Trustworthy Learning and Reasoning in Complex Domains

Federico Cerutti — federico.cerutti@unibs.it

Augmenting human sensemaking abilities
to achieve causal insights and foresights

(a.k.a. *situational understanding*)

Overture.	A brief historical case.
Act I.	On conjectures, refutations, and argumentation.
Act II.	There is no certain datum in the world.
Act III.	Interesting problems are complex.
Epilogue.	

expre... the use of the intellectual... to learn: the use of the intellect; to be informed by another; to learn. prehend; to be informed by another; to learn. **understanding**, un-dėr-stand′ing, *a.* Intelligent; knowing; skilful.—*n.* The act of one who understands; comprehension; apprehension; discernment; knowledge; clear insight; the faculty or power by which one understands; the faculty of the human mind otherwise known as the intellect; the power of thinking and reasoning; intelligence; agreement of two or more persons; agreement between two or more persons; anything mutually understood or agreed upon.

understate, un-dėr-stāt′, *v.t.* To state too low; to state or represent less strongly than the truth will bear.

understatement, un-dėr-stāt′ment, *n.* The act of understating; a statement under...

determine
not certai
undeter
not restra
undevia
viating;
ciple, or
undiges
by the st
arranged
undignl
fied; sho
undilut
or mixe
any adm
undine
A water

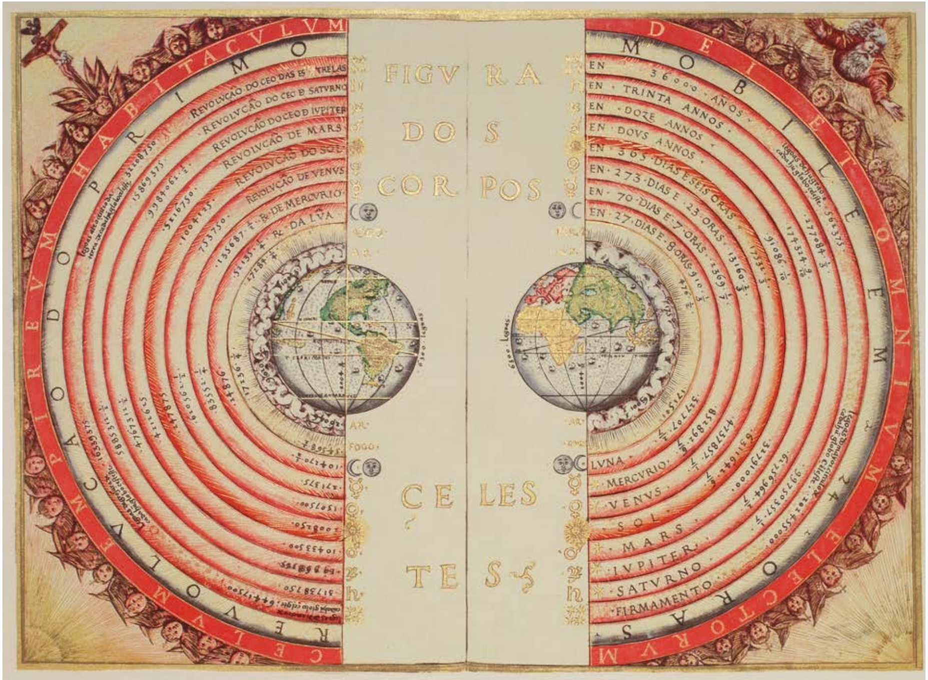


Image: Wikipedia

Empiricism

All hypotheses and theories must be tested against observations of the natural world, rather than resting solely on a priori reasoning, intuition, or revelation.

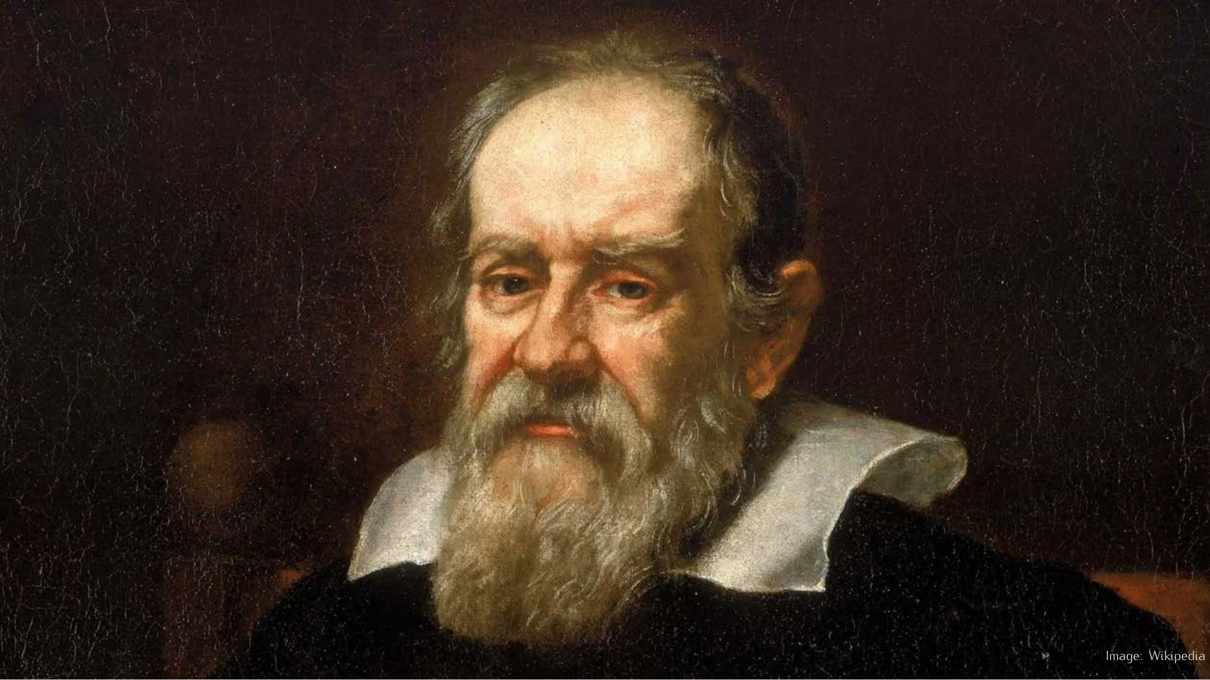
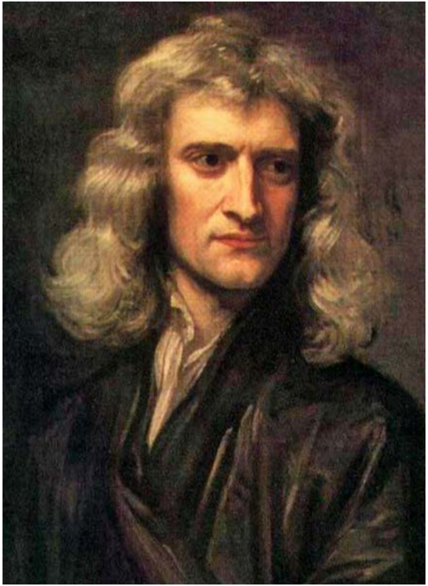
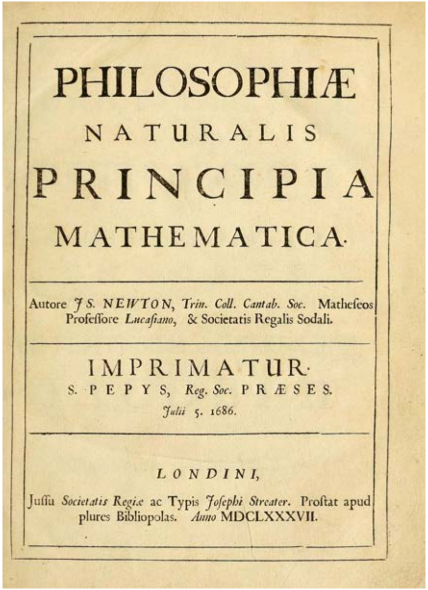


Image: Wikipedia

Image: Wikipedia





PHILOSOPHIÆ
NATURALIS
PRINCIPIA
MATHEMATICA.

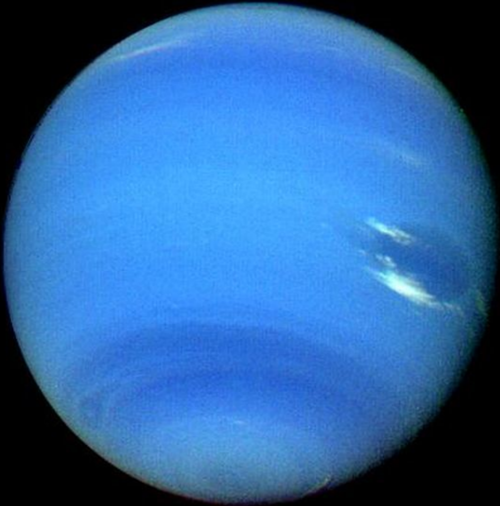
Autore *JS. NEWTON*, *Trin. Coll. Cantab. Soc.* Mathefeos Profeffore *Lucafiano*, & Societatis Regalis Sodali.

IMPRIMATUR.
S. PEPYS, *Reg. Soc.* PRÆSES.
Julii 5. 1686.

LONDINI,

Juffu *Societatis Regiæ* ac Typis *Jofephi Streater*. Proftat apud plures Bibliopolas. *Anno* MDCLXXXVII.

Image: Wikipedia



https://www.jpl.nasa.gov/spaceimages/details.php?id=PIA02210

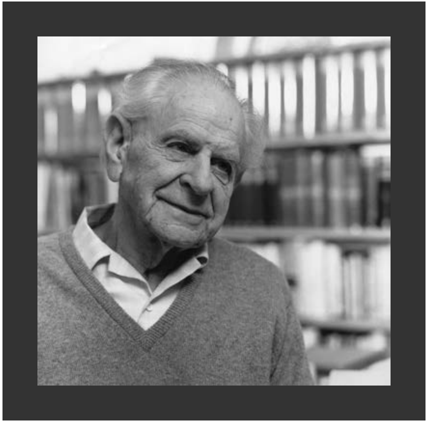

Image: Wikipedia

The path of the planet Uranus did not conform to the path predicted by Newton's law of gravitation in presence of the known planets.

Explanations:

- Human/instrument measure error
- Newton's laws are mistaken
- An invisible magic teapot caused the perturbation in order to show the *hubris* of modern science
- ...
- Newton's laws—confirmed by a significant amount of evidence—are correct and the perturbation is caused by another, unknown, planet

Image: Wikipedia



Scientific theories are capable of being refuted: they are falsifiable

Verification and falsification are different processes:

- No accumulation of confirming instances is sufficient
- Only one contradicting instance suffices to refute a theory

Scientific theories are tentative

Image: Wikipedia

Overture.	A brief historical case.
Act I.	On conjectures, refutations, and argumentation.
Act II.	There is no certain datum in the world.
Act III.	Interesting problems are complex.
Epilogue.	

Does MMR vaccination cause autism?

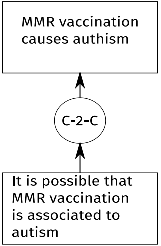
Argument from Correlation to Cause

Correlation Premise: There is a positive correlation between A and B.

Conclusion: A causes B.

- **CQ1:** Is there really a correlation between A and B?
- **CQ2:** Is there any reason to think that the correlation is any more than a coincidence?
- **CQ3:** Could there be some third factor, C, that is causing both A and B?

Walton, Reed, Macagno, Argumentation Schemes, CUP, 2008


MMR vaccination
causes authism
C-2-C
It is possible that
MMR vaccination
is associated to
autism

Early report

Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

Summary

investigated a

Introduction

saw several children who, after a

Findings Onset of behavioural symptoms was associated, by the parents, with measles, mumps, and rubella vaccination in eight of the 12 children, with measles infection in one child, and otitis media in another. All 12 children had intestinal abnormalities, ranging from lymphoid nodular hyperplasia to aphthoid ulceration. Histology showed patchy chronic inflammation in the colon in 11 children and reactive ileal lymphoid hyperplasia in seven, but no granulomas. Behavioural disorders included autism (nine), disintegrative psychosis (one), and possible postviral or vaccinal encephalitis (two). There were no focal neurological abnormalities and MRI and EEG tests were normal. Abnormal laboratory results were significantly raised urinary methylmalonic acid compared with age-matched controls (p=0·003), low haemoglobin in four children, and a low serum IgA in four children.

Support

Child	Behavioural diagnosis	Exposure identified by parents or doctor	Interval from exposure to first behavioural symptom	Features associated with exposure	Age at onset of first symptom	
					Behaviour	Bowel
1	Autism	MMR	1 week	Fever/delirium	12 months	Not known
2	Autism	MMR	2 weeks	Self injury	13 months	20 months
3	Autism	MMR	48 h	Rash and fever	14 months	Not known
4	Autism? Disintegrative disorder?	MMR	Measles vaccine at 15 months followed by slowing in development. Dramatic deterioration in behaviour immediately after MMR at 4·5 years	Repetitive behaviour, self injury, loss of self-help	4·5 years	18 months
5	Autism	None—MMR at 16 months	Self-injurious behaviour started at 18 months		4 years	
6	Autism	MMR	1 week	Rash & convulsion; gaze avoidance & self injury	15 months	18 months
7	Autism	MMR	24 h	Convulsion, gaze avoidance	21 months	2 years
8	Post-vaccinial encephalitis?	MMR	2 weeks	Fever, convulsion, rash & diarrhoea	19 months	19 months
9	Autistic spectrum disorder	Recurrent otitis media	1 week (MMR 2 months previously)	Disinterest; lack of play	18 months	2·5 years
10	Post-viral encephalitis?	Measles (previously vaccinated with MMR)	24 h	Fever, rash & vomiting	15 months	Not known
11	Autism	MMR	1 week	Recurrent "viral pneumonia" for 8 weeks following MMR	15 months	Not known
12	Autism	None—MMR at 15 months	Loss of speech development and deterioration in language skills noted at 16 months			Not known

MMR=measles, mumps, and rubella vaccine.

Table 2: **Neuropsychiatric diagnosis**

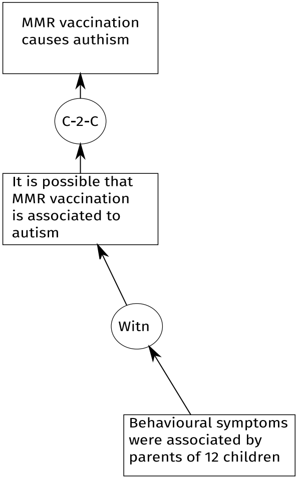
What else should be true if the causal link is true?

with autoimmune hepatitis.

We did not prove an association between measles, mumps, and rubella vaccine and the syndrome described. Virological studies are underway that may help to resolve this issue.

If there is a causal link between measles, mumps, and rubella vaccine and this syndrome, a rising incidence might be anticipated after the introduction of this vaccine in the UK in 1988. Published evidence is inadequate to show whether there is a change in incidence or a link with measles, mumps, and rubella vaccine. A genetic predisposition to autistic-spectrum disorders is suggested by over-representation in boys and a greater concordance rate in monozygotic than in dizygotic twins. In the context of susceptibility to infection, a genetic association

From *Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children* by Wakefield et al, The Lancet, 1998


MMR vaccination
causes authism
C-2-C
It is possible that
MMR vaccination
is associated to
autism
Witn
Behavioural symptoms
were associated by
parents of 12 children

The New England Journal of Medicine

VOLUME 347 | NOVEMBER 7, 2002 | NUMBER 19

A POPULATION-BASED STUDY OF MEASLES, MUMPS, AND RUBELLA VACCINATION AND AUTISM

KREESTEN MELDGAARD MADSEN, M.D., ANDERS HVIID, M.SC., MOGENS VESTERGAARD, M.D., DIANA SCHENDEL, PH.D., JAN WOHLFAHRT, M.SC., POUL THORSEN, M.D., JØRN OLSEN, M.D., AND MADS MELBYE, M.D.

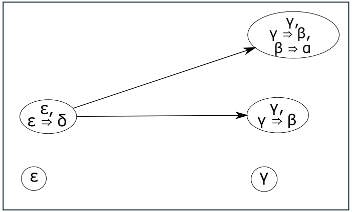
There was no association between the age at the time of vaccination, the time since vaccination, or the date of vaccination and the development of autistic disorder.

Conclusions This study provides strong evidence against the hypothesis that MMR vaccination causes autism. (N Engl J Med 2002;347:1477-82.)

Support

Results Of the 537,303 children in the cohort (representing 2,129,864 person-years), 440,655 (82.0 percent) had received the MMR vaccine. We identified 316 children with a diagnosis of autistic disorder and 422 with a diagnosis of other autistic-spectrum disorders. After adjustment for potential confounders, the relative risk of autistic disorder in the group of vaccinated children, as compared with the unvaccinated group, was 0.92 (95 percent confidence interval, 0.68 to 1.24), and the relative risk of another autistic-spectrum disorder was 0.83 (95 percent confidence interval, 0.65 to 1.07).

From *A Population-based Study of Measles, Mumps, and Rubella Vaccination and Autism* by Madsen et al, The New England Journal of Medicine, 2002

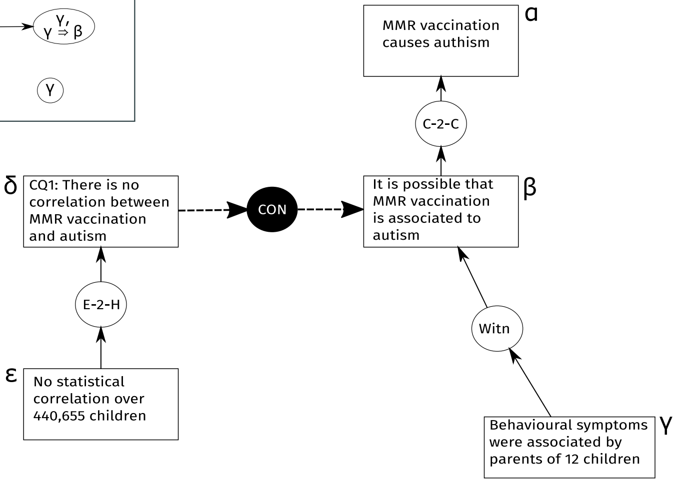


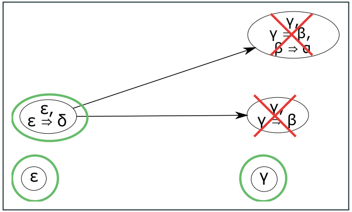
$$\beta \implies \alpha$$

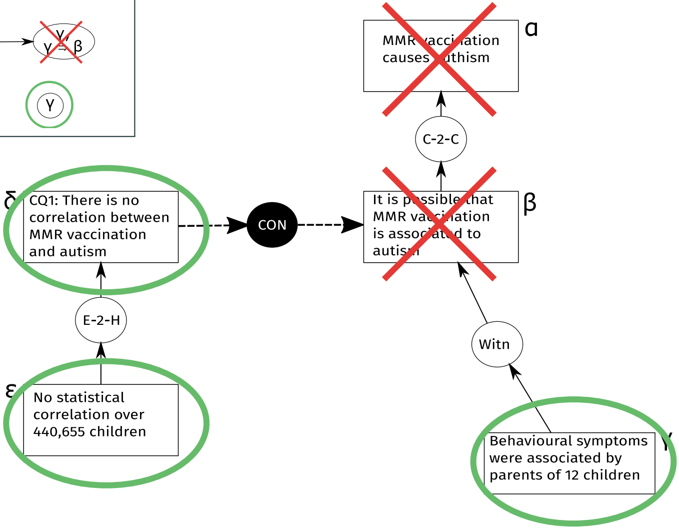
$$\gamma \implies \beta$$

$$\epsilon \implies \delta$$

$$\delta \in \overline{\beta}$$







$$\beta \implies \alpha$$

$$\gamma \implies \beta$$

$$\epsilon \implies \delta$$

$$\delta \in \overline{\beta}$$

Results (tiny summary)

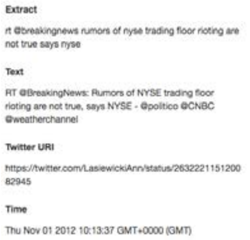
HCI Assessment of argumentation semantics against human intuition (ECAI 2014)

Algorithms Efficient algorithms and ensemble approaches (KR 2014, AAAI 2015, ECAI 2016, KER 2018, IJAR 2018, AIJ 2019, IJCAI 2021)

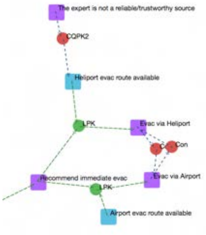
Impact Implementation in the CISpaces.org online system (AAMAS 2015, SPIE 2018, COMMA 2018, JURIX 2018, AI^3 2021)

CISpaces.org

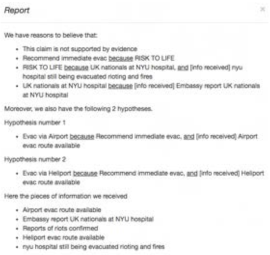
Fact extraction from Twitter



Argumentation graph manipulation



Natural Language Generation for Automatic Reporting



Available for use by professional analysts in the US Army Research Laboratory, and the UK Joint Forces Intelligence Group

TRL4: validation in a laboratory environment

`https://tiresia.unibs.it/cispaces`

F. Cerutti, T. J. Norman, A. Toniolo, and S. E. Middleton. CISpaces.org: from Fact Extraction to Report Generation. COMMA 2018, 269–281, 2018.

Overture.	A brief historical case.
Act I.	On conjectures, refutations, and argumentation.
Act II.	There is no certain datum in the world.
Act III.	Interesting problems are complex.
Epilogue.	

Qualification problem

> *For example, the successful use of a boat to cross a river requires, if the boat is a rowboat, that the oars and rowlocks be present and unbroken, and that they fit each other. Many other qualifications can be added, making the rules for using a rowboat almost impossible to apply, and yet anyone will still be able to think of additional requirements not yet stated.*

J. McCarthy, “Circumscription—A Form of Nonmonotonic Reasoning,” AIJ, 13 (12): 2739, 1980.

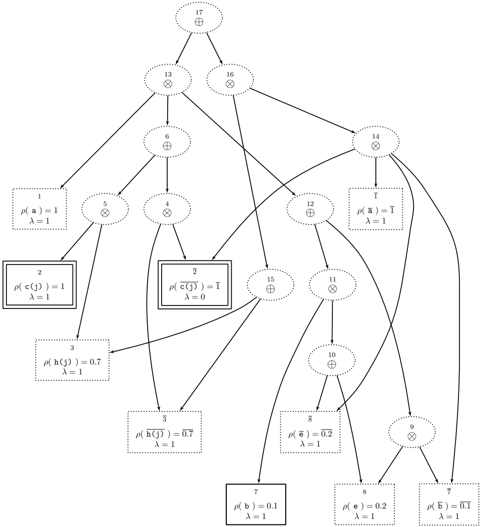
Uncertainty

Reliability of the Source	
A	Completely reliable
B	Usually reliable
C	Fairly reliable
D	Not usually reliable
E	Unreliable
F	Reliability cannot be judged

Credibility of the Information	
1	Confirmed by other sources
2	Probably true
3	Possibly true
4	Doubtful
5	Improbable
6	Truth cannot be judged

```
0.1::burglary.
0.2::earthquake.
0.7::hears_alarm(john).
alarm :- burglary.
alarm :- earthquake.
calls(john) :- alarm, hears_alarm(john).
evidence(calls(john)).
query(burglary).
```

$$\mathbf{alarm} \leftrightarrow \mathbf{burglary} \vee \mathbf{earthquake}$$
$$\mathbf{calls(john)} \leftrightarrow \mathbf{alarm} \wedge \mathbf{hears\_alarm(john)}$$
$$\mathbf{calls(john)}$$



Where numbers come from?

# Day	Earthquake
1	T
2	T
3	F
4	F
5	F
6	F
7	F
8	F
9	F
10	F

π: true—unknown—probability of earthquake in a given period of time

Let y be the number of occurrence of earthquake per period of time (y = 2)

From Bayes' theorem, we can estimate the posterior distribution of π given the data on the basis of a prior: $g(\pi|y) \propto g(\pi) \cdot f(y|\pi)$

The conjugate of a binomial is the Beta distribution. If:

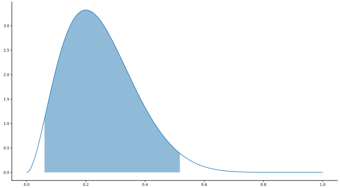
$$g(\pi; a, b) = Beta(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)+\Gamma(b)} \pi^{a-1}(1-\pi)^{b-1}$$

then: $g(\pi|y) = Beta(y+a, n-y+b)$

If $a = b = 1$ (uniform prior), then $\boxed{g(\pi|y) = Beta(y+1, n-y+1)}$

In the example, $g(\pi|y = 2, n = 10) = Beta(3, 9)$

$X_1 \sim Beta(3, 9)$



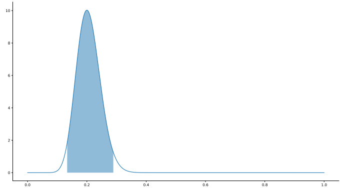
$E[X_1] = 0.2500$

$Var(X_1) = 1.4423 \cdot 10^{-2}$

95% Confidence Interval:

[0.0602, 0.5178]

$X_2 \sim Beta(21, 81)$



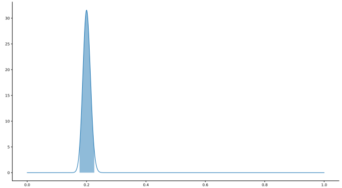
$E[X_2] = 0.2059$

$Var(X_2) = 1.5873 \cdot 10^{-3}$

95% Confidence Interval:

[0.1336, 0.2891]

$X_3 \sim Beta(201, 801)$



$E[X_3] = 0.2006$

$Var(X_3) = 1.5988 \cdot 10^{-4}$

95% Confidence Interval:

[0.1764, 0.2259]

Although $E[X_1] \simeq E[X_2] \simeq E[X_3] \simeq 0.2$

they represent remarkably different random variables

* Y-axes of the graphs are misaligned for better graphical representation.

Microsoft Human-AI Interaction Guidelines

Guideline 1: Make clear what the system can do.

Guideline 2: Make clear how well the system can do what it can do.

. . .

S. Amershi et. al., "Guidelines for Human-AI Interaction," CHI 2019

EU Requirements of Trustworthy AI

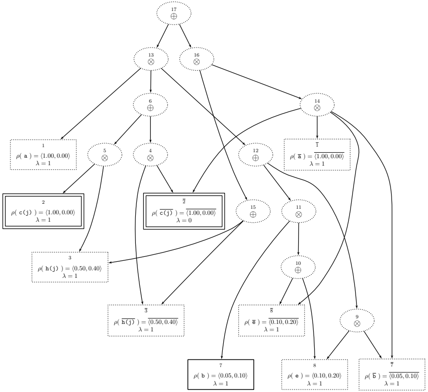
Human agency and oversight
Technical robustness and safety
Privacy and data governance
Transparency
Diversity, non-discrimination, and fairness
Societal and environmental wellbeing
Accountability

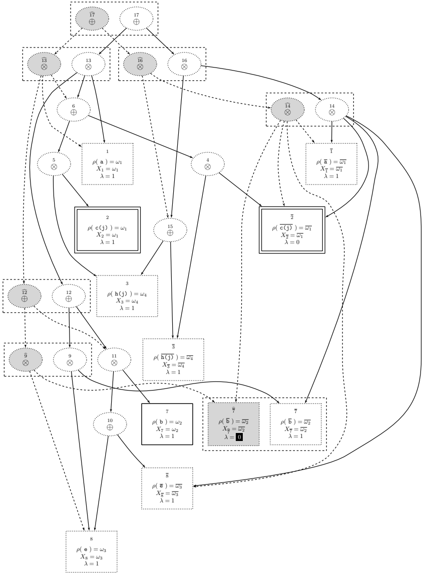
EUROPEAN COMMISSION, 2019. High-Level Expert Group on Artificial Intelligence.

```
ω2 :: burglary.
ω3 :: earthquake.
ω4 :: hears_alarm(john).
alarm :- burglary.
alarm :- earthquake.
calls(john) :- alarm, hears_alarm(john).
evidence(calls(john)).
query(burglary).
```

Identifier	Beta parameters
ω_1	Beta(∞, 1)
$\overline{\omega_1}$	Beta(1, ∞)
ω_2	Beta(2, 18)
$\overline{\omega_2}$	Beta(18, 2)
ω_3	Beta(2, 8)
$\overline{\omega_3}$	Beta(8, 2)
ω_4	Beta(3.5, 1.5)
$\overline{\omega_4}$	Beta(1.5, 3.5)

Cerutti, Kaplan, Kimmig, Şensoy, Handling Epistemic and Aleatory Uncertainties in Probabilistic Circuits, Under Submission, 2021, https://arxiv.org/abs/2102.10865





Cerutti, Kaplan, Kimmig, Şensoy, Handling Epistemic and Aleatory Uncertainties in Probabilistic Circuits, Under Submission, 2021, **https://arxiv.org/abs/2102.10865**

Let n be a $\oplus$-gate over C nodes, its children

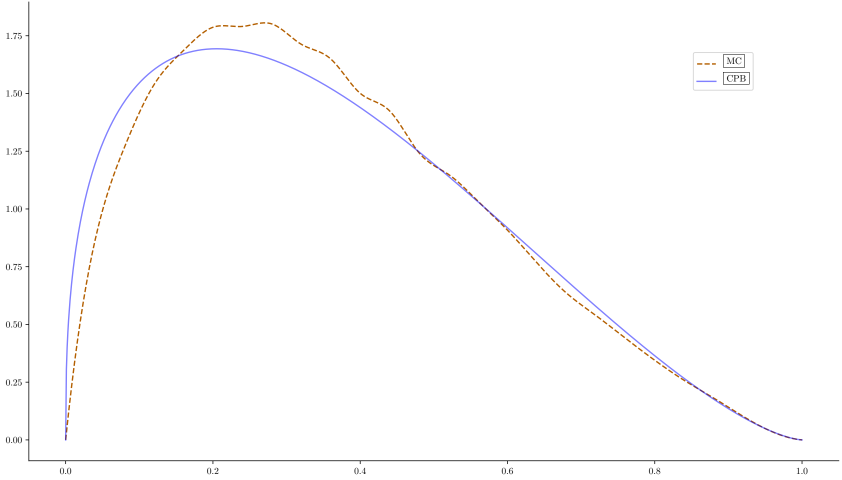
$$
\begin{aligned}
\mathbb{E}[X_n] &= \sum_{c \in C} \mathbb{E}[X_c], \\
\mathrm{cov}[X_n] &= \sum_{c \in C} \sum_{c' \in C} \mathrm{cov}[X_c, X_{c'}], \\
\mathrm{cov}[X_n, X_z] &= \sum_{c \in C} \mathrm{cov}[X_c, X_z] \quad \text{for } z \in \widehat{N_A} \setminus \{n\}
\end{aligned}
$$

Let n be a $\otimes$-gate over C nodes, its children

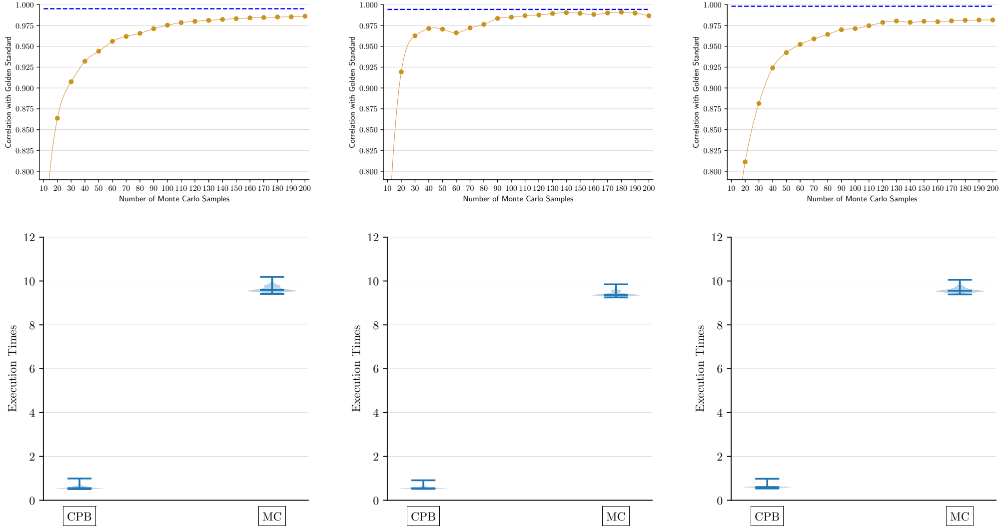
$$
\begin{aligned}
\mathbb{E}[X_n] &= \prod_{c \in C} \mathbb{E}[X_c], \\
\mathrm{cov}[X_n] &\simeq \sum_{c \in C} \sum_{c' \in C} \frac{\mathbb{E}[X_n]^2}{\mathbb{E}[X_c]\mathbb{E}[X_{c'}]} \mathrm{cov}[X_c, X_{c'}], \\
\mathrm{cov}[X_n, X_z] &\simeq \sum_{c \in C} \frac{\mathbb{E}[X_n]}{\mathbb{E}[X_c]} \mathrm{cov}[X_c, X_z] \quad \text{for } z \in \widehat{N_A} \setminus \{n\}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\frac{X_r}{X_{\widehat{r}}}\right] &\simeq \frac{\mathbb{E}[X_r]}{\mathbb{E}[X_{\widehat{r}}]}, \\
\mathrm{cov}\left[\frac{X_r}{X_{\widehat{r}}}\right] &\simeq \frac{1}{\mathbb{E}[X_{\widehat{r}}]^2} \mathrm{cov}[X_r] + \frac{\mathbb{E}[X_r]^2}{\mathbb{E}[X_{\widehat{r}}]^4} \mathrm{cov}[X_{\widehat{r}}] - 2 \frac{\mathbb{E}[X_r]}{\mathbb{E}[X_{\widehat{r}}]^3} \mathrm{cov}[X_r, X_{\widehat{r}}].
\end{aligned}
$$

Cerutti, Kaplan, Kimmig, Şensoy, Handling Epistemic and Aleatory Uncertainties in Probabilistic Circuits, Under Submission, 2021, **https://arxiv.org/abs/2102.10865**

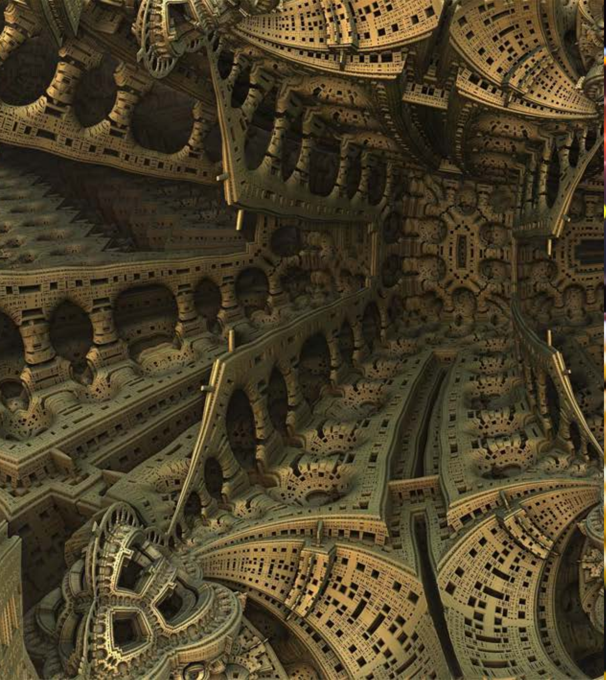


Cerutti, Kaplan, Kimmig, Şensoy, Handling Epistemic and Aleatory Uncertainties in Probabilistic Circuits, Under Submission, 2021, **https://arxiv.org/abs/2102.10865**



Cerutti, Kaplan, Kimmig, Şensoy, Handling Epistemic and Aleatory Uncertainties in Probabilistic Circuits, Under Submission, 2021, **https://arxiv.org/abs/2102.10865**

Overture.	A brief historical case.
Act I.	On conjectures, refutations, and argumentation.
Act II.	There is no certain datum in the world.
Act III.	Interesting problems are complex.
Epilogue.	




Major
Minor
Ackn.
4
5

A Trustworthy Loss Function

Classification becomes regression outputting pieces of evidences in favour of different classes

Expected squared error (aka Brier score) with $\mathrm{Dir}(\boldsymbol{m}_i \mid \boldsymbol{\alpha}_i)$ (prior for a Multinomial) penalising the divergence from the uniform distribution:

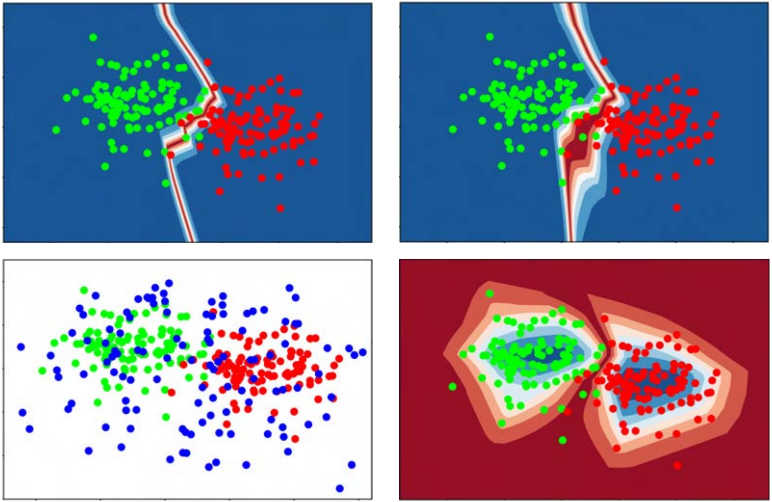
$$\mathcal{L} = \sum_{i=1}^{N} \mathbb{E}[\|\boldsymbol{y}_i - \boldsymbol{m}_i\|_2^2] + \lambda_t \sum_{i=1}^{N} \mathrm{KL}\left(\mathrm{Dir}(\boldsymbol{\mu}_i \mid \widetilde{\boldsymbol{\alpha}}_i) \,\|\, \mathrm{Dir}(\boldsymbol{\mu}_i \mid \mathbf{1}) \right)$$

where:

- λ_t avoid premature convergence to the uniform distribution;
- $\widetilde{\boldsymbol{\alpha}}_i = \boldsymbol{y}_i + (\mathbf{1} - \boldsymbol{y}_i) \cdot \alpha_i$ are the Dirichlet parameters the neural network in a forward pass has put on the wrong classes, and the idea is to minimise them as much as possible.
- $\mathrm{KL}\left(\mathrm{Dir}(\boldsymbol{\mu}_i \mid \widetilde{\boldsymbol{\alpha}}_i) \,\|\, \mathrm{Dir}(\boldsymbol{\mu}_i \mid \mathbf{1}) \right) = \ln\left(\frac{\Gamma(\sum_{k=1}^{K} \widetilde{\alpha}_{i,k})}{\Gamma(K) \prod_{k=1}^{K} \Gamma(\widetilde{\alpha}_{i,k})} \right) + \sum_{k=1}^{K} (\widetilde{\alpha}_{i,k} - 1) \left[\psi(\widetilde{\alpha}_{i,k}) - \psi\left(\sum_{j=1}^{K} \widetilde{\alpha}_{i,j} \right) \right]$ where $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \ln(\Gamma(x))$ is the *digamma* function

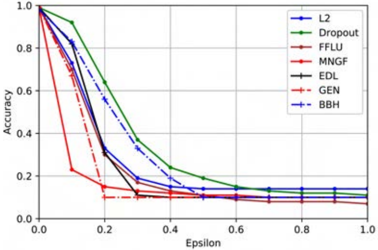
Şensoy, Kaplan, and Kandemir. "Evidential deep learning to quantify classification uncertainty." NeurIPS. 2018.

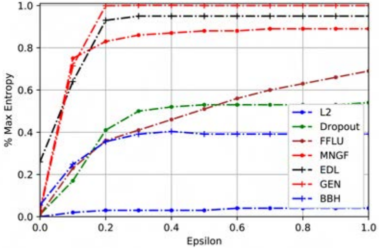
EDL + GAN for adversarial training



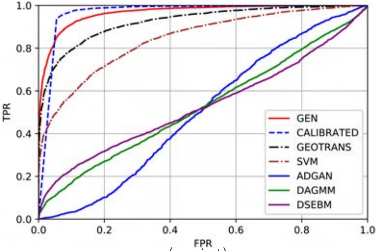
Şensoy, Kaplan, Cerutti, and Saleki. "Uncertainty-Aware Deep Classifiers using Generative Models." AAAI 2020

Robustness against FGS

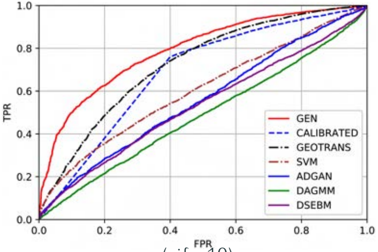




Anomaly detection



(mnist)



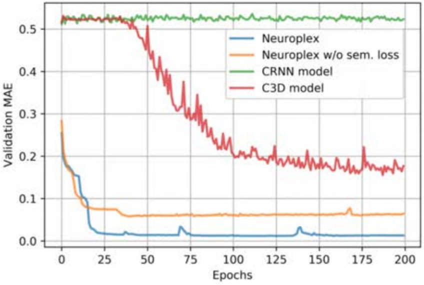
(cifar10)

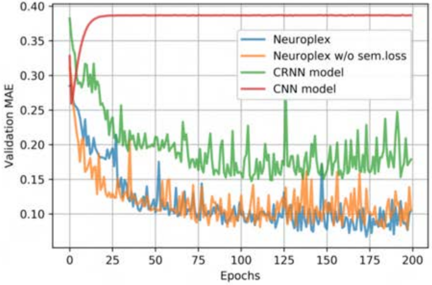
Şensoy, Kaplan, Cerutti, and Saleki. "Uncertainty-Aware Deep Classifiers using Generative Models." AAAI 2020

Roig Vilamala et. al. "A Hybrid Neuro-Symbolic Approach for Complex Event Processing (Extended Abstract)." In ICLP2020.

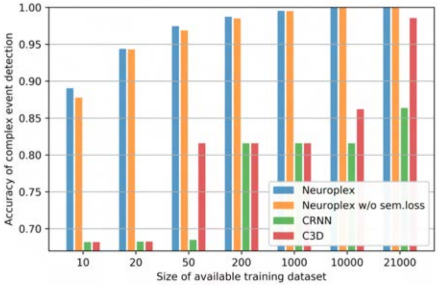
Xing et. al. "Neuroplex: Learning to Detect Complex Events in Sensor Networks through Knowledge Injection." In SenSys2020.

NeuroPLEX





	Sim. 1	Sim. 2	Sim. 3	Sim. 4	Sim. 5
Window Length	10	20	30	3	2
# of Uniq Events	10	10	10	3	3
# of CE	4	4	7	5	4
Avg. CE Length	2.8	2.8	3.43	2	2
Neuroplex	**99.39%**	**99.56%**	**98.65%**	**100.00%**	99.98 %
Lenet(Neuroplex)	98.87%	99.17%	98.91%	99.84%	99.78%
CRNN model	69.98%	7.79%	1.83%	86.37%	**99.99%**
C3D model	88.47%	83.73%	86.91%	98.56%	99.72%



Xing et. al. "Neuroplex: Learning to Detect Complex Events in Sensor Networks through Knowledge Injection." In SenSys2020.

Overture.	A brief historical case.
Act I.	On conjectures, refutations, and argumentation.
Act II.	There is no certain datum in the world.
Act III.	Interesting problems are complex.
Epilogue.	


www.grundner.co.at
www.grundner.co.at
65
70
85
90
95

Roig Vilamala et. al. “A Hybrid Neuro-Symbolic Approach for Complex Event Processing (Extended Abstract).” In ICLP2020.

Xing et. al. “Neuroplex: Learning to Detect Complex Events in Sensor Networks through Knowledge Injection.” In SenSys2020.

Co-I

S. Chakraborty **IBM Research T. J. Watson** • M. Giacomin **Brescia** • L. Kaplan **US CCDC ARL**
A. Kimmig **KU Leuven** • S. Julier **UCL** • Y. McDermott-Rees **Swansea** • T. Norman **Southampton**
N. Oren **Aberdeen** • G. Pearson **UK MoD Dstl** • A. Preece **Cardiff** • M. Şensoy **Ozyegin**
M. Srivastava **UCLA** • M. Thimm **Hagen** • N. Tintarev **Maastricht** • A. Toniolo **St. Andrews**
M. Vallati **Huddersfield**

Intern/PhD/Post-Doc

C. Allen **Cardiff** • A. Fanelli **Brescia** • L. Garcia **UCLA** • S. Habib **UCL** • C. Hougen **Michigan**
O. Lipinski **Southampton** • K. Mishra **US CCDC ARL** • M. Roig Vilamala **Cardiff** • H. Rose **UCL**
G. Pellier-Hollows **Cardiff** • T. Xing **UCLA** • T. Zanetti **Cardiff**