7. Malmqvist, L., Yuan, T., Nightingale, P., Manandhar, S.: Determining the acceptability of abstract arguments with graph convolutional networks. In: SAFA@ COMMA. pp. 47–56 (2020)

Solving the problems of plurality of causes and of intermixture of effects in Mill's canons of induction

Wolfgang Pietsch[1]

Technische Universität München, Germany wolfgang.pietsch@tum.de

Abstract. A novel solution to the problems of plurality of causes and of intermixture of effects in John Stuart Mill's canons of induction is proposed, which is based on two fundamental ideas. First, causal dependencies are defined in terms of causal relevance and irrelevance rather than in terms of necessary and sufficient conditions. Second, causal statements are relativized to dy-namically changing background conditions rather than a static causal field. The proposed solution improves on existing approaches in terms of both simplicity and adequacy for scientific practice.

Keywords: Mill's canons of induction · method of difference · INUS condition.

Mill's canons of induction (1843) remain one of the most influential inductive frameworks. Perhaps the crucial objection against these inductive methods is that they are not capable of dealing with certain widespread causal structures, in particular the *plurality of causes* and the *intermixture of effects*. Essentially, in a complex cause of the form $(A_1 \wedge A_2) \vee A_3$, the '$\vee A_3$'-part addresses the plurality of causes, while the '$\wedge A_2$'-part addresses the intermixture of effects. Mill believed that his inductive framework is mostly unable to deal with these issues.

The main attempts in the twentieth century to improve on Mill's canons of induction are due to Georg Henrik von Wright (1951), John Mackie (1980, appendix), and Brian Skyrms (2000, Ch. 5). These authors suggest a novel systematization of Mill's canons, which differs substantially from Mill's four (or five) methods. In particular, they attempt to solve the problems of plurality of causes and of intermixture of effects by delineating a range of methods tailored for specific types of causes. For example, Mackie proposes a two-dimensional system of different inductive methods, where the first number designates the allowed complexity of potential causes and the second number the type of method that is employed, e.g. method of difference or of agreement. However, the resulting systematization turns out to be rather complex. Furthermore, the approach does not align well with scientific practice. In particular, in typical experimental procedures, assumptions about the complexity of potential causes are rarely presupposed.

In the following, a solution to the mentioned problems is sketched which deviates in two crucial aspects. First, the proposed solution is based on causal relevance and causal irrelevance as fundamental concepts rather than on necessary and sufficient conditions. Second, causal statements are relativized to dynamically changing background conditions rather than a static causal field as in Mackie's approach.

Without being able to provide full detail in the limited space available here, essentially, 'causal relevance' of a condition A to a phenomenon C with respect to a given background B means: provided that the background conditions B are instantiated, then whenever A is present, C is present, and whenever A is absent, C is absent. Similarly, 'causal irrelevance' of a condition A to a phenomenon C with respect to a given background B means: provided that the background conditions B are instantiated, then C is always present, no matter if A is present or absent (for more precise formulations and a much more thorough analysis, see Pietsch 2022, Ch. 5).

The problem of *intermixture of effects* is tackled by defining: A is a 'causal factor' for a phenomenon C with respect to a background B, if and only if there exists an X such that A is causally relevant to C with respect to a background $B \wedge X$ and irrelevant to $\not C$ with respect to a background $B \wedge \not X$. For example, if a match and combustible material only in combination cause a fire, then the match is causally relevant to the fire with respect to a background, in which combustible material is present. By contrast, the match is causally irrelevant, when combustible material is absent, as the match alone cannot start the fire. Thus, the match is a causal factor for the fire according to the above definition.

The problem of *plurality of causes* is addressed by defining: A is an 'alternative cause' to C with respect to a background B, if and only if there exists an X such that A is causally relevant to C with respect to a background $B \wedge \not X$, but causally irrelevant to C with respect to a background $B \wedge X$. For example, if both lightning and an explosion are independent causes of a fire, then lightning is causally relevant to the fire as long as there is no explosion. By contrast, in the case of an explosion, lightning is causally irrelevant, since the explosion already causes the fire independent of the presence or absence of lightning. Thus, lightning is an alternative cause.

The above solution to the problems of plurality of causes and of intermixture of effects is simpler than the proposals by von Wright, Mackie and Skyrms. In particular, two fundamental methods suffice, one for determining causal relevance, the other for determining causal irrelevance. More complex causal dependencies can be defined based solely on these two fundamental methods. Furthermore, there is no need for any assumptions regarding the complexity of potential causes. Thus, the proposed solution more closely aligns with scientific practice, e.g. experimental procedures, but also machine learning methods (Pietsch 2021).

References

1. Mackie, J.L.: The Cement of the Universe. Clarendon Press, Oxford (1980)

2. Mill, J. S.: A System of Logic, Ratiocinative and Inductive. Longmans, Green & Co, London (1843)
3. Pietsch, W.: Big Data. Cambridge University Press, Cambridge (2021)
4. Pietsch, W.: On the Epistemology of Data Science. Conceptual Tools for a New Inductivism. Philosophical Studies Series, vol 148. Cham, Springer (2022)
5. Skyrms, B.: Choice and Chance: An Introduction to Inductive Logic. Wadsworth/Thomson Learning, Belmont (2000)
6. von Wright, G.H.: A Treatise on Induction and Probability. Routledge, New York (1951)

Towards Realisability of Rankings-based Semantics*

Kenneth Skiba[1] [0000−0003−1250−8920]

Artificial Intelligence Group, University of Hagen, Germany
kenneth.skiba@fernuni-hagen.de

In recent years, *abstract argumentation frameworks* (AF) [4] have gathered research interest as a model for argumentative reasoning. They are a model for rational decision-making in presence of conflicting information. *Arguments* and *attacks* are represented as nodes and edges, respectively, of a directed graph, i.e. an argument a *attacking* argument b is represented as a directed edge from a to b. In a scenario of strategic argumentation, an agent wants to persuade an opponent. One way to find a persuasion strategy is by considering the strength of each argument, since stronger arguments have a higher chance to persuade the opponent. Hence, *ranked-based semantics* were introduced (see [3, 1] for an overview). These functions define a preorder based on the acceptability degree of each argument s.t. we can state that an argument a is "stronger" than an argument b.

Consider a scenario where an agent observes other agents discussing. Based on this observation she prepares arguments and a strength assessment of each argument. However, the agent has no knowledge about the underlying AF. So, the underlying AF has to be established first.

In this work, we discuss the question whether there exists an AF with the observed strength assessment. So, for a given ranking-based semantics ρ and a preorder r can we find an AF, which exactly has r as its ranking when applying ρ? This problem is known as *realisability* and was already investigated for *extension semantics*, which specify when a set of arguments is considered jointly acceptable. Dunne et.al. [5] have shown that there are sets of arguments for which we cannot find an AF s.t. these sets are considered jointly acceptable.

Example 1. Assume the ranking $a \succ b \succ c \succ d$ and the ranking-based semantics *Categoriser* defined by [2]. So, a is preferred over b, b over c, and d is the weakest argument. The Categoriser semantics is defined via a ranking $\succeq^{Cat}$ on A s.t. for $a, b \in A$, $a \succeq^{Cat} b$ holds iff $Cat(a) \geq Cat(b)$, and $Cat : A \to]0, 1]$ is defined with $Cat(a) = \frac{1}{1+\sum_{b \in a^-} Cat(b)}$. Where A is the set of arguments, R the set of attacks between two arguments, and a^- the set of arguments attacking a. We want to find an AF s.t. the Categoriser semantics returns $a \succ b \succ c \succ d$ as the corresponding ranking. One such an AF would be $AF_r = \{\{a, b, c, d\},\ \{(a, b), (a, c), (a, d), (d, b), (d, c), (b, c)\}\}$ as depicted in Figure 1.

* The research reported here was supported by the Deutsche Forschungsgemeinschaft under grant 423456621.

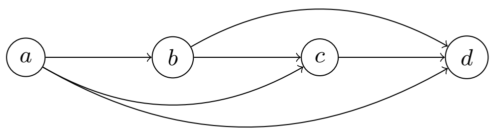


Fig. 1. AF_r constructed in Example 1.

In Example 1, we have seen that, we can find an AF for a given ranking based on the Categoriser semantics. Now, we can ask whether it is always possible to find such an AF. Formally, for a given ranking r, can we find an AF s.t. this AF has r as its corresponding ranking, when applying the Categoriser semantics? Indeed, it is possible, consider the following construction: Given a ranking r based on the Categoriser semantics, we construct an $AF_r = (A, R)$ in the following way: Every argument a appearing in r is part of A; If $a \succ^{Cat} b$, then $(a, b) \in R$. So, if an argument a is ranked better than argument b, then a attacks b. An example use of the construction can be found in Example 1. We can show that the Categoriser semantics applied to AF_r will always return r. Therefore, for every ranking based on the Categoriser semantics we can find an AF.

What does these results mean for our agent? If she uses the Categoriser semantics as her reasoning formalism, then she can always find an AF with her strength assessment. We can construct other AFs with the same ranking based on the Categoriser semantics. Based on this observation we can define an *equivalence* notion. Two AFs are *ρ-equivalent* iff applying ρ returns the same ranking for both AFs.

References

1. Amgoud, L., Ben-Naim, J.: Ranking-based semantics for argumentation frameworks. In: Proc. SUM'13 (2013)
2. Besnard, P., Hunter, A.: A logic-based theory of deductive arguments. Artif. Intell. (2001)
3. Bonzon, E., Delobelle, J., Konieczny, S., Maudet, N.: A comparative study of ranking-based semantics for abstract argumentation. In: Proc. of AAAI'16 (2016)
4. Dung, P.M.: On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. Artificial Intelligence (1995)
5. Dunne, P.E., Dvořák, W., Linsbichler, T., Woltran, S.: Characteristics of multiple viewpoints in abstract argumentation. Artificial Intelligence **228**, 153–178 (2015)

Neurosymbolic Learning: On Generalising Relational Visuospatial and Temporal Structure

Jakob Suchan[1] and Mehul Bhatt[2]

[1] German Aerospace Center (DLR), Germany
[2] Örebro University, Sweden

CoDesign Lab – Cognition. AI. Interaction. Design.
info@codesign-lab.org – https://codesign-lab.org

Abstract. We position ongoing research aimed at developing a general framework for structured spatio-temporal learning from multimodal human behavioural stimuli. The framework and its underlying general, modular methods serve as a model for the application of integrated (neural) visuo-auditory processing and (semantic) relational learning foundations for applications (primarily) in the behavioural sciences.

Keywords: Representation learning and grounding · Relational learning · Cognitive vision and perception · Multimodality · Dynamic Visuospatial Imagery

High-level perceptual sensemaking of multimodal human behavioural stimuli is foundational to diverse cognitive assistive technologies and autonomous perception & interaction systems [3, 4]. Multimodal sensemaking, at a level of descriptive and analytical complexity that matches cognitive human performance and expectations, is also crucial for the development of next-generation AI technologies and artefacts —concerned with agency, assistance and autonomy— where human-centred considerations of personalisation, normative behaviour, explanation, empathy, trust, responsibility are at the core.

Multimodal Learning: A Neurosymbolic Foundation. In this position statement, we summarise aspects of ongoing research in 'Cognitive Vision and Perception' [4] aimed at developing a general framework for structured spatio-temporal learning from multimodal human behavioural stimuli, e.g., consisting of dynamic visuospatial and auditory features. With an emphasis on formalisations of spatio-linguistically rooted relations of *space, time and motion*, the framework is geared towards supporting high-level learning of *deep semantic* relational spatio-temporal structure —by means of inductive generalisation— from low-level stimuli (typically) emanating from embodied human interactions in everyday naturalistic settings. At the crux of the proposed framework are general and modularly developed foundational spatio-temporal learning methods intended to serve as a neurosymbolic model for with integrated (neural learning based) visuo-auditory processing and (semantics based) relational learning synergistically serving as a foundational backbone in diverse applications such

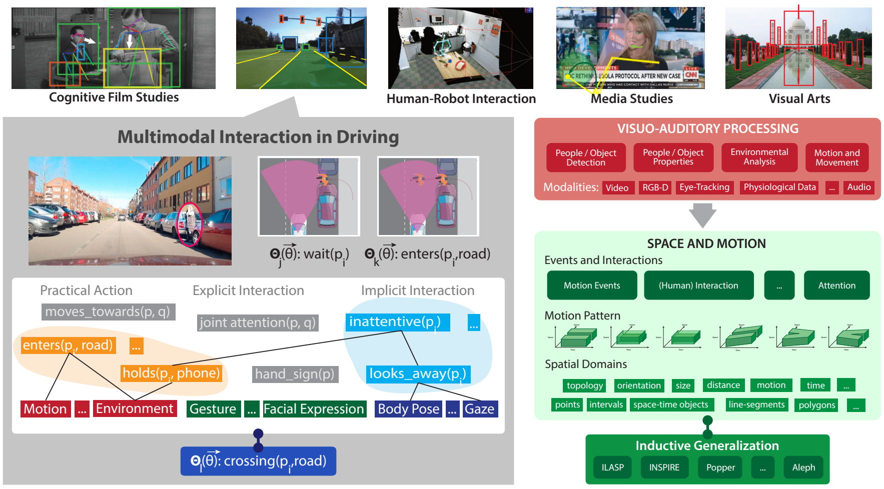


Fig. 1. Relational Spatio-Temporal Structure of Embodied Multimodal Interaction

as behavioural research in psychology (e.g., visual perception), studies in multimodal interaction, HRI and social robotics.

Relational Space-Time Generalisation: A Case for Commonsense Relational learning by inductive generalisation in the context of Aritificial Intelligence (AI) and Machine Learning (ML) is a well-established area of research. Beyond the specific context of Al and ML, the topics of knowledge discovery, explanatory reasoning, hypothesis formation, and decision making assume a far broader significance from philosophical, logical, and cognitive perspectives.

Our approach to relational generalisation from multimodal observations —in so far as this position statement is concerned— is driven by inductive learning based on a logical / knowledge representation and reasoning approach under *constraint logic* and *answer set* learning settings. At the core of the multimodal learning framework are commonsense characterisations of space and motion primitives suited for the grounding of embodied interaction specific multimodal interactional features pertaining to people, objects, gaze, body pose, visual fixation measured via eye-tracking, speech / auditory features such as those relevant to intonation (Fig. 1). For relational (space-time) learning [11], in scope are systems such as ILASP [8], Inspire [9], Popper [5], ALEPH [10]. From an applied viewpoint, the ongoing research is motivated by demonstrating the significance and value of (inductive) generalisation as a means to learn the relational spatio-temporal structure underlying multimodal data pertaining to embodied human interactions in diverse empirical research contexts where the ability to induce high-level, semantic, declaratively explainable behaviour models is of interest. In essence, the learnt behavioural models pertain to some aspect of everyday human activity and interaction.

Acknowledgments

We acknowledge partial funding by the *Federal Ministry for Economic Affairs and Climate Action* (BMWK) as part of the *SituWare* project (reference no. 19A19011C).

Bibliography

[1] Bhatt, M.: Reasoning about space, actions and change: A paradigm for applications of spatial reasoning. In: Qualitative Spatial Representation and Reasoning: Trends and Future Directions. IGI Global, USA (2012)

[2] Bhatt, M., Guesgen, H.W., Wölfl, S., Hazarika, S.M.: Qualitative spatial and temporal reasoning: Emerging applications, trends, and directions. Spatial Cogn. Comput. **11**(1), 1–14 (2011). https://doi.org/10.1080/13875868.2010.548568, https://doi.org/10.1080/13875868.2010.548568

[3] Bhatt, M., Kersting, K.: Semantic interpretation of multi-modal human-behaviour data - making sense of events, activities, processes. Künstliche Intell. / Artificial Intelligence **31**(4), 317–320 (2017)

[4] Bhatt, M., Suchan, J.: Cognitive vision and perception. In: ECAI 2020 - 24th European Conference on Artificial Intelligence. Frontiers in Artificial Intelligence and Applications, vol. 325, pp. 2881–2882. IOS Press (2020)

[5] Cropper, A., Morel, R.: Learning programs by learning from failures. Mach. Learn. **110**(4), 801–856 (2021)

[6] Dubba, K.S.R., Cohn, A.G., Hogg, D.C., Bhatt, M., Dylla, F.: Learning relational event models from video. J. Artif. Intell. Res. **53**, 41–90 (2015)

[7] Kondyli, V., Suchan, J., Bhatt, M.: Grounding Embodied Multimodal Interaction: Towards Behaviourally Established Semantic Foundations for Human-Centered AI. In: First International Workshop on Knowledge Representation for Hybrid Intelligence (KR4HI 2022)., part of International Conference on Hybrid Human-Artificial Intelligence (HHAI 2022), Amsterdam, The Netherlands (2022)

[8] Law, M., Russo, A., Broda, K.: Inductive learning of answer set programs. In: Fermé, E., Leite, J. (eds.) Logics in Artificial Intelligence - 14th European Conference, JELIA 2014, Funchal, Madeira, Portugal, September 24-26, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8761, pp. 311–325. Springer (2014)

[9] Schüller, P., Benz, M.: Best-effort inductive logic programming via fine-grained cost-based hypothesis generation - the inspire system at the inductive logic programming competition. Mach. Learn. **107**(7), 1141–1169 (2018)

[10] Srinivasan, A.: The Aleph Manual (2001), http://web.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/

[11] Suchan, J., Bhatt, M., Schultz, C.: Deeply semantic inductive spatio-temporal learning. In: Cussens, J., Russo, A. (eds.) Proceedings of the 26th International Conference on Inductive Logic Programming (Short papers), London, UK, 2016. CEUR Workshop Proceedings, vol. 1865, pp. 73–80. CEUR-WS.org (2016)