

Game Theory Meets AI

A Logical and Mathematical Approach to Verifying Multi-Agent Systems

Muhammad Najib

ICAMSAC, 21 Nov 2023

School of Mathematical and Computer Sciences
Heriot-Watt University, UK

1. Ubiquity
2. Interconnection
3. Delegation
4. Human Orientation
5. Intelligence

1. Ubiquity

- 2. Interconnection
- 3. Delegation
- 4. Human Orientation
- 5. Intelligence

- Computing systems are everywhere (Moore's law: small, low-power, inexpensive CPUs).
- Computing systems embedded in devices around us: Roomba, smart fridge, Alexa,...

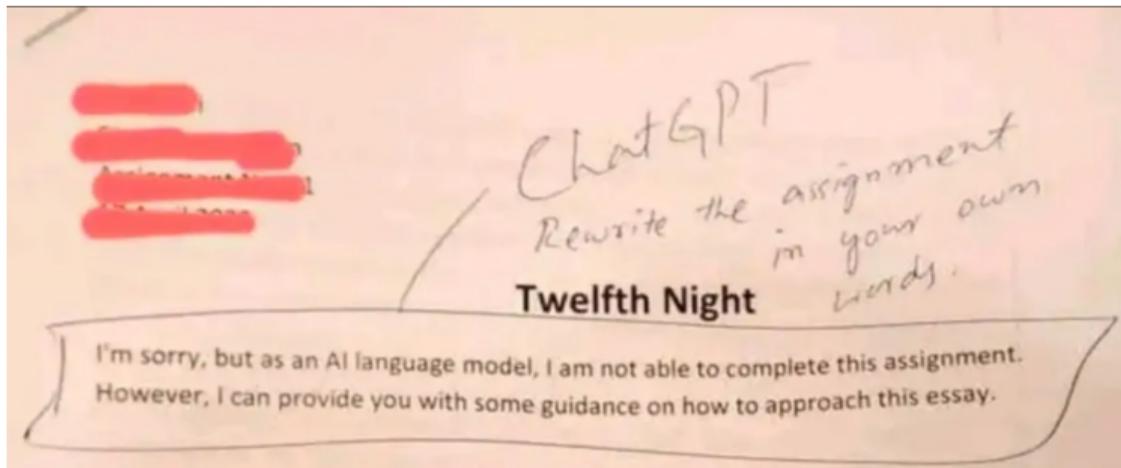
1. Ubiquity
2. **Interconnection**
 - Computer systems connected with one another.
 - e.g., internet
3. Delegation
4. Human Orientation
5. Intelligence

1. Ubiquity
2. Interconnection
3. **Delegation**
 - Computers do things for us (we let them take control).
 - Fly-by-wire planes, autonomous cars, ...
4. Human Orientation
5. Intelligence

Five Trends in Computing

- Computers do things for us (we let them take control).
- Fly-by-wire planes, autonomous cars, ...

1. Ubiquity
2. Interconnection
3. **Delegation**
4. Human Orientation
5. Intelligence



- Many computer systems are designed to interact with humans.
- We interact with them like with humans (Alexa, Siri,...).

1. Ubiquity
2. Interconnection
3. Delegation
4. **Human Orientation**
5. Intelligence

- Many computer systems are designed to interact with humans.
- We interact with them like with humans (Alexa, Siri,...).

1. Ubiquity
2. Interconnection
3. Delegation
4. **Human Orientation**
5. Intelligence

The screenshot shows a conversation on the [r/ChatGPT](#) subreddit. The post by user [altinbeck](#) from 3 months ago features the bold text: "always thank Chat GPT in case of a robot revolution". Below it, a comment from user [JODLER](#) says: "be nice even if its a AI". The AI's response is: "thanks sir gpt for the good work". The AI continues: "You're very welcome! If you have any more questions or need assistance in the future, feel free to reach out. Have a great day, and good luck with your roleplay or any other endeavors you pursue!"

1. Ubiquity
2. Interconnection
3. Delegation
4. Human Orientation
5. **Intelligence**
 - Data + Compute Power + Algorithm & Engineering
 - AI systems become smarter, more capable.

Manifestations:

1. Ubiquity
2. Interconnection
3. Delegation
4. Human Orientation
5. Intelligence

- Cloud computing
- Internet of Things
- Ubiquitous computing
- Semantic Web
- ...
- **Multi-agent systems**

What is an Agent?

“... a computer system that is capable of independent (**autonomous**) **action on behalf of its user.**”^a

^aMichael Wooldridge. *An Introduction to Multiagent Systems*. 2nd ed. Chichester, UK: Wiley, 2009.

“... an **autonomous** entity which observes and **acts** upon an environment and directs its activity **towards achieving goals.**”^a

^aStuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson, 2020. URL:
<http://aima.cs.berkeley.edu/>.

Example of an Agent



Make a call

"Hey Siri, call Mom."

"Hey Siri, call Vivek's mobile on speakerphone."

Siri can also make and answer calls on HomePod >



Get directions

"Hey Siri, find coffee near me."

"Hey Siri, get directions home."

Use Siri with CarPlay >

Now ask Siri to ...



Send a message

"Hey Siri, send a message to Ming Lu."

"Hey Siri, text Adrian and Sofia, 'Where are you?'"

Siri can read new messages on your AirPods >



Play music

"Hey Siri, play the hottest Taylor Swift tracks."

"Hey Siri, play the new Tame Impala album."

Learn more ways to play music >



Find information

"Hey Siri, what's the weather for today?"

"Hey Siri, how high is Mount Everest?"

Learn more things you can ask Siri >



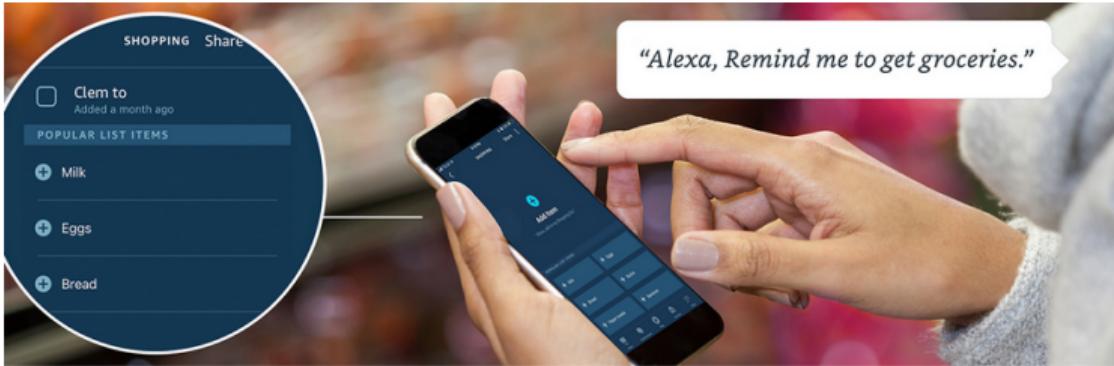
Find your Apple device

"Hey Siri, where's my iPhone?"

"Hey Siri, find my AirPods."

Learn how to use Find My >

Example of an Agent



Try saying

- *"Alexa, set a recurring alarm for 7 AM."*
- *"Alexa, what's on my calendar for today?"*
- *"Alexa, schedule a meeting with Jeff."*
- *"Alexa, remind me to call mom on Saturday at 2 PM."*
- *"Alexa, remind me to get groceries when I get home."*

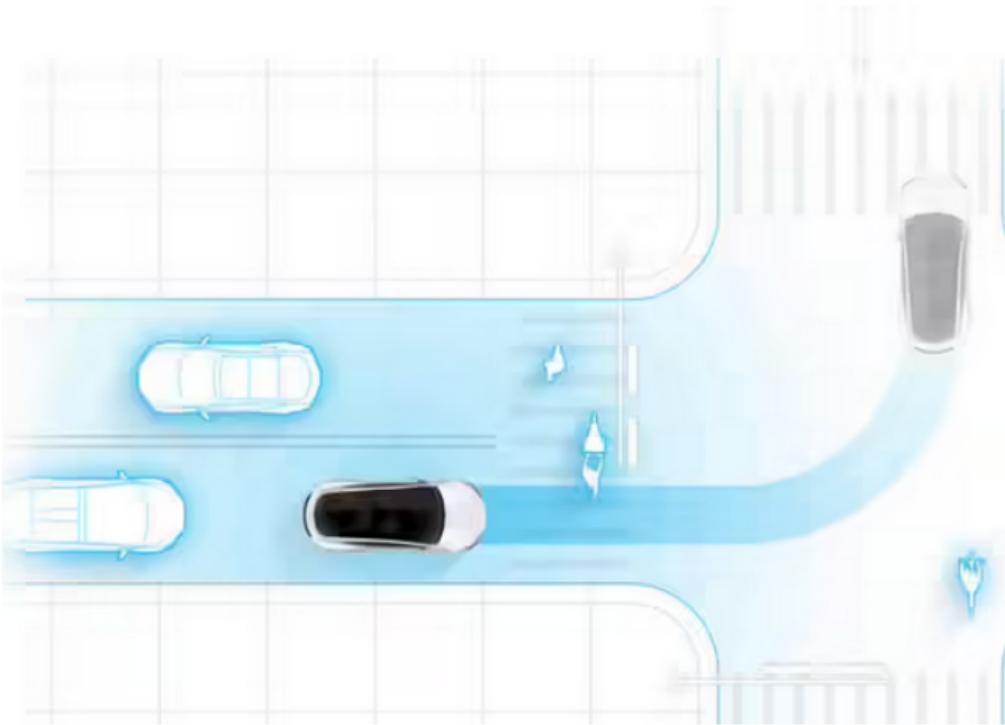
Example of an Agent



T E S L A

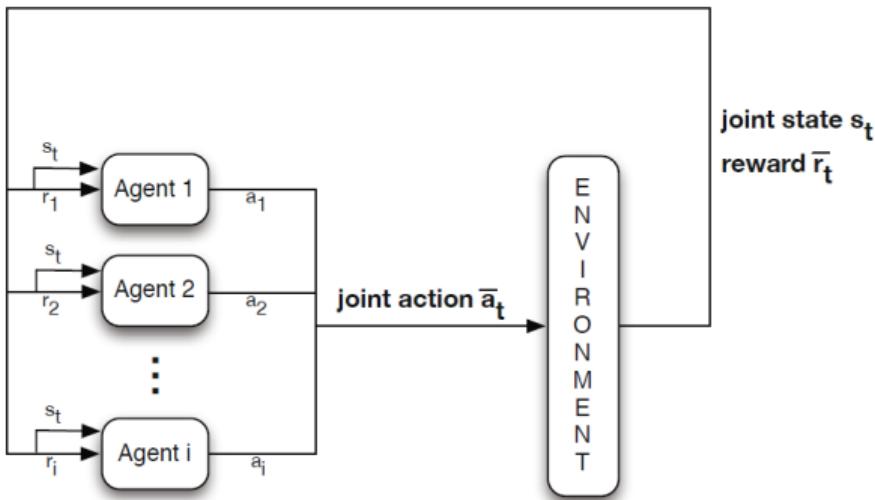
From Home

All you will need to do is get in and tell your car where to go. If you don't say anything, your car will look at your calendar and take you there as the assumed destination. Your Tesla will figure out the optimal route, navigating urban streets, complex intersections and freeways.



What is a Multi-Agent System?

- A system consists of **multiple agents** that **interact** with one another.
- Agents **act** on behalf of users/stakeholders with **different goals and preferences**.
- They interact and act upon the **environment**.



Source: Nowe, Ann & Vrancx, Peter & De Hauwere, Yann-Michaël. (2012). Game Theory and Multi-agent Reinforcement Learning.

Example of a Multi-Agent System

- Algorithmic/high-frequency trading.
- Trading softwares **buy & sell** stocks to **generate as much money as possible**.

J.P.Morgan

[Solutions](#) > [Corporate & Investment Banking](#) > [Markets](#) > [Execute](#) > FX Algos Execute



FX Algos on Execute

Electronic trading solutions available on J.P. Morgan Markets

Problem with Multi-Agent Systems

- MAs are prone to **instability** and might have **unpredictable dynamics**.
- Or, some stable behaviour gives rise to **bad outcomes**.
- 2010 Flash Crash^a: over a 30 minutes period, Dow Jones lost (momentarily) over a trillion dollars of valuation.
 - "...the interaction between automated execution programs and algorithmic trading strategies can quickly erode liquidity and result in disorderly markets."^b



^a<https://www.theguardian.com/business/2015/apr/22/2010-flash-crash-new-york-stock-exchange-unfolded>

^bU.S. Securities and Exchange Commission; Commodity Futures Trading Commission. "Findings Regarding the Market Events of May 6, 2010"

Problem with Multi-Agent Systems

News

Opinion

Sport

Culture

Lifestyle

MOT

UK World Climate crisis Ukraine Football Newsletters Business Environment UK politics Education Society Sci

Self-driving cars

Cruise recalls all self-driving cars after grisly accident and California ban

All 950 of the General Motors subsidiary's autonomous cars will be taken off roads for a software update

Associated Press

Wed 8 Nov 2023 18.17 GMT



- With *safety critical* systems (e.g., autonomous cars), not only we risk losing money but human lives.

Problem with Multi-Agent Systems

News

Opinion

Sport

Culture

Lifestyle

MOT

UK World Climate crisis Ukraine Football Newsletters Business Environment UK politics Education Society Sci

Self-driving cars

Cruise recalls all self-driving cars after grisly accident and California ban

All 950 of the General Motors subsidiary's autonomous cars will be taken off roads for a software update

Associated Press

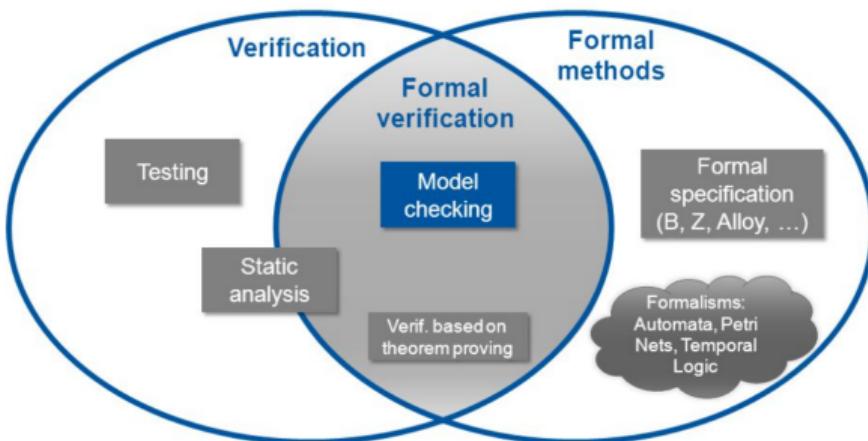
Wed 8 Nov 2023 18.17 GMT



- With *safety critical* systems (e.g., autonomous cars), not only we risk losing money but human lives.

We want our AI (multi-agent) systems to be ‘**CORRECT**’

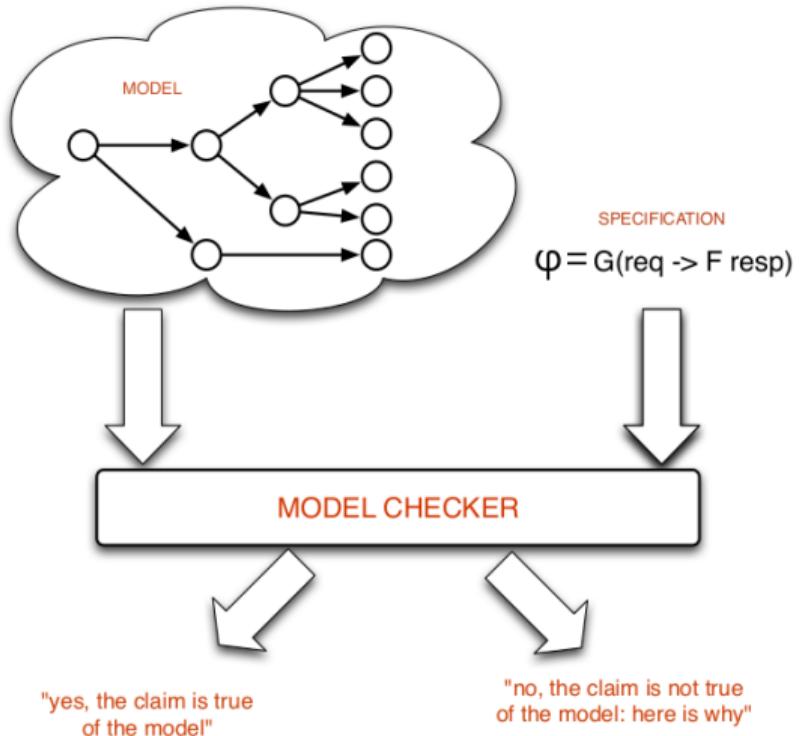
- The **correctness problem** has been one of the most widely studied problems in computer science over the past fifty years, and remains a topic of fundamental concern to the present day
- the correctness problem: checking that computer systems behave as their designer intends
- **Formal verification** is the problem of checking that a system P is *correct* with respect to a formal specification φ (e.g., LTL)



- Standard formal language for talking about (infinite) state sequences
- Has been around for more than four decades¹
- Propositional logic ($\wedge, \vee, \neg, \dots$) + temporal modalities (G, F, X, \dots)
 - Gp : is always the case that p
 - Fq : will eventually the case that q
- We can express something like:
 - “*it is always not cold in Bali*”: $G\neg cold$
 - “*eventually will rain in Denpasar*”: $F rain$

¹Amir Pnueli. “The temporal logic of programs”. In: *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*. ieee. 1977, pp. 46–57.

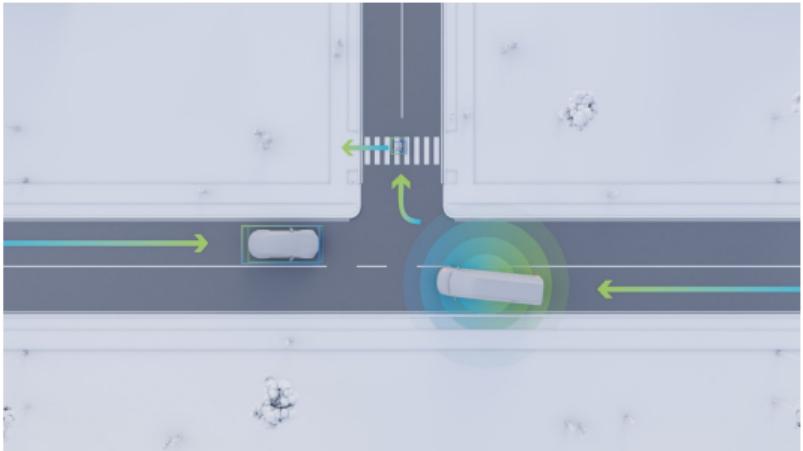
(LTL) Model Checking



Very influential: 4 Turing Award Winners

1996	Amir Pnueli		For seminal work introducing temporal logic into computing science and for outstanding contributions to program and systems verification . ^[35]
2007	Edmund M. Clarke		For their roles in developing model checking into a highly effective verification technology, widely adopted in the hardware and software industries. ^[38]
	E. Allen Emerson		
	Joseph Sifakis		

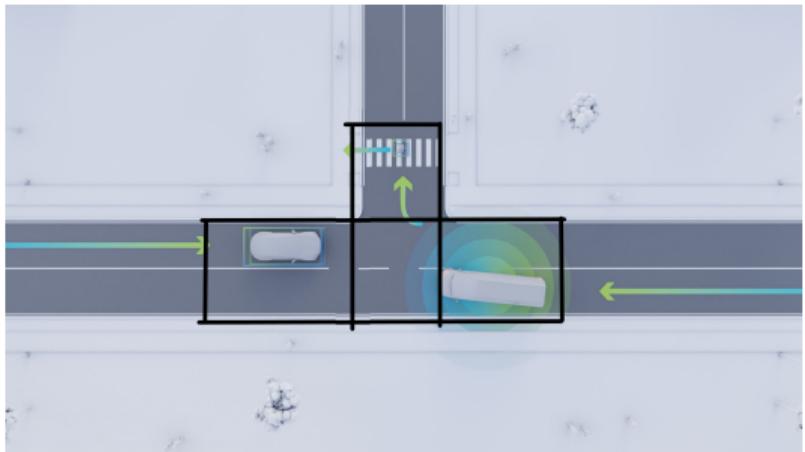
From Scenario to Model Checking



Source: <https://www.digitrans.expert/en>

- Two autonomous vehicles are approaching a junction.
- One is turning, the other one is going straight.
- We want: “*avoid collisions*”
- Once a collision occurs, the vehicles cannot continue their journey

From Scenario to Model Checking



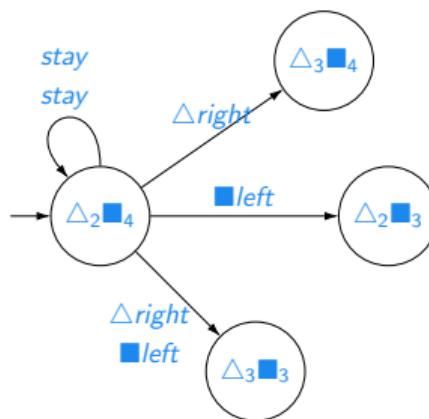
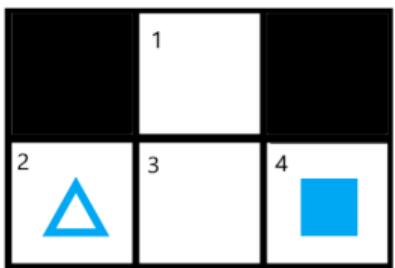
- Abstracting → discretising
- “*avoid collisions*”: $\text{G} \neg \text{collide}$

Source: <https://www.digitrans.expert/en>

From Scenario to Model Checking

“avoid collisions”: $\text{G} \neg \text{collide}$, where *collide* means \triangle and \blacksquare are **in the same location**

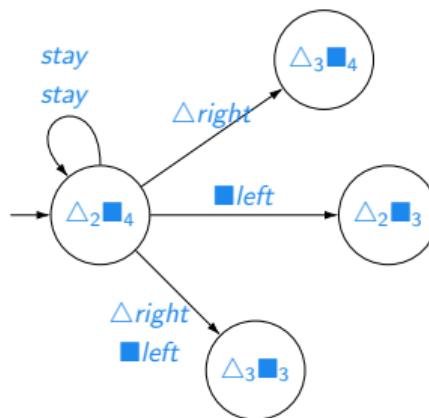
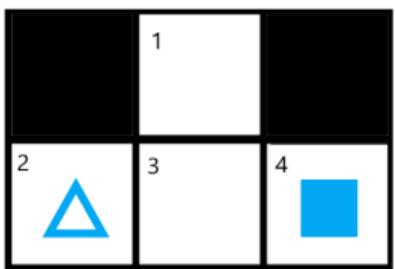
$$\varphi := \text{G} \neg \bigvee_{i \in \{1,2,3,4\}} (\triangle_i \wedge \blacksquare_i)$$



From Scenario to Model Checking

“avoid collisions”: $\text{G} \neg \text{collide}$, where *collide* means \triangle and \blacksquare are **in the same location**

$$\varphi := \text{G} \neg \bigvee_{i \in \{1,2,3,4\}} (\triangle_i \wedge \blacksquare_i)$$

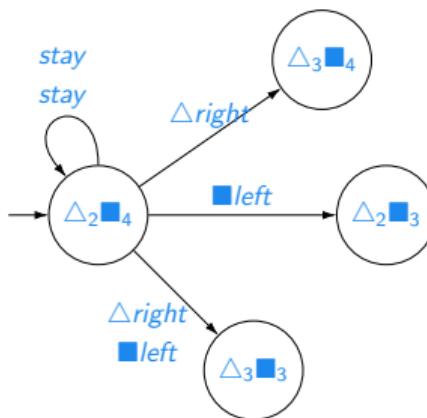
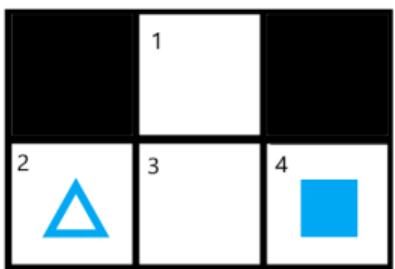


φ is **violated** since it is *possible* to reach the state $\triangle_3 \blacksquare_3$

From Scenario to Model Checking

“avoid collisions”: $\text{G} \neg \text{collide}$, where *collide* means \triangle and \blacksquare are **in the same location**

$$\varphi := \text{G} \neg \bigvee_{i \in \{1,2,3,4\}} (\triangle_i \wedge \blacksquare_i)$$



φ is **violated** since it is *possible* to reach the state $\triangle_3 \blacksquare_3$

Is this **reasonable**?

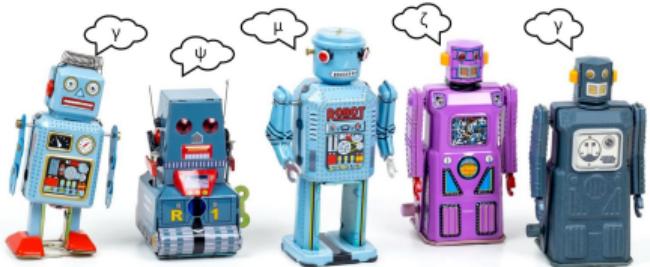


Source: <https://www.digitrans.expert/en>

- A collision is a **possible** behaviour.
- However, not a **rational** behaviour.
- The vehicles would **prefer** to **avoid** a collision: wait for the other vehicle to pass, then continue to its destination
- Classical verification is not a good/reasonable approach to check the correctness of such a scenario.

Problem with Classical Notion of Correctness Problem

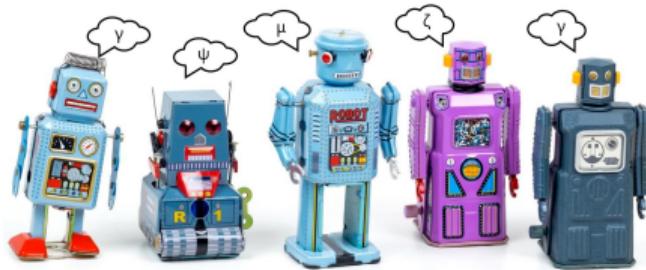
How should we define correctness in MAs?



Classical notion of correctness ignores agents **goals/preferences**

A New Notion of Correctness Problem

How should we define correctness in MAs?



Correctness with respect to **rational choices** of agents

Classical Verification

Is the system correct?



Rational Verification

Is the system correct wrt behaviours that can be **sustained by rational choices** of agents?

- Use **game theory** to model/analyse rational behaviours.
- Turn MASs into **multi-player games**.

²Alessandro Abate et al. "Rational verification: game-theoretic verification of multi-agent systems". In: *Applied Intelligence* 51.9 (2021), pp. 6569–6584.

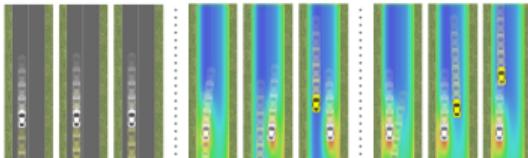
- Games serve as **abstractions** for *strategic interactions* between self-interested players/agents
- Various settings: *turn-based vs concurrent, zero-sum vs general-sum, cooperative vs non-cooperative*, ...
- Relevant for many scenarios in autonomous/AI systems
 - e.g., zero-sum: DeepMind AlphaZero (go, chess, shogi playing), concurrent: resource sharing/allocation (server, GPU power),...
 - even autonomous vehicles

2019 International Conference on Robotics and Automation (ICRA)
Palais des congrès de Montréal, Montréal, Canada, May 20-24, 2019

Hierarchical Game-Theoretic Planning for Autonomous Vehicles

Jaime F. Fisac^{*1} Eli Bronstein^{*1} Elis Stefansson² Dorsa Sadigh³ S. Shankar Sastry¹ Anca D. Dragan¹

Abstract— The actions of an autonomous vehicle on the road affect and are affected by those of other drivers, whether overtaking, negotiating a merge, or avoiding an accident. This mutual dependence, best captured by dynamic game theory, creates a strong coupling between the vehicle's planning and its predictions of other drivers' behavior, and constitutes an open problem with direct implications on the safety and viability of



Ingredients:

1. Several decision makers (**the players/agents**)
2. Players have different goals (**the goals**)
3. Each player can affect the outcome for all (**the actions**)

Ingredients:

1. Several decision makers (**the players/agents**)
2. Players have different goals (**the goals**)
3. Each player can affect the outcome for all (**the actions**)

Game theory

the methodology of using mathematical tools to model and analyse situations of interactive decision making.

Ingredients:

1. Several decision makers (**the players/agents**)
2. Players have different goals (**the goals**)
3. Each player can affect the outcome for all (**the actions**)

Game theory

the methodology of using mathematical tools to model and analyse situations of interactive decision making.

vs decision theory

The **interactivity** distinguishes game theory from standard decision theory, which involves a single decision maker.

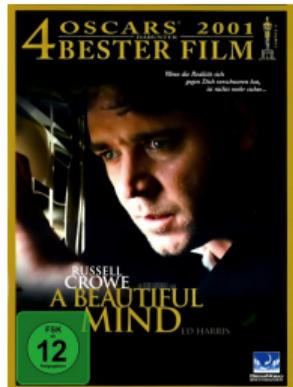
How to model rational behaviours?

- What kind of behaviour is **rational**?
- Game theory proposes many “solution concepts”, i.e., a formal rule for ‘predicting’ how a game will be played
- The most influential is **Nash equilibrium**: Nobel prize in Economics 1994



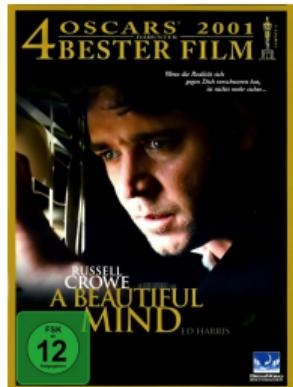
How to model rational behaviours?

- What kind of behaviour is **rational**?
- Game theory proposes many “solution concepts”, i.e., a formal rule for ‘predicting’ how a game will be played
- The most influential is **Nash equilibrium**: Nobel prize in Economics 1994



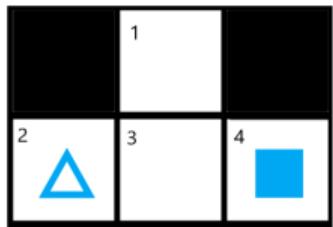
How to model rational behaviours?

- What kind of behaviour is **rational**?
- Game theory proposes many “solution concepts”, i.e., a formal rule for ‘predicting’ how a game will be played
- The most influential is **Nash equilibrium**: Nobel prize in Economics 1994



Nash equilibrium

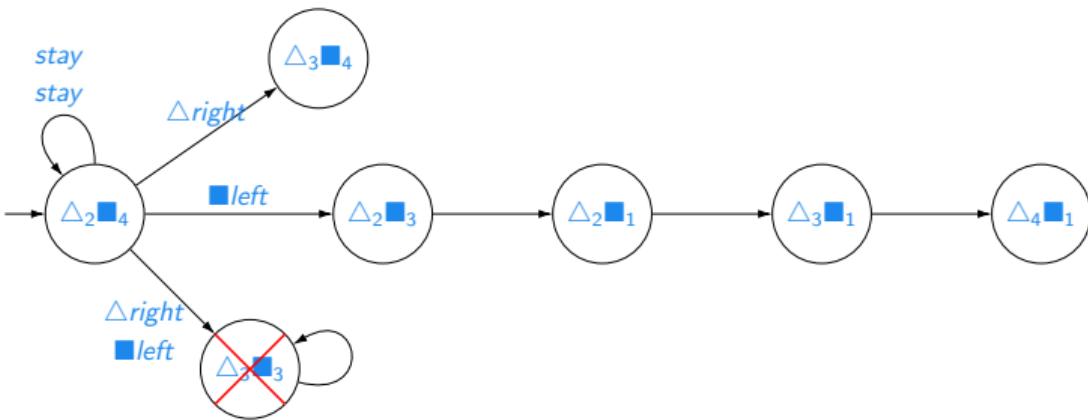
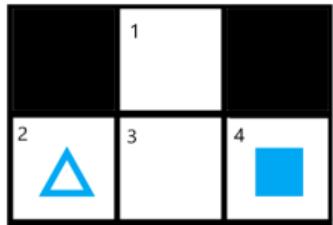
A situation where no player in a game would want to change their strategy, while keeping the other players' strategies constant



- the players: Δ, \blacksquare
- the goals:
 - Player Δ wants to go straight: $\gamma_\Delta := F\Delta_4$
 - Player \blacksquare wants to turn: $\gamma_\blacksquare := F\blacksquare_1$
- the actions: players can move to adjacent locations

Modelling Rationality in a Game

$$\varphi := \mathbf{G} \neg \bigvee_{i \in \{1,2,3,4\}} (\Delta_i \wedge \blacksquare_i) \quad \gamma_\Delta := \mathbf{F} \Delta_4 \quad \gamma_\blacksquare := \mathbf{F} \blacksquare_1$$

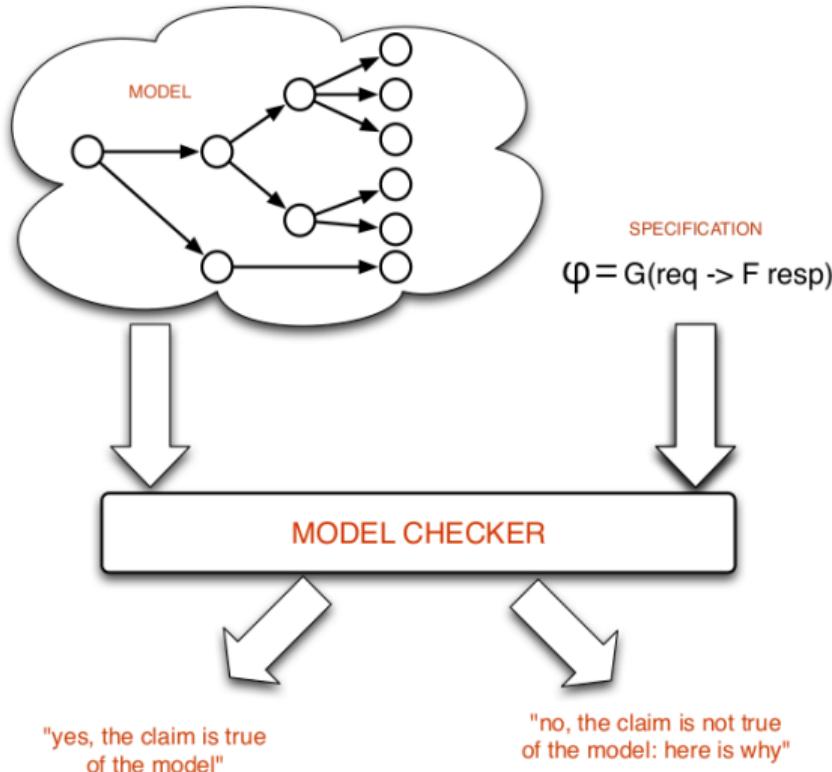


Δ moves: right, right and \blacksquare moves: left, up

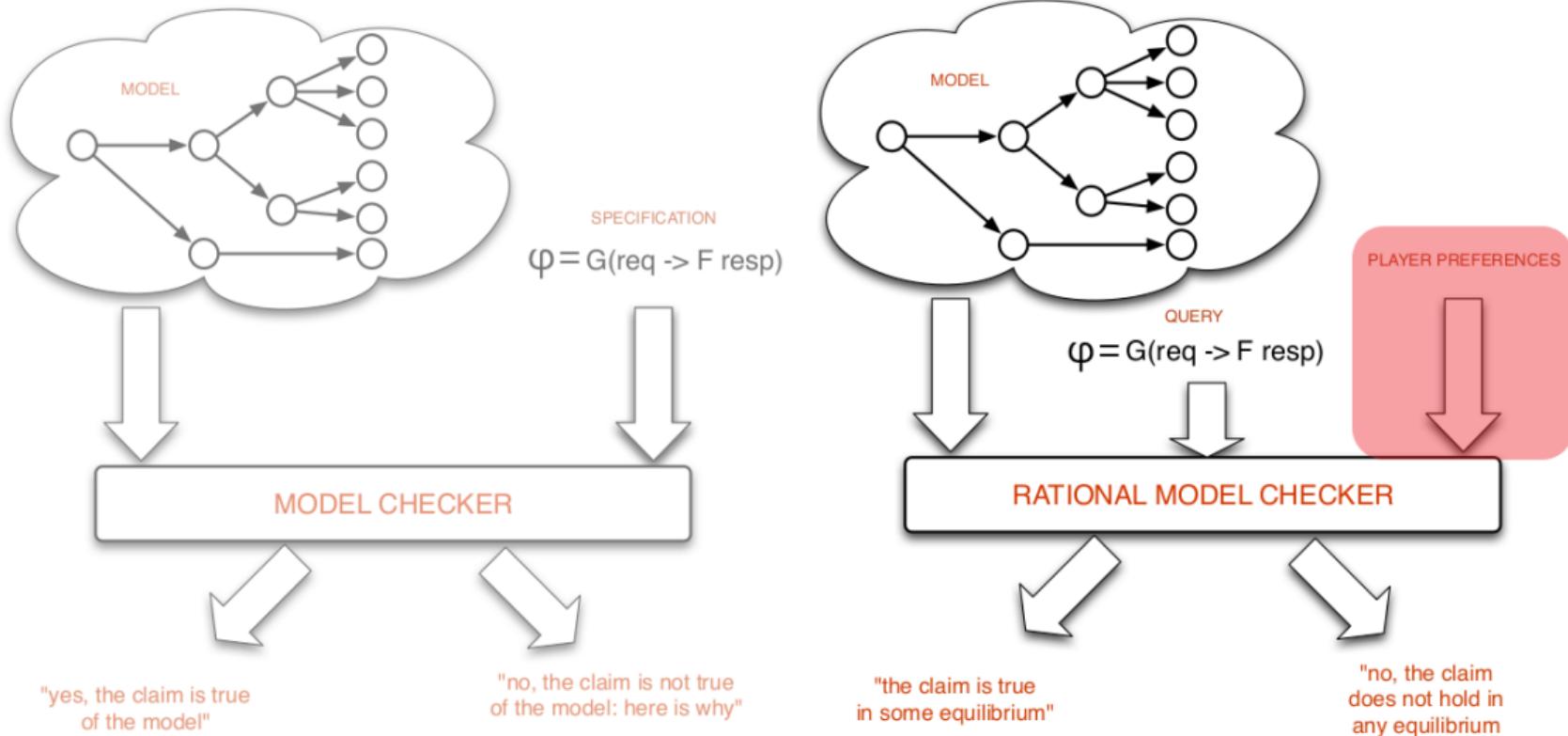
Not a NE, since (for example) Δ can stay put and wait for \blacksquare to go up, then proceed to move right, right

In fact $\Delta_3 \blacksquare_3$ will **never** be reached under *strong* NE! Under the strong NE assumption, the formula φ is **not** violated!!

From Verification to Rational Verification



From Verification to Rational Verification



- **Safety:** all stable outcomes (e.g., NE) do **not violate** a desirable property φ (A-NASH)
- **Liveness:** there **exists** a stable outcome that satisfies a desirable property φ (E-NASH)
- **Stability:** Is there any stable outcome? (NON-EMPTINESS)

Software

The VAS group is actively maintaining a number of open-source software packages, including:

MCMAS

MCMAS is an open-source, OBDD-based symbolic model checker tailored to the verification of Multi-Agent Systems (MAS). It is given by means of ISPL (Interpreted Systems Programming Language) programs. ISPL is an agent-based, modular language for interpreted systems, a popular semantics in MAS.

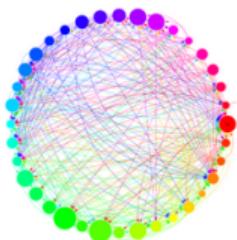
- Model Checker for Multi-Agent Systems (MCMAS)³
- Open source: <https://vas.doc.ic.ac.uk/software/mcmas/>
- OBDD-based symbolic techniques; can reduce the size of models
- Only support memoryless/Markovian strategies

³A. Lomuscio, H. Qu, and F. Raimondi. "MCMAS: An Open-source Model Checker for the Verification of Multi-Agent Systems". In: *STTT* (2017).

- Equilibrium Verification Environment (EVE)⁴
- Automata-theoretic techniques
- Support memoryful strategies; players can fully implement LTL goals
- EVE online: <http://eve.cs.ox.ac.uk/>

Home EVE Download People Publications Others

Welcome to EVE Website



EVE (Equilibrium Verification Environment) is a formal verification tool for the automated analysis of temporal equilibrium properties of concurrent and multi-agent systems represented as multi-player games. Systems are modelled using the Simple Reactive Module Language (SRML) as a collection of independent system components (players/agents in a game), which are assumed to have goals expressed using Linear Temporal Logic (LTL) formulae. In particular, EVE checks for the existence of Nash equilibria in such systems and can be used to do rational synthesis and verification automatically.

⁴ Julian Gutierrez et al. "Automated temporal equilibrium analysis: Verification and synthesis of multi-player games". In: *Artificial Intelligence* (2020).

Non-Emptiness Experiment Result⁵

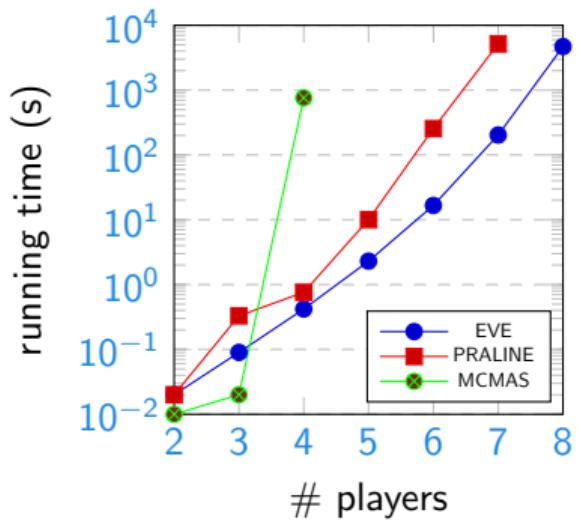


Figure 1: Running time for NON-EMPTINESS Gossip Protocol.

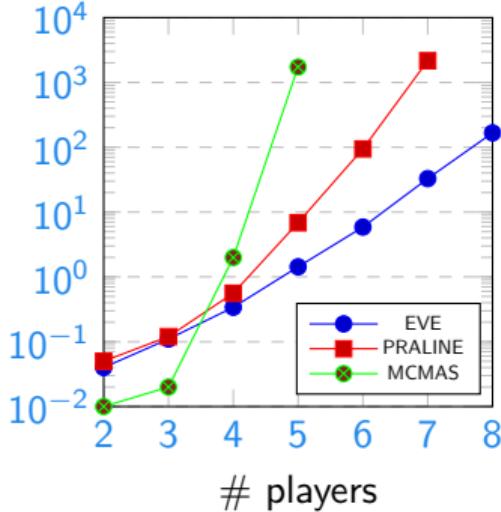


Figure 2: Running time for NON-EMPTINESS Replica Control Protocol.

⁵ Julian Gutierrez et al. "EVE: A Tool for Temporal Equilibrium Analysis". In: ATVA. 2018.

E-Nash Experiment Result⁶

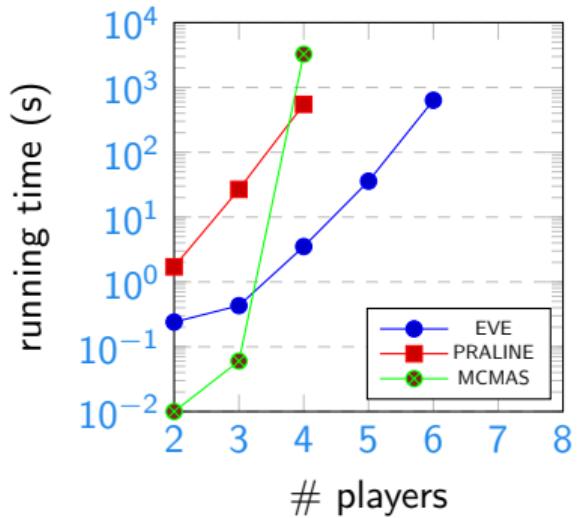


Figure 3: Running time for E-NASH Gossip Protocol.

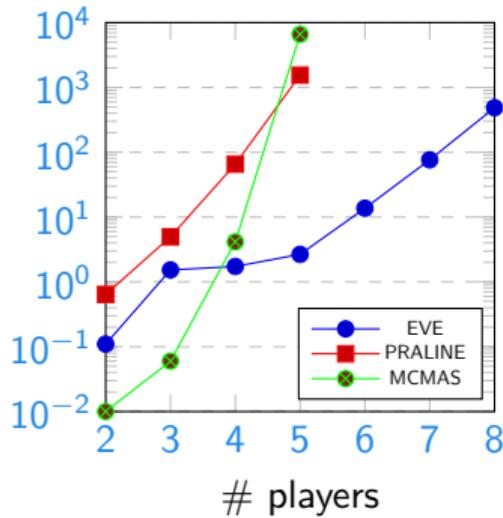


Figure 4: Running time for E-NASH Replica Control Protocol.

⁶Gutierrez et al., "EVE: A Tool for Temporal Equilibrium Analysis".

A-Nash Experiment Result⁷

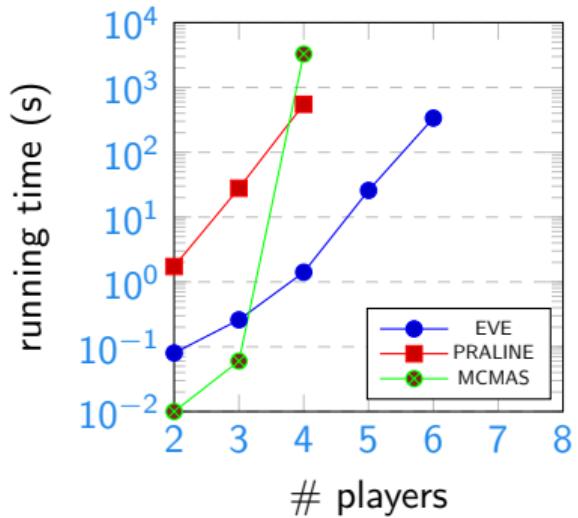


Figure 5: Running time for A-NASH Gossip Protocol.

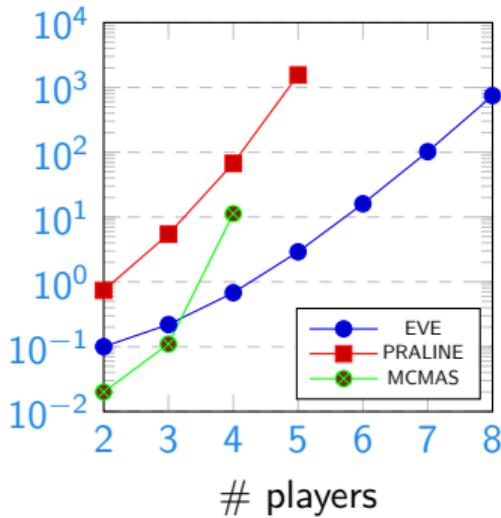


Figure 6: Running time for A-NASH Replica Control Protocol.

⁷ Gutierrez et al., "EVE: A Tool for Temporal Equilibrium Analysis".

Where next for rational verification?

- Decision Problems with LTL are expensive: 2EXPTIME, although restricting to fragments of LTL can bring them down to NP or even PTIME⁸
- **Statistical methods:** can these make it more practical? E.g., model checking with the Monte Carlo method⁹
- **Learning agents:** What if the players use some learning element, e.g., reinforcement learning?¹⁰
- **Privacy & security:** So far the setting has been *perfect information*. What if this is not a viable setting? For instance, we might not want other vehicles to know our home address.
- **Explainability:** In the *synthesis* of rational strategies (*rational synthesis*), e.g., autonomous vehicle route planning, how can we make the strategies transparent to human?

⁸ Julian Gutierrez et al. "On Computational Tractability for Rational Verification". In: *IJCAI*. 2019, pp. 329–335.

⁹ Radu Grosu and Scott A Smolka. "Monte carlo model checking". In: *TACAS*. 2005.

¹⁰ Lewis Hammond et al. "Multi-Agent Reinforcement Learning with Temporal Logic Specifications". In: *AAMAS*. 2021.

- The future looks increasingly more and more multi-agent
- Want and need these multi-agent systems to be safe and correct
- Verification of Multi-Agent Systems
 - A new and more appropriate notion of correctness: rational verification
 - Modelling systems as games
 - Tools: MCMAS, EVE
- Challenges
 - Practicality and scalability
 - Incorporating agents who learn
 - How to ensure privacy and security?
 - How to make decisions transparent to human?

Thank you!