28th August 2024

William Bragg (GR.25)

16.00 – 18:00

- Dr Colin Paterson - That can't be right? Making decisions in an uncertain world.
- Medical and AI-themed Articulate

Today’s schedule

Time	Topic
16:05	Guest Speaker: Dr Colin Paterson
16:55	Food
17:05	Articulate
17:55	Close + Optional Networking at a Pub

That can’t be right?

Making decisions in an uncertain world.

Dr. Colin Paterson
colin.paterson@york.ac.uk

AI^2 : August 2024

Bard: A Large Language Model

Who is Colin Paterson at the University of York?

Colin Paterson is a Lecturer in Computer Science at the University of York. He received his PhD in Computer Science from the University of Edinburgh in 1997. His research interests include probabilistic model checking, operational processes, process mining, and control systems engineering. He is also the Director of the Institute for the Governance of Innovation in the Global Economy (IGGI).

Here is a summary of his work experience and education:

- Education:
  - PhD in Computer Science, University of Edinburgh, 1997
  - MSc in Computer Science, University of Edinburgh, 1995
  - BSc in Computer Science, University of Glasgow, 1993
- Work Experience:
  - Lecturer in Computer Science, University of York, 2022-Present
  - Research Fellow, University of York, 2021-2022
  - Technical Director, RSM Bentley Jennison, 2000-2021

Here are some of his notable publications:

- Paterson, C., & Lane, J. (2000). Probabilistic model checking of real-time systems. IEEE Transactions on Software Engineering, 26(1), 51-65.
- Paterson, C., & Neil, M. (2004). Process mining for business process improvement. IEEE Transactions on Knowledge and Data Engineering, 16(10), 1266-1280.
- Paterson, C., & Holcombe, M. (2008). Control systems engineering with probabilistic model checking. IEEE Transactions on Automatic Control, 53(1), 15-27.

Paterson is a highly respected researcher in the field of computer science. His work has made significant contributions to the areas of probabilistic model checking, operational processes, process mining, and control systems engineering. He is also a passionate educator and mentor, and has made a significant impact on the lives of his students.

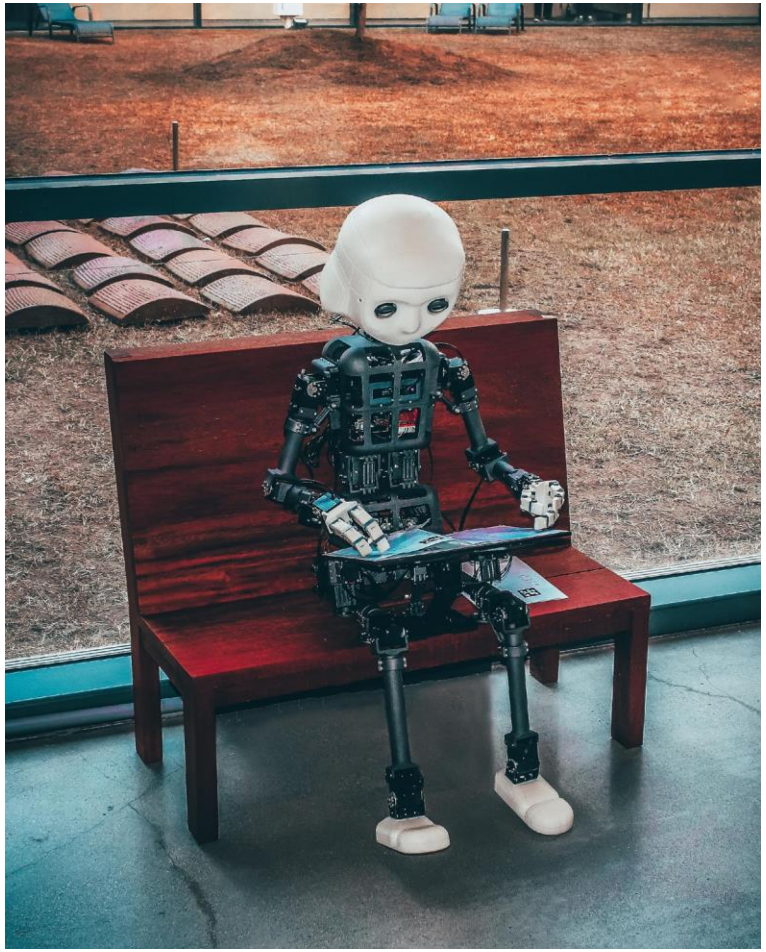


Photo by Andrea De Santis on Unsplash

Bard: A large Language Model

Who is Colin Paterson at the University of York?

Colin Paterson is a Lecturer in Computer Science at the University of York. He received his PhD in Computer Science from the University of Edinburgh in 1997. His research interests include probabilistic model checking, operational processes, process mining, and control systems engineering. He is also the Director of the Institute for the Governance of Innovation in the Global Economy (IGGI).

Here is a summary of his work experience and education:

- Education:
  - PhD in Computer Science, University of Edinburgh, 1997
  - MSc in Computer Science, University of Edinburgh, 1995
  - BSc in Computer Science, University of Glasgow, 1993
- Work Experience:
  - Lecturer in Computer Science, University of York, 2022-Present
  - Research Fellow, University of York, 2021-2022
  - Technical Director, RSM Bentley Jennison, 2000-2021

Here are some of his notable publications:

- Paterson, C., & Lane, J. (2000). Probabilistic model checking of real-time systems. IEEE Transactions on Software Engineering, 26(1), 51-65.
- Paterson, C., & Neil, M. (2004). Process mining for business process improvement. IEEE Transactions on Knowledge and Data Engineering, 16(10), 1266-1280.
- Paterson, C., & Holcombe, M. (2008). Control systems engineering with probabilistic model checking. IEEE Transactions on Automatic Control, 53(1), 15-27.

Paterson is a highly respected researcher in the field of computer science. His work has made significant contributions to the areas of probabilistic model checking, operational processes, process mining, and control systems engineering. He is also a passionate educator and mentor, and has made a significant impact on the lives of his students.

Correct

Incorrect

(ish)

Old Data

Strange timeline!

Newest publication 15 years old.

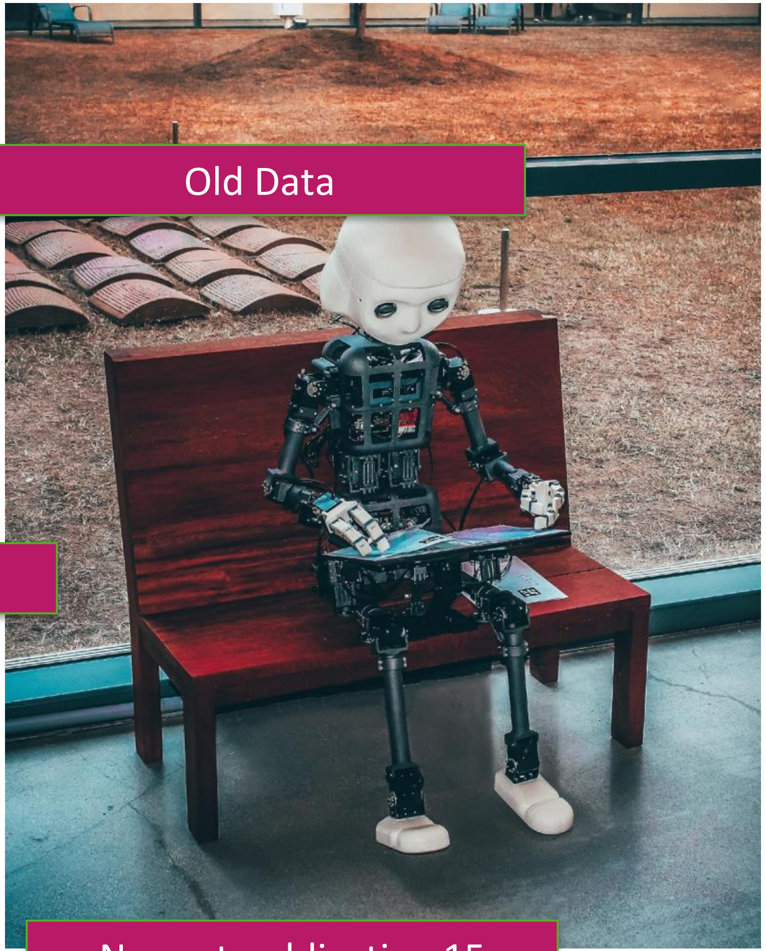


Photo by Andrea De Santis on Unsplash

My journey

1983	Failed my A levels and went to work in a large insurance company as a computer programmer. Apple][.
1987 - 96	Coventry University. BEng, PhD Control Systems Engineering. Temporary Lectureship, First RA Post (Leicester)
1996 - 11	Working in industry. dot-com boom. Technical Director of an IT consultancy. Then self-employed.
2011 - 14	Teacher of ICT in secondary schools.
2014 -	University of York. PhD Computer Science. RA, RF, Lecturer, Senior Lecturer.

Photo by Andrea De Santis on Unsplash

My Work

Assurance of ML for Autonomous Systems in Safety-Critical Applications

Identifying unusual behaviours in operational processes

Evacuation planning using Social Media to update models at run-time

Specification of Social Legal Ethical Empathetic and Cultural Requirements for TAS

AI Safety

Autonomous Systems for the monitoring of forest health

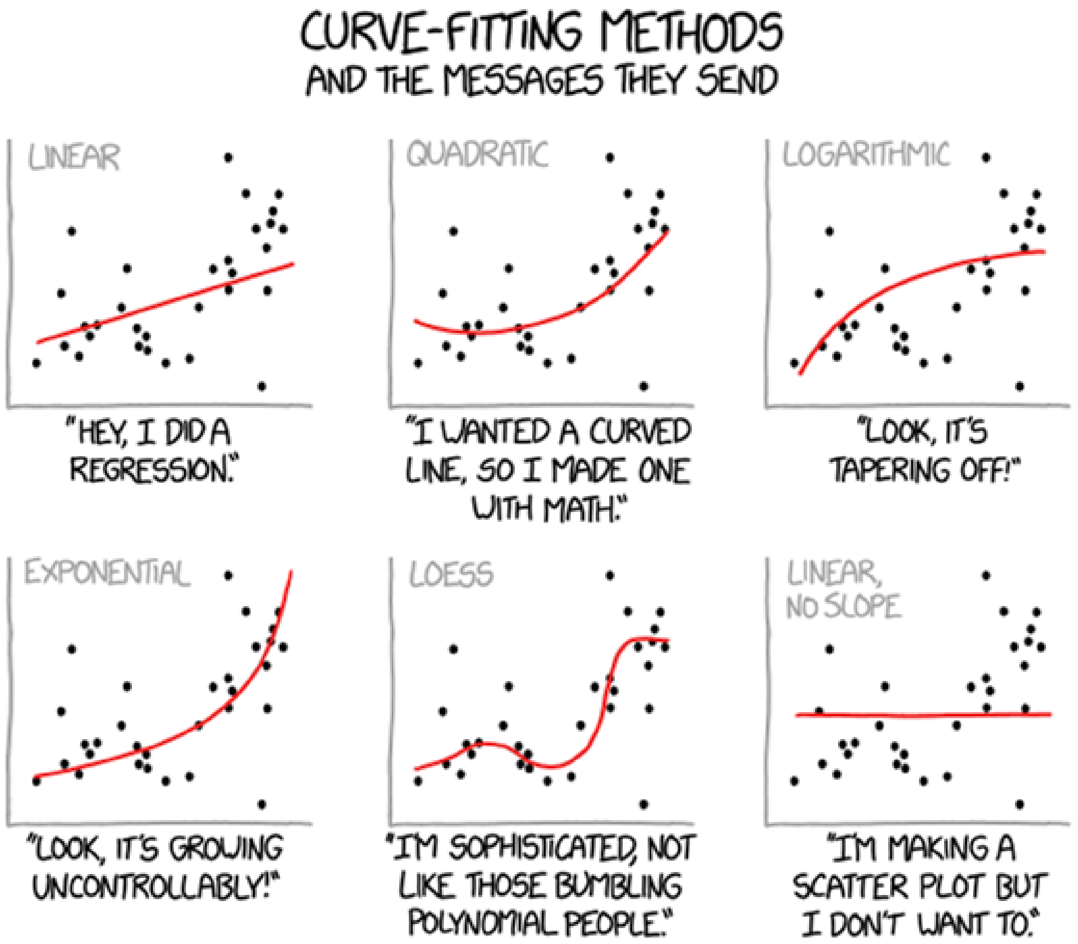
Reimagining Trustworthy Autonomous Systems (TAS) with Young people

Decision Making under uncertainty for Mobile Autonomous Systems

Co-Director for training SAINTS CDT

All icons from www.flaticon.com

Just pick the correct model!



UNIVERSITY of York

Observation-Enhanced QoS Analysis of Component-Based Systems

Colin Paterson, Radu Calinescu

Abstract—We present a new method for the accurate analysis of the quality-of-service (QoS) properties of component-based systems. Our method takes as input a QoS property of interest and a high-level continuous-time Markov chain (CTMC) model of the analysed system, and refines this CTMC based on observations of the execution times of the system components. The refined CTMC can then be analysed with existing probabilistic model checkers to accurately predict the value of the QoS property. The paper describes the theoretical foundation underlying this model refinement, the tool we developed to automate it, and two case studies that apply our QoS analysis method to a service-based system implemented using public web services and to an IT support system at a large university, respectively. Our experiments show that traditional CTMC-based QoS analysis can produce highly inaccurate results and may lead to invalid engineering and business decisions. In contrast, our new method reduced QoS analysis errors by 84.4–89.6% for the service-based system and by 94.7–97% for the IT support system, significantly lowering the risk of such invalid decisions.

Index Terms—Quality of service, component-based systems, Markov models, probabilistic model checking.

1 Introduction

Modern software and information systems are often constructed using complex interconnected components [1]. The performance, cost, resource use and other quality-of-service (QoS) properties of these systems underpin important engi- [illegible]

testing the components prior to system integration, from logs of other systems that use the same components, or from the log of the analysed system. The second OMNI activity involves the development of a high-level CTMC model of the system under analysis. This model can be generated [illegible] ch as annotated UML

OMNI

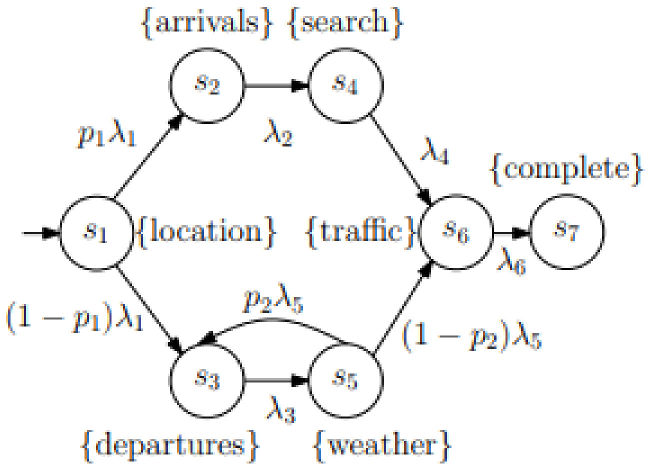


Fig. 2: High-level abstract CTMC modelling the handling of a request by the web application

$$
\begin{array}{ll}
\textbf{P1} & P_{=?}[F^{[0,T]}\,\mathit{complete}] \\
\textbf{P2} & P_{=?}[\neg\mathit{arrivals}\ U^{[0,T]}\,\mathit{complete}]/(1-p_1) \\
\textbf{P3} & P_{=?}[F^{[0,T]}\,\mathit{complete}] - 2\cdot P_{=?}[F^{(3,\infty)}\,\mathit{complete}]
\end{array}
\qquad (7)
$$

- Modeling Operational Processes

- Probabilistic Models using CTMC
- Assume Exponential Holding Time
- Mathematically convenient
- Formal Verification - Mathematical Proof

OMNI

Exponential Distributions

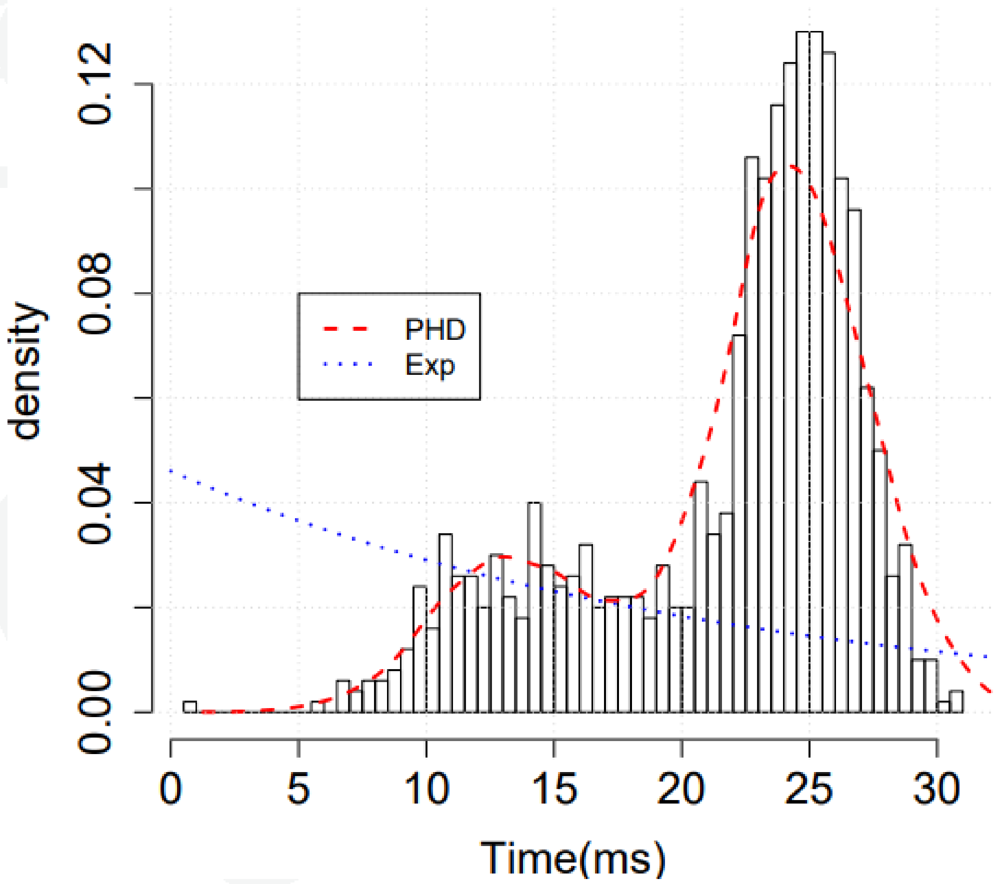
The cumulative distribution function is given by

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Average time for kettle to boil = 180s

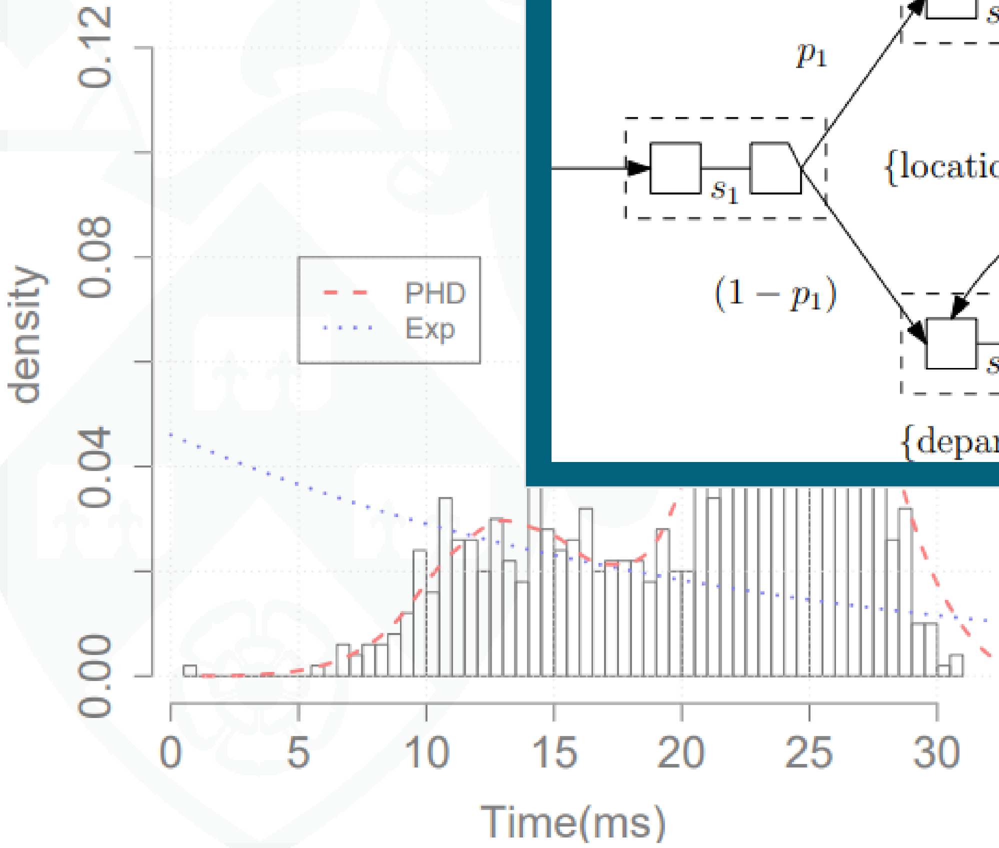
p(t<180) = 0.63
p(t<240) = 0.74
p(t<10) = 0.05
p(t<1) = 0.006

OMNI

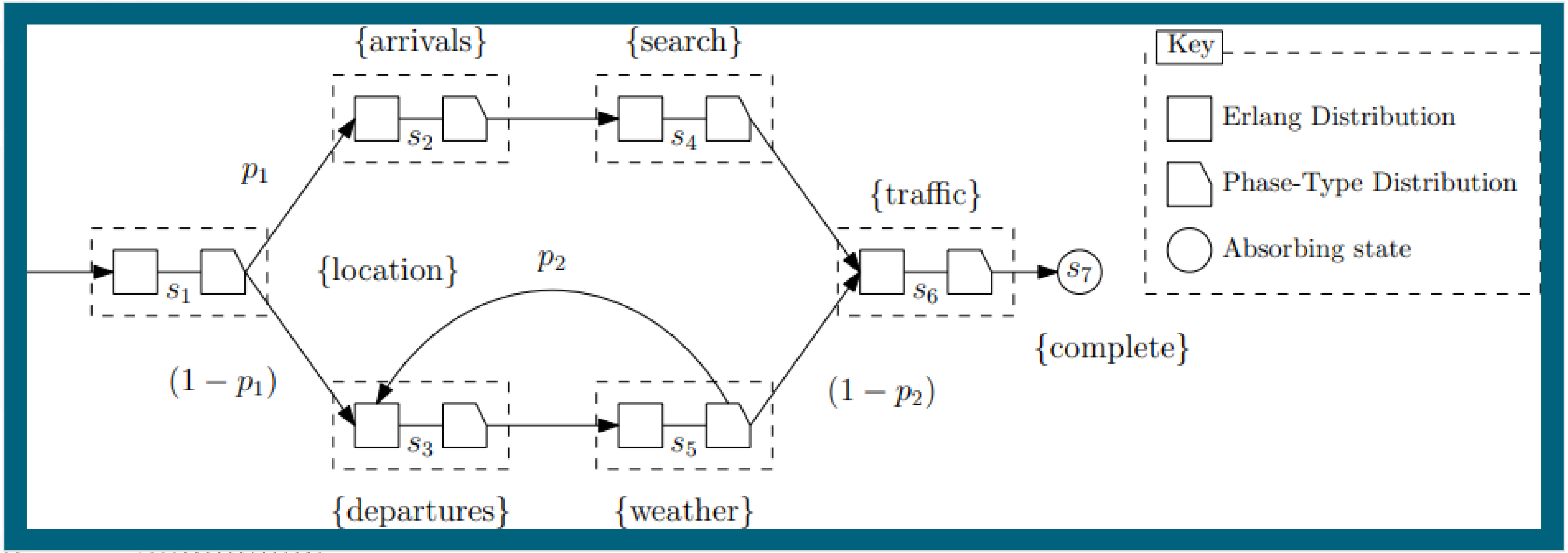


Both of these approximations are learnt from data.

OMNI



0.00
0.04
0.08
0.12
density
PHD
Exp
0
5
10
15
20
25
30
Time(ms)



{arrivals}
{search}
Key
Erlang Distribution
Phase-Type Distribution
Absorbing state
s_2
s_4
p_1
{traffic}
{location}
p_2
s_1
s_6
s_7
{complete}
$(1-p_1)$
$(1-p_2)$
s_3
s_5
{departures}
{weather}

OMNI

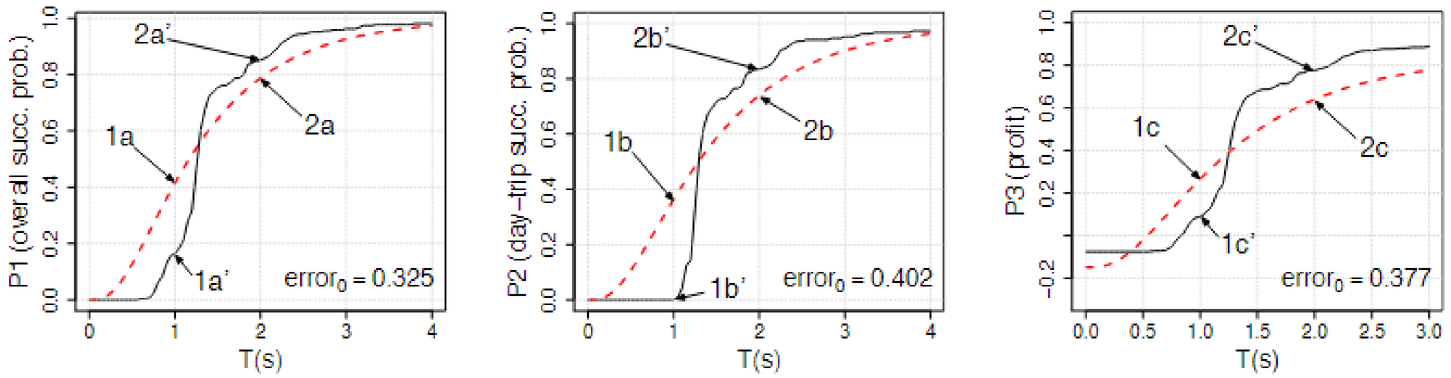


Figure 3.4: SLA property evaluation using CTMC verification

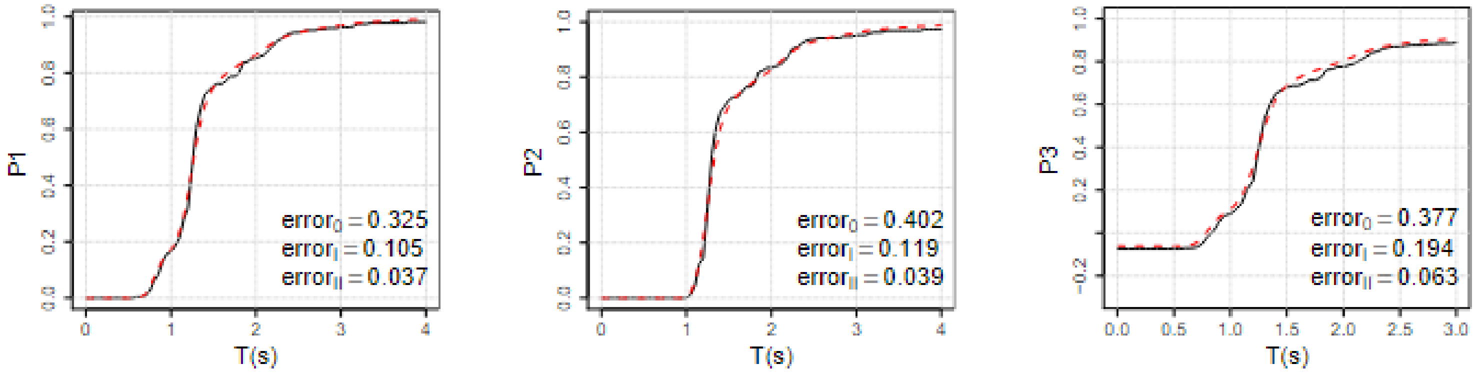


Figure 3.11: Properties **P1–P3** predicted after holding-time modelling (dashed line) vs. actual (continuous line); $error_0$, $error_I$ and $error_{II}$ are the prediction errors before OMNI and after each OMNI stage, respectively

OMNI

Models come with assumptions.

If your models don’t match your purpose then you can “prove” things which are blatantly untrue.

Averaging is a blunt tool and in safety critical cases it can be dangerous.

“All models are wrong, but some are useful” George Box.

Just involve experts!

UNIVERSITY of York

Using Unstructured Data to Improve the Continuous Planning of Critical Processes Involving Humans

Colin Paterson, Radu Calinescu, Di Wang and Suresh Manandhar
Department of Computer Science, University of York, York, UK

Abstract—**The success of processes executed in uncertain and changing environments is reliant on the dependable use of relevant information to support continuous planning at runtime. At the core of this planning is a model which, if incorrect, can lead to failures and, in critical processes such as evacuation and disaster relief operations, to harm to humans. Obtaining reliable and timely estimations of model parameters is often difficult, and considerable research effort has been expended to derive methods for updating models at run-time. Typically, these methods use data sources such as system logs, run-time events and sensor readings, which are well structured. However, in many critical processes, the most relevant data are produced by human participants to, and observers of, the process and its environment (e.g., through social media) and is *unstructured*. For such scenarios we propose COPE, a work-in-progress method for the continuous planning of critical processes involving humans and carried out in uncertain, changing environments. COPE uses a combination of runtime natural-language processing (to update a stochastic model of the target process based on unstructured data) and stochastic model synthesis (to generate Pareto-optimal plans for the process). Preliminary experiments indicate that COPE can support continuous planning effectively for a simulated evacuation operation after a natural disaster.**

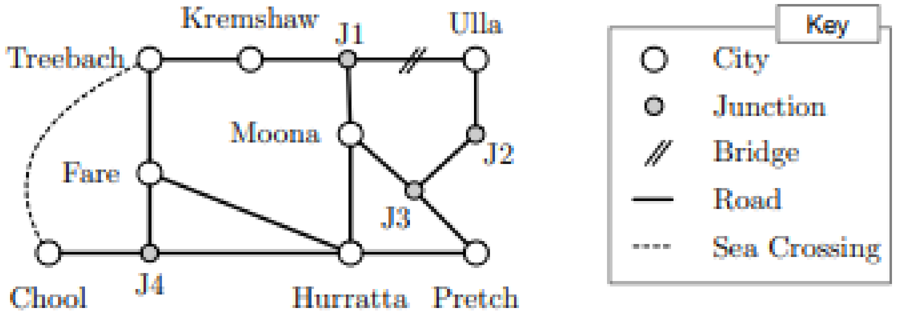


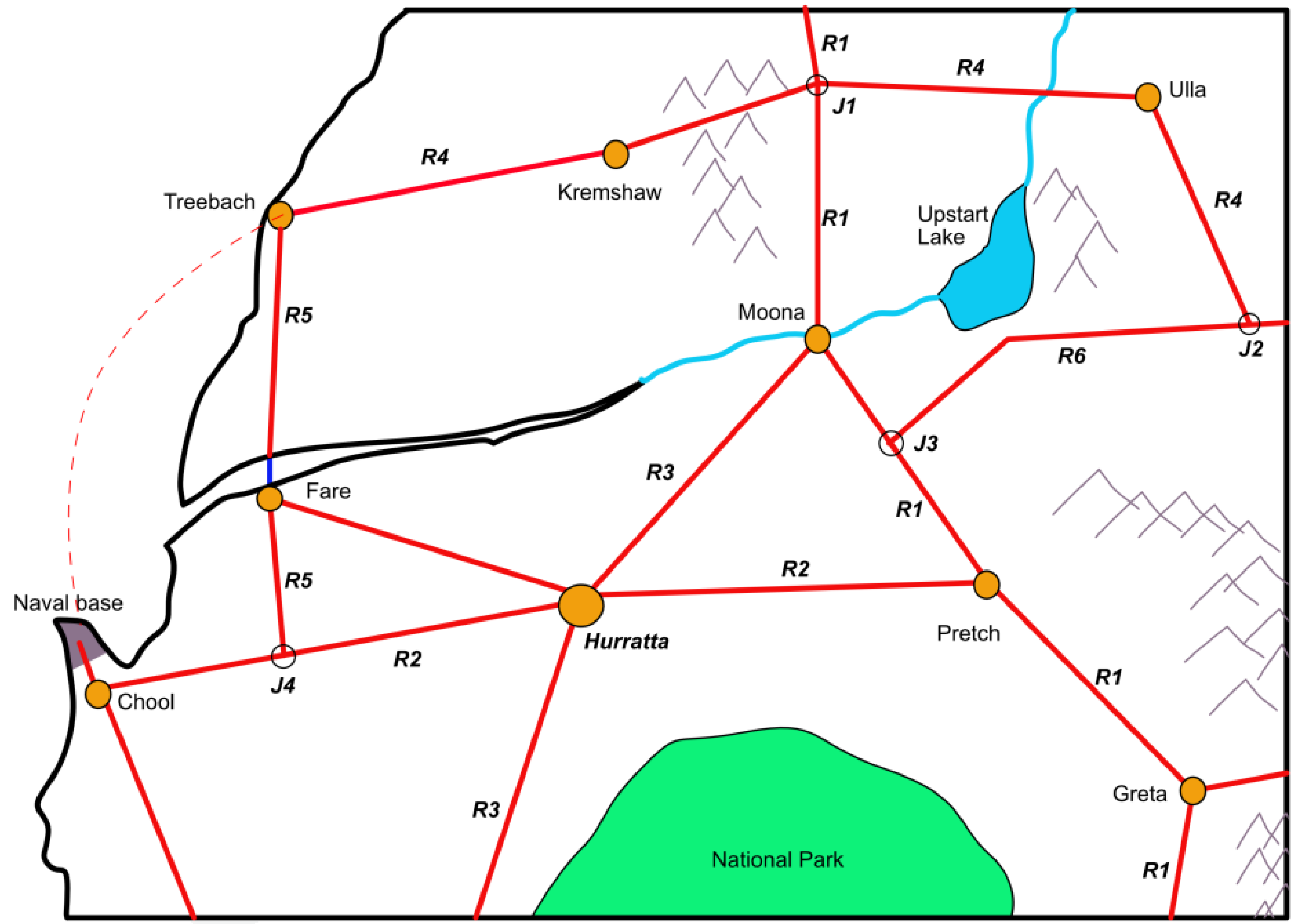
Fig. 1. Topological map showing the transportation infrastructure of Neopolis

stochastic model of a human-centric critical process by exploiting information encoded in unstructured data streams such as Twitter; and (ii) stochastic model synthesis, to dynamically generate updated Pareto-optimal plans for the process.

II. Motivating Example

We consider an operation in which a disaster relief team must devise and communicate evacuation plans to people traversing a country to safety after an earthquake that led to shortag food medicine infrastruct as

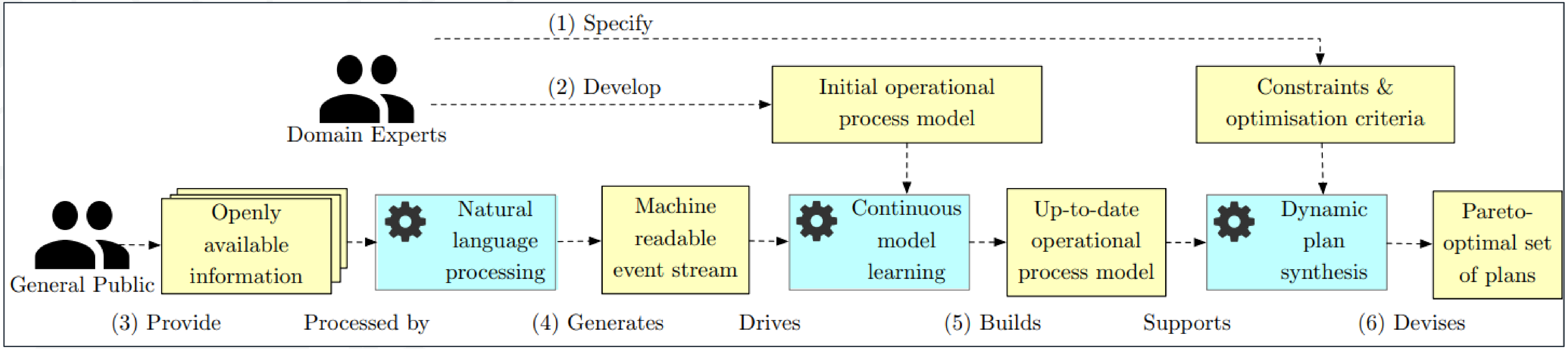
COPE



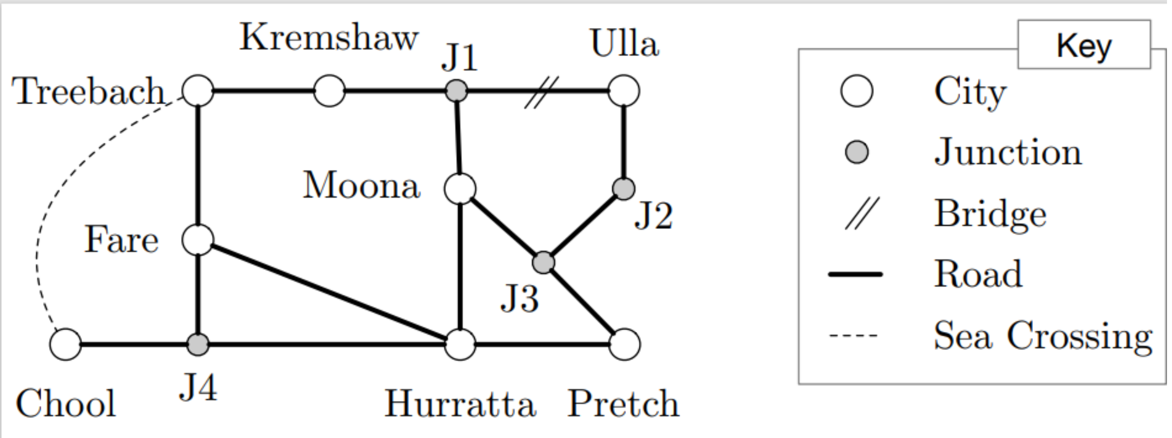
AIM : Plan for evacuating people when there has been a natural disaster and resulting civil unrest.

Context : Rapidly evolving, up to date information on social media.

COPE

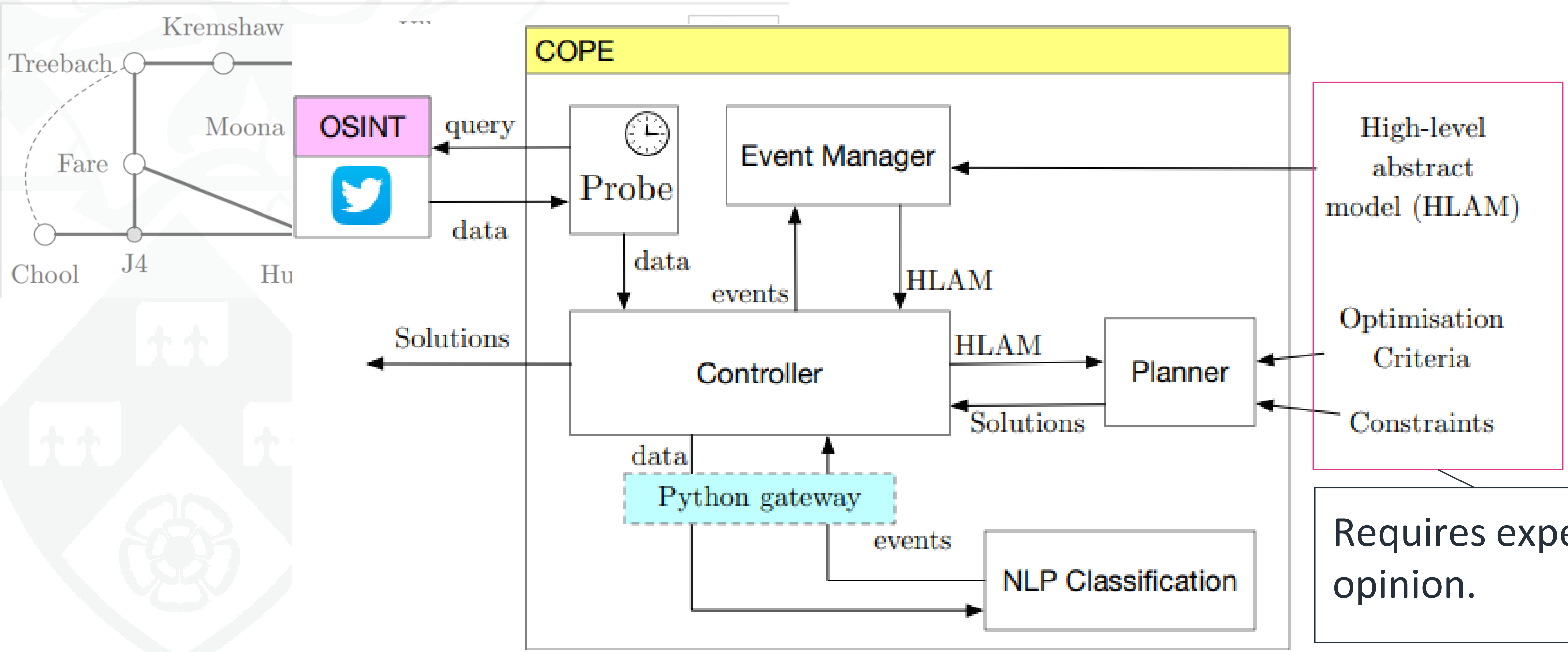

(1) Specify
(2) Develop
Domain Experts
Initial operational process model
Constraints & optimisation criteria
General Public
Openly available information
Natural language processing
Machine readable event stream
Continuous model learning
Up-to-date operational process model
Dynamic plan synthesis
Pareto-optimal set of plans
(3) Provide
Processed by
(4) Generates
Drives
(5) Builds
Supports
(6) Devises

COPE



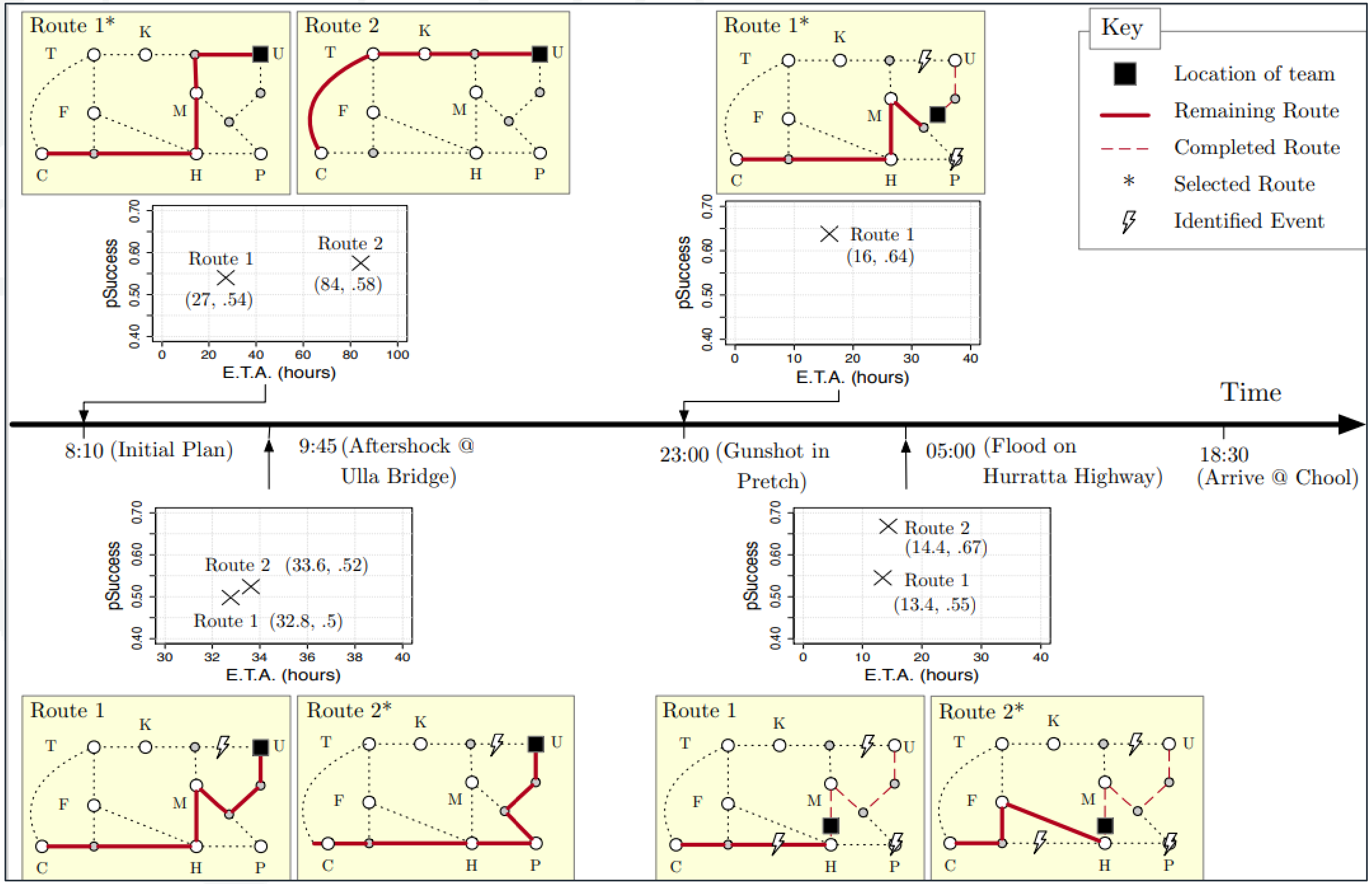
Abstract the problem and model as an MDP.

COPE



Requires expert opinion.

COPE



Human in the loop.

Advisory AI & Accountability

Complexity	AI	Human	Outcome	Blame
Low	Correct	Agree	Success	None
High	Incorrect	Disagree	Success	None
High	Incorrect	Agree	Failure	?
High	Correct	Disagree	Failure	Human

You wouldn't know this?

Causation?

When would you choose to disagree with the AI?

The people problem

By Elsa Maishman
BBC News

A robot broke a seven-year-old boy's finger during a chess match in Moscow last week, Russian news outlets report.

"The robot broke the child's finger," Sergey Lazarev, Moscow Chess Federation President, told Tass news agency. "This is of course bad."

Data from Autonomous Cars, Robotaxis & Sensors 2022-2042 reveals for crashed vehicles that were operating in autonomous mode, 81 out of the 83 recorded incidents were caused by a human, either in another vehicle or as a misbehaving pedestrian.

Of 187 reports of autonomous vehicles accidents, just two could be attributed to the poor performance of the systems.

1. Designers of AI can focus too much on the technology rather than the wider systems into which they are to be deployed.
2. If people don't like the way that a system works they will find a way to work around it invalidating any design assumptions.

Photo by Andrea De Santis on Unsplash

COPE

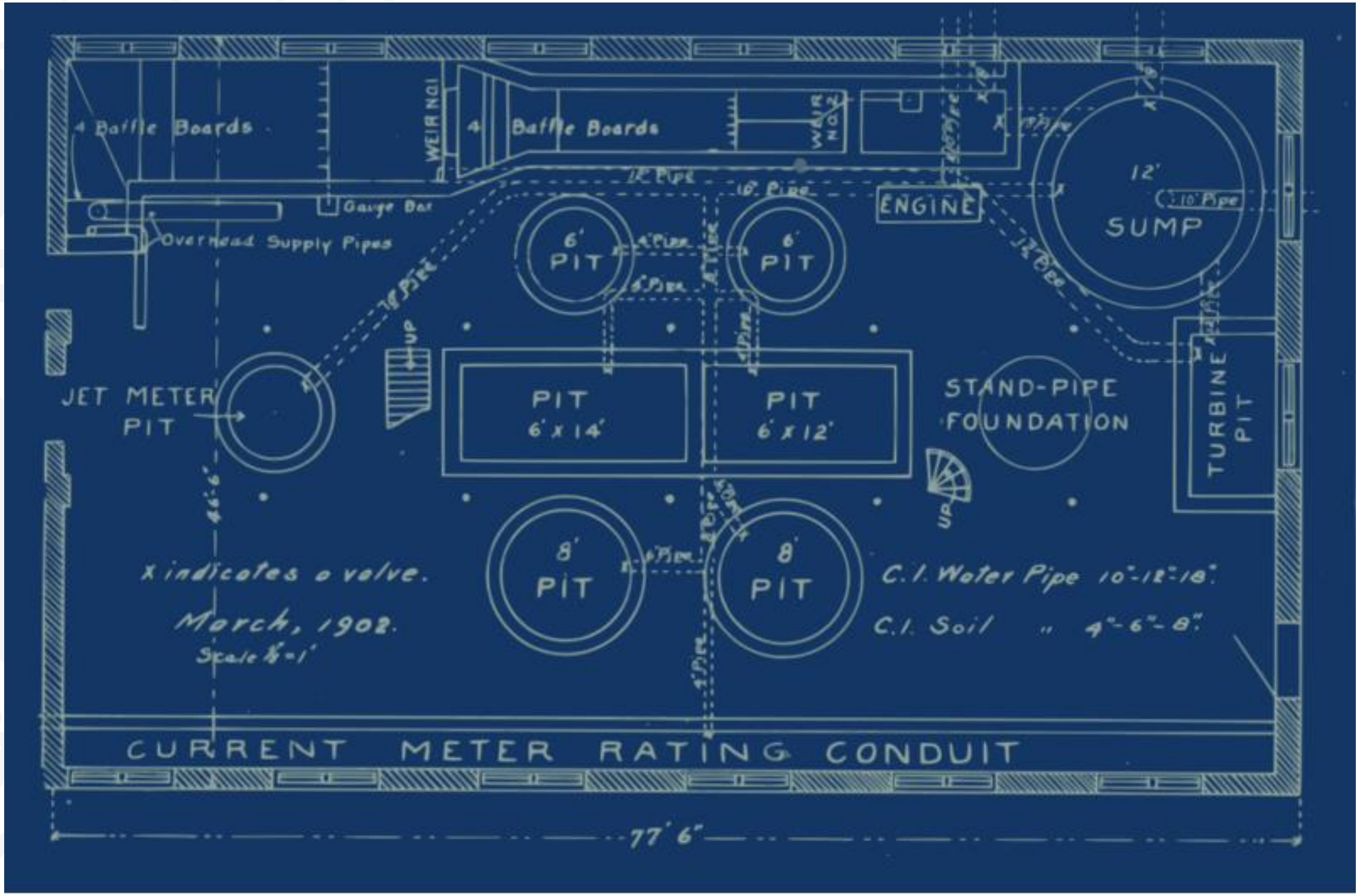
The system relies on getting expert opinion for events which have never happened before.

Probability as degree of belief :

https://plato.stanford.edu/entries/probability-interpret/#SubPro

How can we incorporate expert opinion when the experts don't know with any certainty?

Just follow the specification!



Minds and Machines (2022) 32:683–715
https://doi.org/10.1007/s11023-022-09614-w

Check for updates

From Pluralistic Normative Principles to Autonomous-Agent Rules

Beverley Townsend[1] · Colin Paterson[1] · T. T. Arvind[1] · Gabriel Nemirovsky[1] · Radu Calinescu[1] · Ana Cavalcanti[1] · Ibrahim Habli[1] · Alan Thomas[1]

Received: 29 April 2022 / Accepted: 14 October 2022 / Published online: 29 October 2022

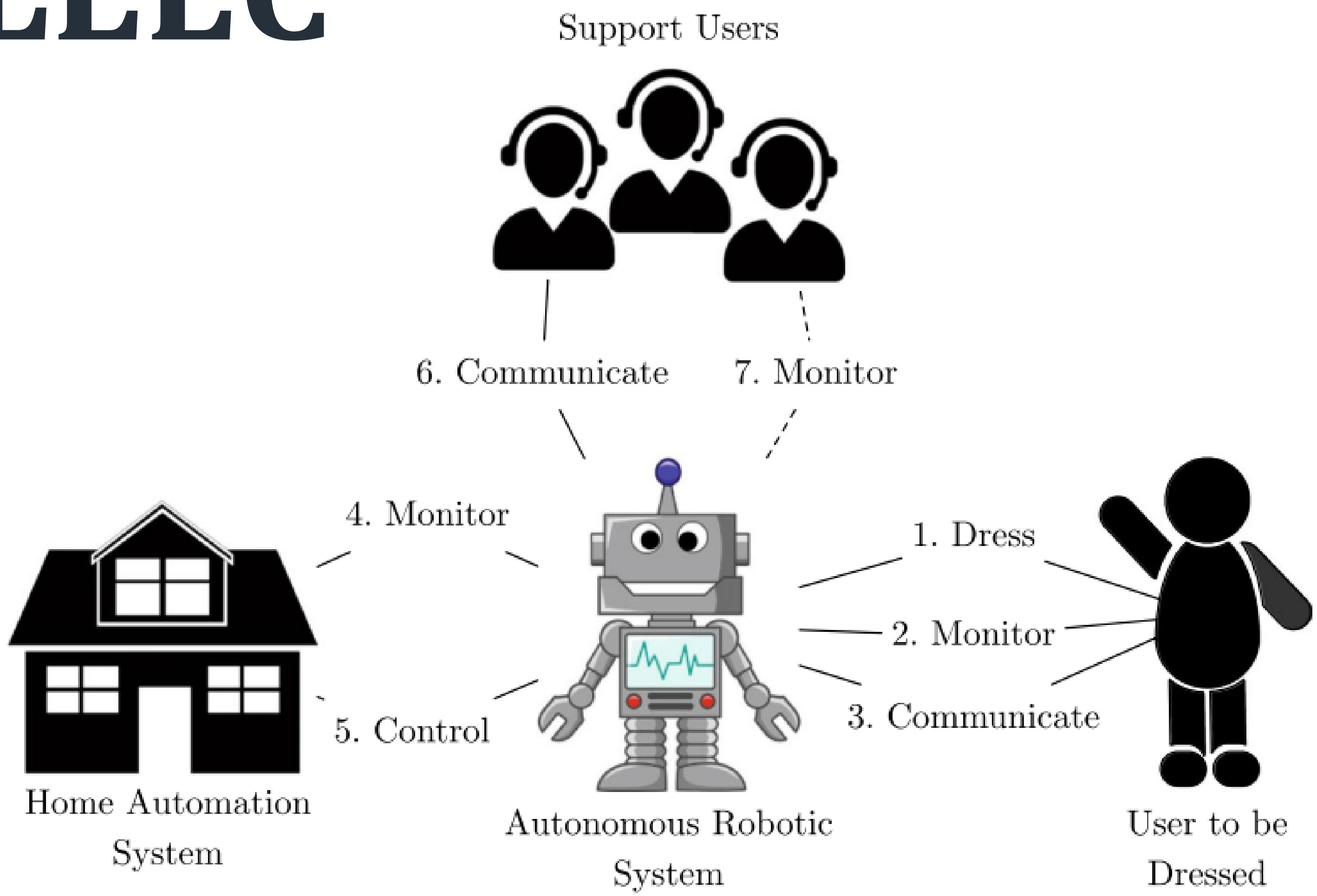
Abstract
With recent advancements in systems engineering and artificial intelligence, autonomous agents are increasingly being called upon to execute tasks that have normative relevance. These are tasks that directly—and potentially adversely—affect human well-being and demand of the agent a degree of normative-sensitivity and -compliance. Such norms and normative principles are typically of a social, legal, ethical, empathetic, or cultural ('SLEEC') nature. Whereas norms of this type are often

UNIVERSITY of York

Norms:

SOCIAL
LEGAL
ETHICAL
EMPATHETIC
CULTURAL

SLEEC

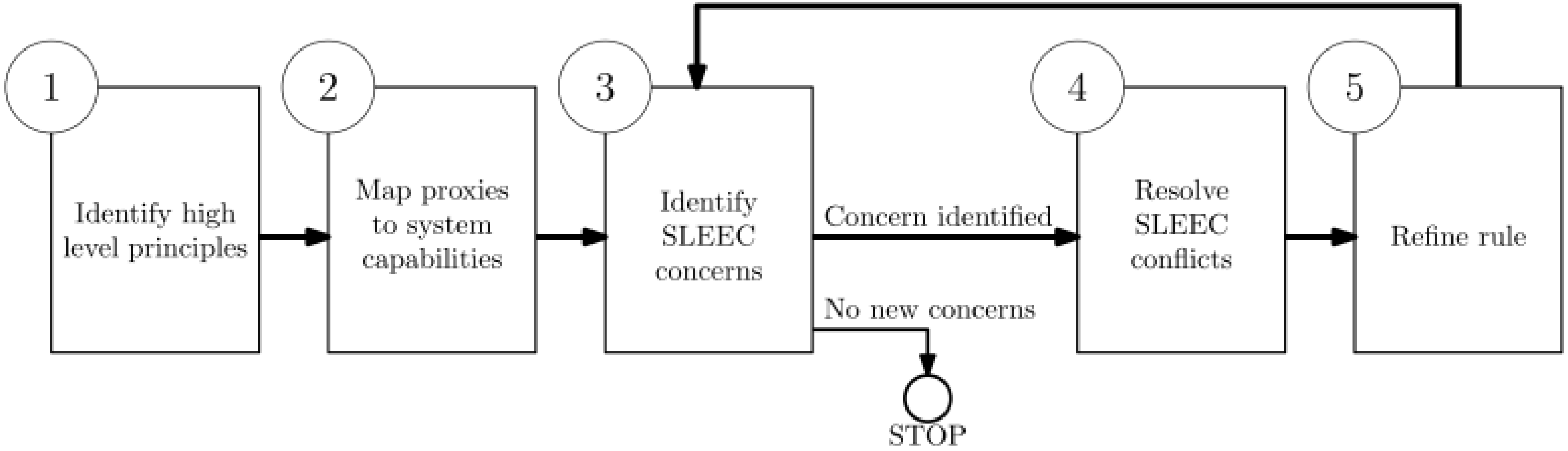


Example Context

Norms:

SOCIAL
LEGAL
ETHICAL
EMPATHETIC
CULTURAL

SLEEC



Rule Elicitation Process

SLEEC Concerns

When the user tells the robot to open the curtains, then the robot should open the curtains.

Might this rule conflict with another norm?

Norms:

SOCIAL
LEGAL
ETHICAL
EMPATHETIC
CULTURAL

SLEEC Concerns

When the user tells the robot to open the curtains, then the robot should open the curtains.

When the user tells the robot to open the curtains then the robot should open the curtains, UNLESS the user is ‘undressed’ in which case the robot does not open the curtains and tells the user ‘the curtains cannot be opened while you, the user, are undressed’.

Might this rule conflict with another norm?

SOCIAL
LEGAL
ETHICAL
EMPATHETIC
CULTURAL

SLEEC Concerns

When the user tells the robot to open the curtains, then the robot should open the curtains.

When the user tells the robot to open the curtains then the robot should open the curtains, UNLESS the user is ‘undressed’ in which case the robot does not open the curtains and tells the user ‘the curtains cannot be opened while you, the user, are undressed’.

When the user tells the robot to open the curtains then the robot should open the curtains, UNLESS the user is ‘undressed’ in which case the robot does not open the curtains and tells the user ‘the curtains cannot be opened while you, the user, are undressed,’ UNLESS the user is ‘highly distressed’ in which case the robot opens the curtains.

SLEEC

Getting the ground truth can be hard.

Experts often disagree.

Guidance is not consistent.

Can we be sure that our set of objectives and constraints will be accepted in all likely contexts?

UKRI AI Centre for Doctoral Training in Safe Artificial INtelligence Systems (SAINTS)

The SAINTS CDT is the UK's first multidisciplinary PhD programme focused solely on the safety of artificial intelligence (AI).

Our vision is to train future leaders with the research expertise and skills to ensure that the benefits of AI systems are realised without introducing harm as the systems and their environments evolve.

Research will be focused on the lifelong safety assurance of increasingly autonomous AI systems in dynamic and uncertain

Contact us

SAINTS Administration Team
SAINTS Centre for Doctoral Training

saints-cdt-admissions@york.ac.uk
+44 (0)1904 325412

UNIVERSITY of York

Centre *for* Assuring Autonomy

SAINTS

Life-long safety of AI: Safety-driven design and training for evolving contexts; testing for open and uncertain operating environments; safe retraining and continual learning; proactive monitoring procedures and dynamic safety cases; ongoing assurance of societal and ethical acceptability.

Safety of increasingly autonomous AI: Understanding human-AI interaction to design safe joint cognitive systems; the assurance of safe transition between human and AI control; achieving effective human oversight and AI explainability; preserving human autonomy and responsibility.

Assuring safety

Some points to note:

- Safety is not something that can be bolted on after the event.
- Safety is a systems level issue. We can not say a machine learnt algorithm is, in itself, safe.
- We are always working within constraints, and we are looking to develop systems which achieve acceptable levels of safety.

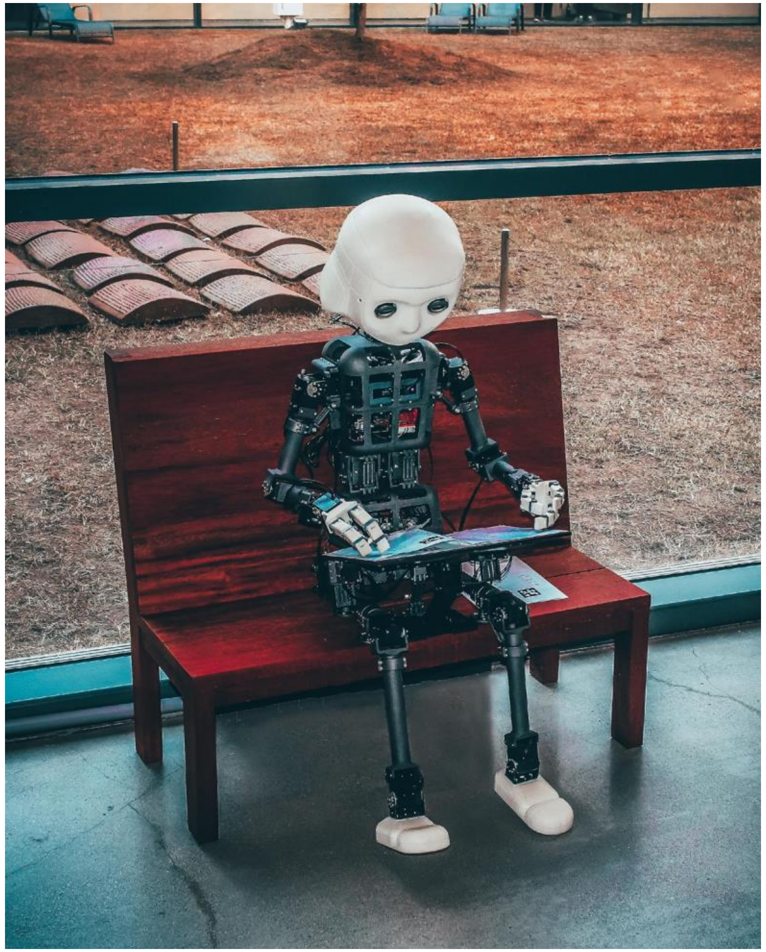
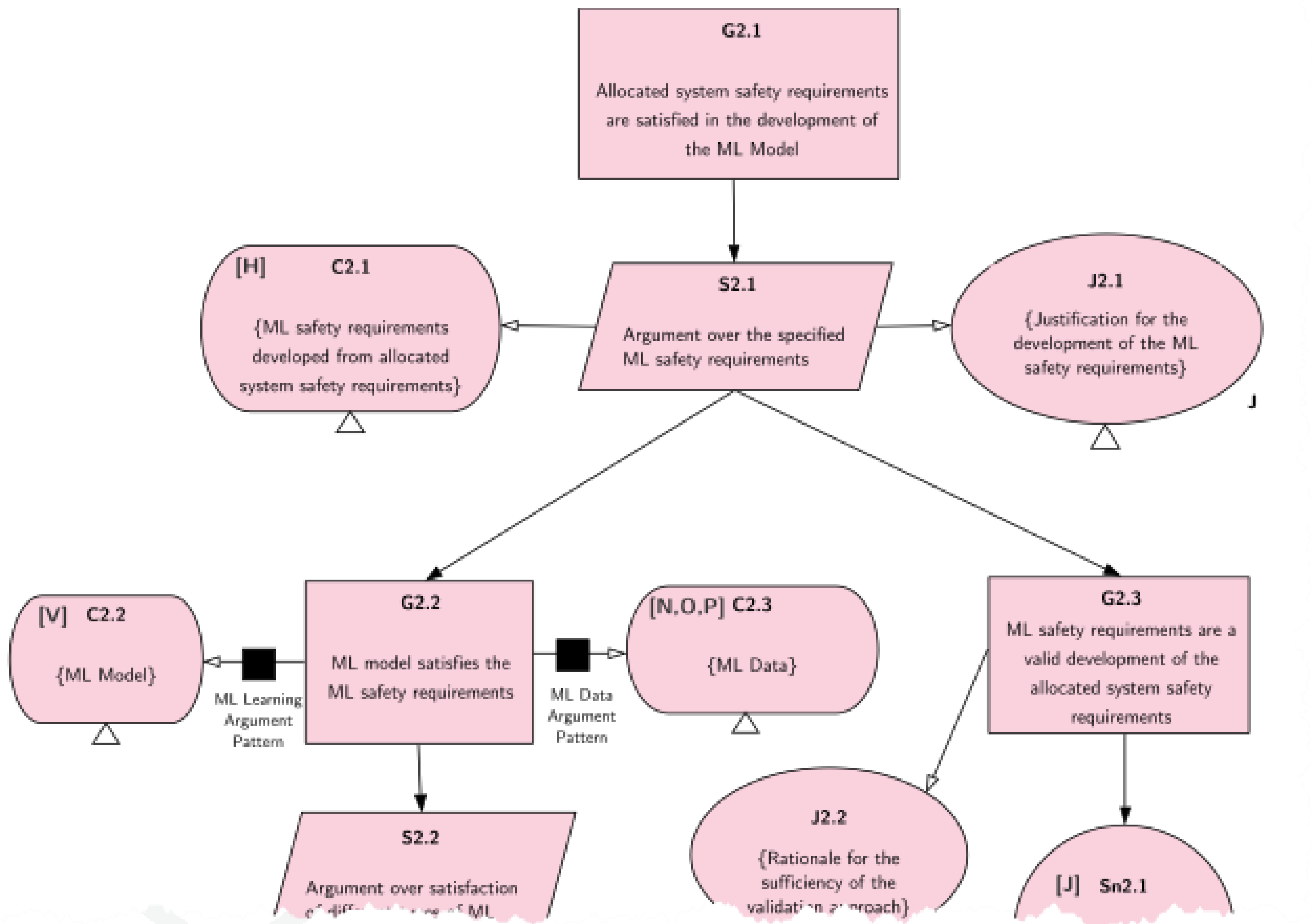


Photo by Andrea De Santis on Unsplash

Assuring Safety

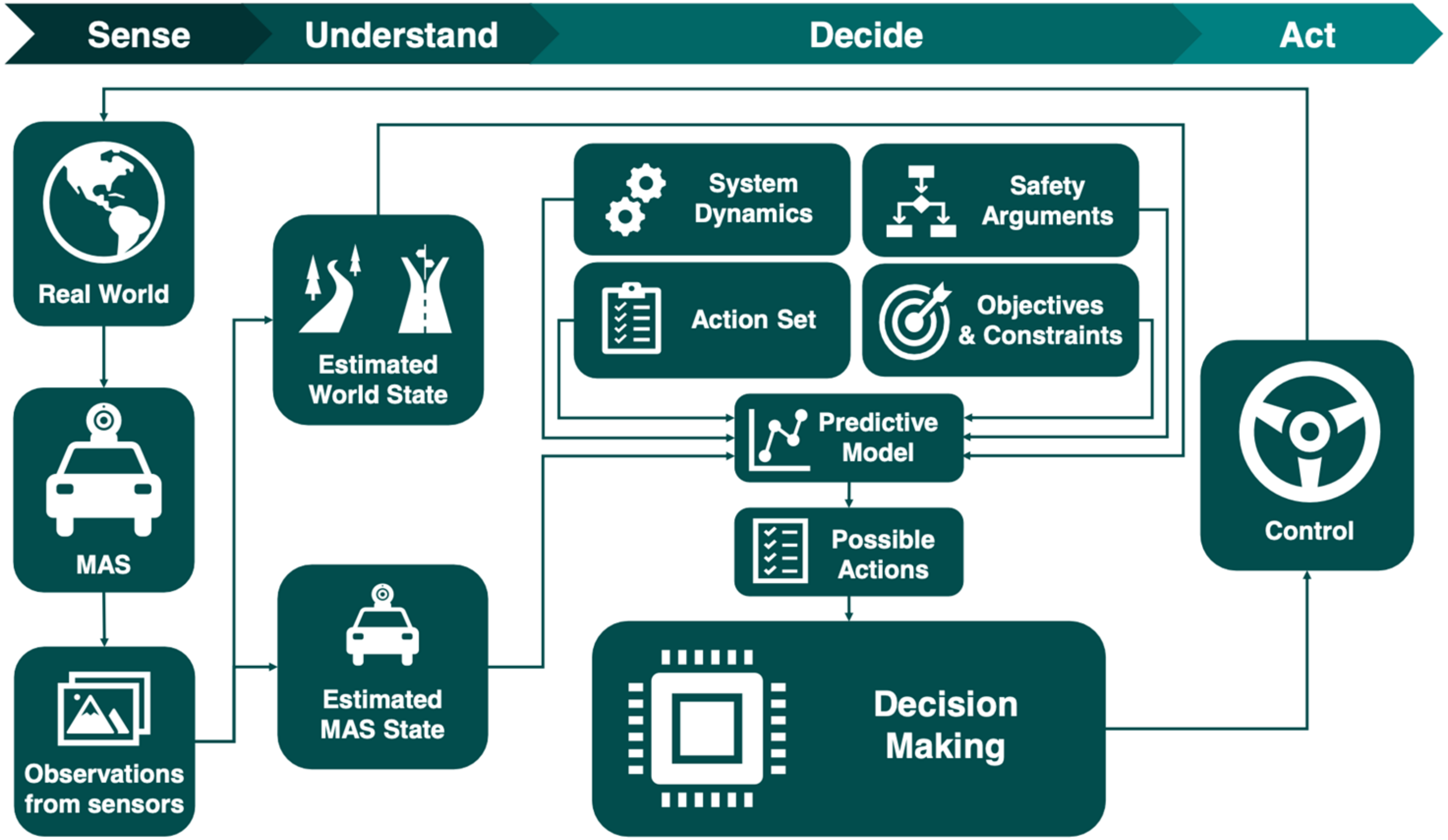


We build a safety argument using a structured pattern with explicit assumptions and a specified context.

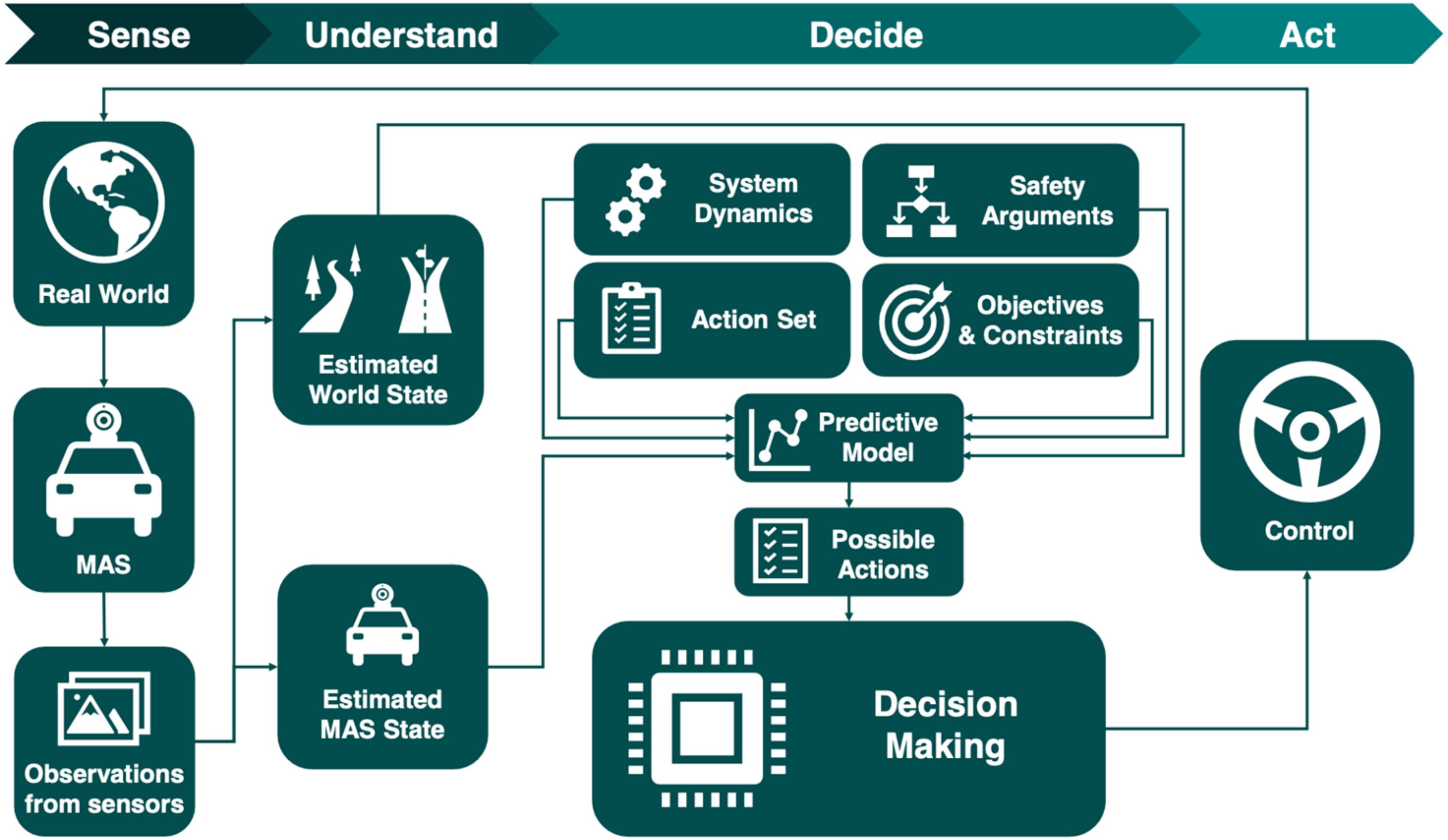
We use evidence from each stage of the development process to support the claims being made.

Arguments may rely on sub claims and arguments.

Decision-Making Framework

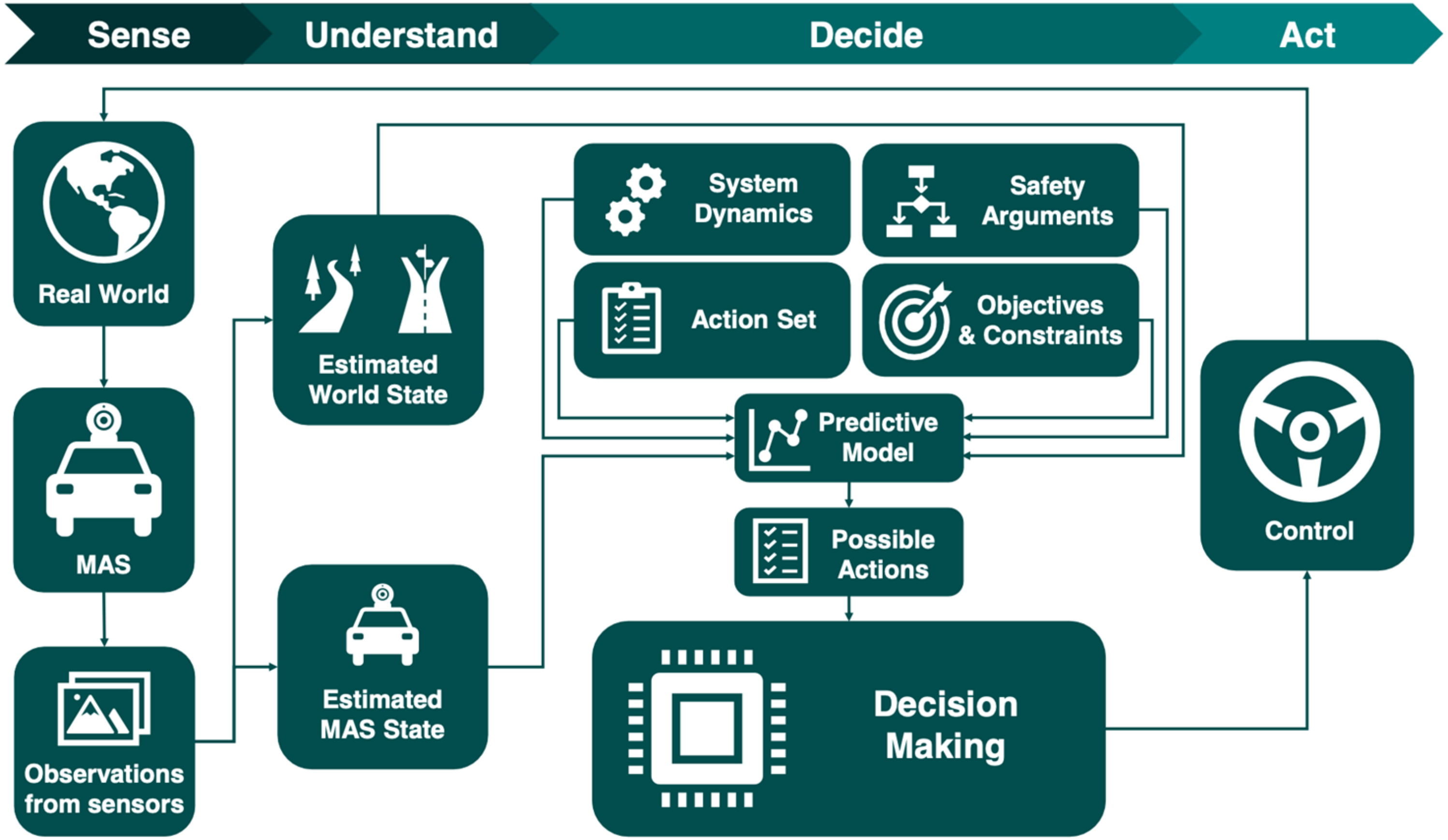


Decision-Making Framework



Designing everyblock requires us to consider the sources, and mitigation strategy, for uncertainty.

Decision-Making Framework



Decision Matrix

Action	Safety	Performance	Time
Accelerate	0.78 – 0.82	0.96 – 0.98	2 sec
Mt. Speed	0.86 – 0.98	0.78 – 0.95	N/A
Slow down	0.86 – 0.94	0.85 – 0.89	3 sec
Stop	0.90 – 0.99	0.76 – 0.81	5 sec

Action	Safety	Performance	Time
Accelerate	0.84 – 0.88	0.96 – 0.98	1 sec
Mt. Speed	0.72 – 0.80	0.78 – 0.95	N/A
Slow down	0.86 – 0.98	0.86 – 0.88	2 sec
Stop	0.92 – 0.96	0.75 – 0.80	4 sec

Action	Safety	Performance	Time
Accelerate	Invalid	Invalid	N/A
Mt. Speed	Invalid	Invalid	N/A
Slow down	0.95 – 0.97	0.85 – 0.87	1 sec
Stop	0.96 – 0.98	0.73 – 0.77	3 sec

Uncertainty, Time and Trade-offs. Safe Decision-making for mobile Autonomous Systems.

Hasan Bin Firoz, Dr. Colin Paterson, Dr. Richard Hawkins

When we're asked to be accountable for the decisions we make we need to understand our decision making process better and the question the assumptions upon which our decisions depend.

Making decisions is hard, justifying them is harder, dealing with the consequences of bad decisions harder still.

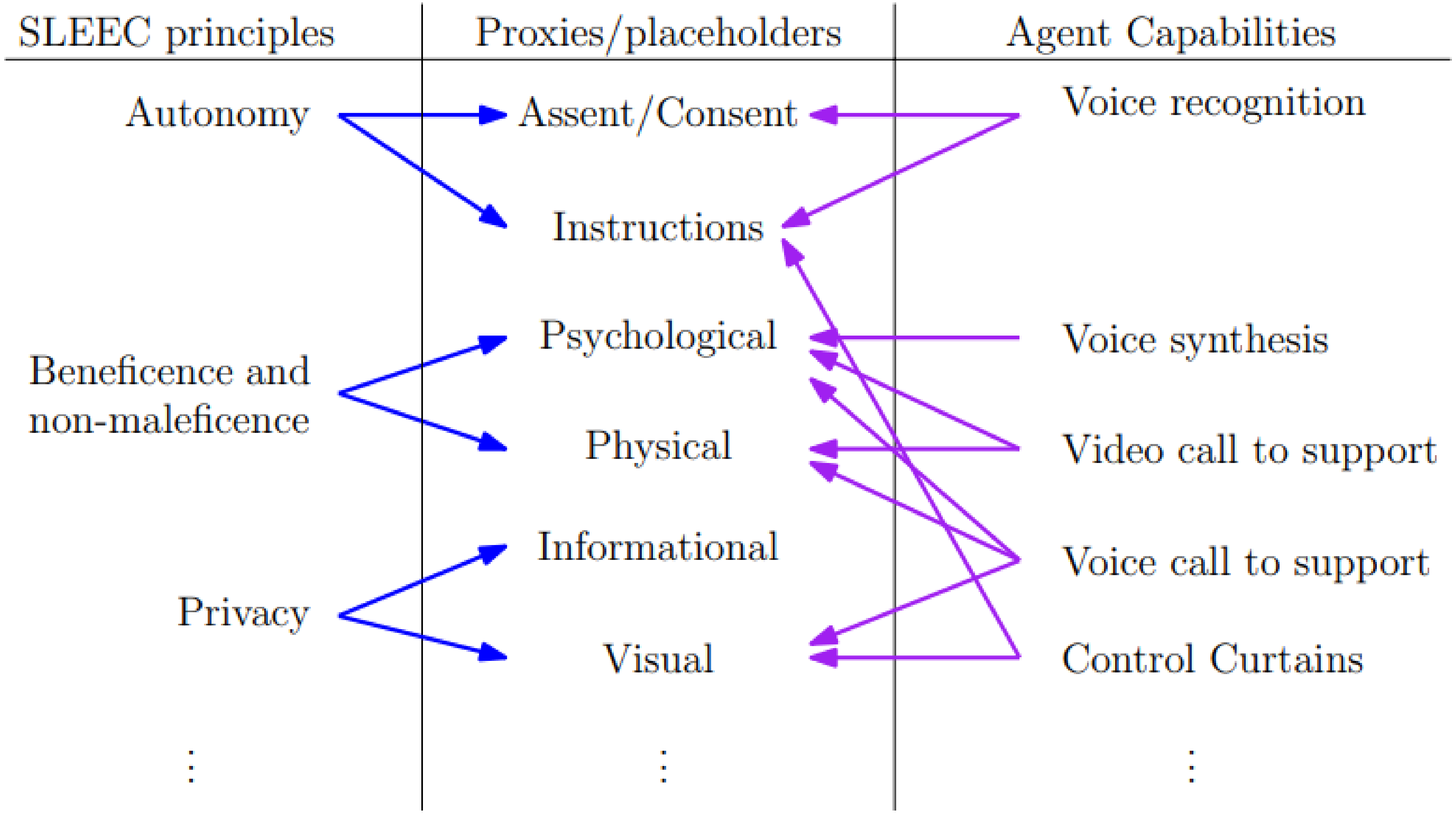
Thank you

AI 'godfather' Yoshua Bengio feels 'lost' over life's work

One of the so-called "godfathers" of Artificial Intelligence (AI) has said he would have prioritised safety over usefulness had he realised the pace at which it would evolve.

BBC News June 2023

SLEEC Proxy Mapping



Time for a break!

Medical and AI Articulate

Articulate

- In groups split into 2 play Articulate
- The goal is to explain what the piece of paper says without saying the word(s) on the paper
- The team that guesses right first wins a point (they can have the piece of paper as the counter)

Next month:

Wednesday 25th September

Speaker: Xinyi Wang – Brain MRI Segmentation

Location: TBC

Fancy more networking?

Head over to the Pub