Towards a Formal Framework for Social Robots with Theory of Mind

Filippos Gouidis[1], Alexandros Vassiliades[1,2], Nena Basina[1] and Theodore Patkos[1]
[1]*Institute of Computer Science, Foundation for Research and Technology, Hellas, Greece*
[2]*School of Informatics, Aristotle University of Thessaloniki, Greece*

Keywords: Social Robotics, Theory of Mind, Epistemic Reasoning, Reasoning about Action, Event Calculus.

Abstract: A key factor of success for future social robotics entities is going to be their ability to operate in tight collaboration with non-expert human users in open environments. Apart from physical skills, these entities will have to exhibit intelligent behavior, in order both to understand the dynamics of the domain they inhabit and to interpret human intuition and needs. In this paper, we discuss work in progress towards developing a formal framework for endowing intelligent autonomous agents with advanced cognitive skills, central to human-machine interaction, such as Theory of Mind. We argue that this line of work can lay the ground for both theoretical and practical research, and present a number of areas, where such a framework can achieve essential impact for future social and intelligent systems.

1 INTRODUCTION

Modeling the behavior and the mental state of others is an essential cognitive ability of humans, central to their social interactions. From a very young age, people unconsciously generate meta-representations associated with what others believe, in addition to their own beliefs, and use these comparative mental models when they attempt to make sense or predict the behavior of others (Apperly, 2012). The processes involved in recognizing that people have different mental states, goals and plans, and in inferring others' mental states, is collectively known as Theory of Mind (ToM).

ToM is also crucial for developing autonomous systems that operate in tight collaboration with humans, in order to anticipate their needs and intentions, and proactively respond to future actions. From the Artificial Intelligence (AI) standpoint, the symbiosis of intelligent agents, such as social and companion robots, with humans introduces a multitude of challenges, at the core of which is the modeling of how the world works, what knowledge humans consider commonsense, and which their own abilities -physical or mental- and the abilities of others are (Marcus and Davis, 2019); or, in the language of cognitive psychologists, this means that the agents need to be equipped with a rich cognitive model.

(a) Top view (b) Observer's perspective

Figure 1: A scene observed from different angles generates diverse beliefs about the existence and position of objects.

Motivation

In this paper, we aim to highlight the importance of endowing social agents with ToM, considering scenarios of everyday life. We also present work in progress towards developing a formal, generic framework for generating agents that can reason about knowledge and causality, using an expressive, as well as efficient, in terms of computational complexity, formalization.

Consider the following toy setting that will motivate our analysis in the sequel: Figure 1a shows a desk in a meeting room with laptops and various items scattered around, such as pens, mugs, etc. The persons working at the office, as well as an assistant robot, may change their position around the desk. Let us assume that, from a given moment on, all entities only have a sideways, and not a top-down view of the desk (Figure 1b). Apparently, for the person sitting

Gouidis, F., Vassiliades, A., Basina, N. and Patkos, T.
Towards a Formal Framework for Social Robots with Theory of Mind.
DOI: 10.5220/0010893300003116
In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022) - Volume 3*, pages 689-696
ISBN: 978-989-758-547-0; ISSN: 2184-433X

in front of an open laptop, any item behind the screen is occluded. The robot, positioned at a different angle, should be able to make simple inferences, such as which objects are visible to each person considering their current positions, as well as more complex inferences, such as whether the position of occluded objects is known, due to the previous positions of the persons around the desk. The robot should also appropriately update the different mental states, based on both the physical (ontic) actions that take place, such as that someone picked up the mobile phone, and the epistemic actions, such as announcements or distractions. For instance, a person concentrated on watching a presentation may not notice certain actions, leading to potentially erroneous beliefs.

While the goal to endow agents with at least basic ToM capabilities, rich cognitive models and the capacity to make commonsense inferences, is not new to the field of AI, existing social-cognitive agents either lack such skills or develop ad hoc solutions that are difficult to generalize or verify. In (Chen et al., 2021) for instance, a deep neural network is developed to predict the long term behavior of an actor with ToM using raw video data; the explainability of the outcome or the verification of the process is rather limited though. Classical AI, based on symbolic methods, has long ago devised expressive formalisms that enable an agent to make epistemic inferences about their own mental state (1^{st}-order beliefs) and about the mental state of the others (2^{nd}-order beliefs) in causal domains (e.g., see (D'Asaro et al., 2020; Schwering et al., 2015; Ma et al., 2013; Shapiro et al., 2011; Ditmarsch et al., 2007; Liu and Levesque, 2005; Davis and Morgenstern, 2005; Scherl, 2003)). The majority of such formalisms is based on the possible worlds model, which although elegant in generating expressive epistemic statements, is well known for the high computational complexity, as well as for certain logical irregularities, such as the logical omniscience problem. Other approaches, as in (Suchan et al., 2018), do model beliefs in formal languages, but adopt a domain-dependent modeling, making it difficult to prove generic properties, e.g., about nested beliefs, action ramifications etc.

Contribution and Impact

The aim of this study is of both theoretical and practical interest. Our main contribution is a formal and declarative implementation of a theory for reasoning about action, knowledge and time for dynamic domains, which does not rely in the possible-worlds semantics. We deliver an axiomatization that has a number of advantages, in comparison to existing frameworks. First, the theory is able to support epistemic reasoning about a multitude of commensense phenomena, such as direct and indirect effects of actions, default knowledge, inertia etc. Second, our implementation enables approximate epistemic reasoning, in order to tackle issues related to high computational complexity. Last, we develop a means to automatically transform non-epistemic domain axiomatizations into a formal encoding with well-defined properties that enables reasoning with belief, thus simplifying the task of the knowledge engineer when modeling the dynamics of causal domains.

We argue that such a system can impact various aspects of practical research in fields related to social robotics and computer vision, especially for interpreting scenes that involve human-machine interaction. Omitting the technical details, we discuss cases that signal how an agent with ToM can prove beneficial in a range of situations, from intuitive communication and advanced decision making to the analysis of human-object interaction videos.

Next, we introduce the main formalisms that form the basis for our framework (Section 2), and present our methodology and initial implementation results (Section 3). Section 4 showcases a number of areas, where such a framework can accomplish impact. The paper concludes in Section 5 with remarks on the directions of future research that lies ahead.

2 BACKGROUND

Our framework builds on and extends two formalisms, a discrete time non-epistemic dialect of the Event Calculus, capable of modeling a multitude of commonsense phenomena, and an epistemic extension of this dialect that does not rely on the possible worlds semantics.[1]

2.1 Non-epistemic Notions

Reasoning about actions, change and causality is an active field of research since the early days of AI. Among the various formalisms that have been proposed is the Event Calculus (EC) (Kowalski and Sergot, 1986; Miller and Shanahan, 2002), a well-established technique for reasoning about causal and narrative information in dynamic environments. It is a

[1]Epistemic logics represent knowledge, i.e., facts that are true, while doxastic logics are used for reasoning about potentially erroneous beliefs of agents. Although our main goal is to model an agent's belief state, we occasionally refer to knowledge for convenience, as commonly done in relevant literature too, but without necessarily being restricted to epistemic logics exclusively.

Table 1: Event Calculus Types of Formulae.

Domain Signature		
$\mathcal{F},\mathcal{E},\mathcal{T}$	Fluents, Events and Timepoints	E.g., $f, f_i, e, e_i, \mathbb{N}_0$
Axioms		
$\mathcal{DEC}$	Domain-independent Axioms	See (Mueller, 2015)
Σ	Positive Effect Axioms	$\bigwedge[(\neg)holdsAt(f_i,T)] \Rightarrow initiates(e,f,T)$
Σ	Negative Effect Axioms	$\bigwedge[(\neg)holdsAt(f_i,T)] \Rightarrow terminates(e,f,T)$
Δ_2	Trigger Axioms	$\bigwedge[(\neg)holdsAt(f_i,T)]\wedge$ $\bigwedge[(\neg)happens(e_j,T)] \Rightarrow happens(e,T)$
Γ	Initial State and Observations	$holdsAt(f,0), \neg holdsAt(f_1,1), ...$
Δ_1	Event Occurrences	$happens(e,0), happens(e_1,3), ...$

many-sorted first-order language for reasoning about action and change, which explicitly represents temporal knowledge. It also relies on a non-monotonic treatment of events.

Many EC dialects have been proposed over the years; for our purposes, we will use the non-epistemic discrete time Event Calculus dialect (DEC), axiomatized in (Mueller, 2015). Formally, DEC defines a sort $\mathcal{E}$ of *events* indicating changes in the environment, a sort $\mathcal{F}$ of *fluents* denoting time-varying properties and a sort $\mathcal{T}$ of *timepoints*, used to implement a linear time structure. The calculus applies the *principle of inertia* for fluents, in order to solve the frame problem, which captures the property that things tend to persist over time unless affected by some event. For instance, the fluent $faces(Agent, Orientation)$ indicates the point of view of an agent, while the event $turnsTowards(Agent, Orientation)$ denotes a change in orientation.[2]

A set of predicates express which fluents hold when ($holdsAt \subseteq \mathcal{F} \times \mathcal{T}$), which events happen ($happens \subseteq \mathcal{E} \times \mathcal{T}$), which their effects are ($initiates$, $terminates$, $releases \subseteq \mathcal{E} \times \mathcal{F} \times \mathcal{T}$) and whether a fluent is subject to the law of inertia or released from it ($releasedAt \subseteq \mathcal{F} \times \mathcal{T}$). For example, $initiates(e,f,T)$ means that if action e happens at som timepoint T it gives cause for fluent f to be true at timepoint $T+1$.

The commonsense notions of persistence and causality are captured in a set of *domain independent* axioms, referred to as $\mathcal{DEC}$, that define the influence of events on fluents and the enforcement of inertia for the *holdsAt* and *releasedAt* predicates. In brief, $\mathcal{DEC}$ states that a fluent that is not released from inertia has a particular truth value at a particular time if at the previous timepoint either it was given a cause to take that value or it already had that value.

In addition to domain independent axioms, a particular *domain axiomatization* describes the commonsense domain of interest (Σ and Δ_2 set of axioms), observations of world properties at various times (Γ axioms) and a narrative of known world events (Δ_1 axioms) (see Table 1). Action occurrences, as well as their effects may be context-dependent, i.e., they may depend on preconditions. For instance, the domain effect axiom

$holdsAt(faces(A,O),T) \Rightarrow$
$terminates(turnsTowards(A,O_{new}),faces(A,O),T)\wedge$
$initiates(turnsTowards(A,O_{new}),faces(A,O_{new}),T)$

implements the change in orientation of an agent, when the event *turnsTowards* occurs.

2.2 Epistemic Notions

To support reasoning about the mental state of agents, theories like $\mathcal{DEC}$ need to be extended with epistemic modalities (e.g., *knows*, *believes*), in order to represent the properties of both ontic and epistemic fluents and events. The epistemic extensions enable the reasoning agent to make inferences even in cases when the state of preconditions is unknown upon action occurrence. Lately, a number of epistemic EC dialects have been proposed, most of which rely on the possible-worlds semantics to assign meaning to the epistemic notions, e.g., (Ma et al., 2013; D'Asaro et al., 2020). This semantics provide intuitiveness and highly expressive models, but come at a cost: the computational complexity is exponential to the number of unknown parameters, while certain counter-intuitive assumptions, such as logical omniscience, need to be tolerated. Moreover, although in principle nested beliefs can be supported, most existing implementations of these formalisms are limited to 1^{st}-order epistemic statements.

[2]Variables start with a upper-case letter and are implicitly universally quantified, unless otherwise stated. Predicates and constants start with a lower-case letter.

Table 2: The ASP modules that constitute the epistemic EC reasoner.

	Non-epistemic	1^{st}-order ToM	2^{nd}-order ToM
Domain-independent Axioms	$\mathcal{DEC}$	Core $\mathcal{DECKT}$ Hidden Causal Dependencies	2^{nd}-order $\mathcal{DECKT}$
Domain-dependent Axioms	Domain Axiomatization	Meta-domain Axiomatization	
	Initial State Observations	Initial State Observations	Initial State Observations

The Discrete time Event Calculus Knowledge Theory ($\mathcal{DECKT}$) on the other hand, first proposed in (Patkos and Plexousakis, 2009), is an epistemic extension of $\mathcal{DEC}$ that adopts a deductive approach to modeling knowledge. Rather than producing knowledge by contrasting the truth value of fluents that belong to different possible worlds, $\mathcal{DECKT}$ defines a set of meta-axioms that, in brief, capture the following: i) when an action occurs, if all preconditions of an effect axiom triggered by this action are known, the effect will also become known, ii) if at least one precondition is known not to hold, no belief change regarding the effect will occur; iii) in all other cases, i.e., when at least one precondition is unknown, but none is known not to hold, then the state of the effect will become unknown too; at the same time, a causal dependency, called hidden causal dependency (HCD), will be created between the unknown precondition(s) and the effect. The idea behind HCDs is that if it turns out that the unknown preconditions did indeed hold, then so should the effect, given that no action affected these fluents in-between. $\mathcal{DECKT}$ also axiomatizes the conditions under which such causal dependencies are expanded or eliminated, considering the interplay of the effects of events as time progresses.

The theory is sound and complete with respect to possible-worlds theories under specific assumptions, e.g., deterministic domains. The explicit treatment of epistemic fluents as ordinary domain fluents introduces advantages, as we explain next. Yet, there are certain limitations, which we wish to overcome with our current work. First, to the best of our knowledge, the only implementation of $\mathcal{DECKT}$ to date is a rule-based system (see (Patkos et al., 2016)) with procedural, rather than declarative semantics; in this work, we deliver an encoding in the language of Answer Set Programming (ASP), based on formal, stable models semantics. Second, $\mathcal{DECKT}$ only models knowledge, without any support for nested knowledge statements; our implementation offers the ability to expand the formalism with nested statements. This encoding lays the ground for modeling also belief, rather than knowledge. Third, our implementation of $\mathcal{DECKT}$ helps perform approximate epistemic reasoning, a task that is not trivial for possible world-based implementations, offering sound but potentially incomplete inferences, to alleviate computational complexity issues. Last, as we show next, we also axiomatize epistemic events, such as *notices*, not supported by the original theory.

3 METHODOLOGY

3.1 The Cognitive Model

The constituent parts of our approximate epistemic EC reasoner are presented in Table 2. The logical program is broken down into modules (rulesets), each of which corresponds to a particular set of axioms with well-specified properties.[3] The encoding of all axiomatizations has been done in the Answer Set Programming (ASP) language (Gelfond and Lifschitz, 1988; Marek and Truszczynski, 1999). ASP is a declarative problem solving paradigm oriented towards complex combinatorial search problems. A domain is represented as a set of logical rules, whose models, called answer sets, correspond to solutions to a reasoning task. Sets of such rules, or answer set programs, come with an intuitive, well-defined semantics, having its roots in research in knowledge representation, in particular non-monotonic reasoning. Our system implements a translation of all the EC theories into ASP rules, which are then executed by the state-of-the-art Clingo ASP reasoner[4].

As shown in Table 2, there are three sets of modules, one for non-epistemic reasoning, one for 1^{st}-order epistemic inferencing and a third one for 2^{nd}-order, nested epistemic statements. Each set contains a domain-independent axiomatization, needed for implementing the appropriate commonsense behavior, regardless of the domain of interest. For the first set, this module is the encoding of the $\mathcal{DEC}$ set of axioms. The second part splits $\mathcal{DECKT}$ into the core $\mathcal{DECKT}$ set and the HCD axioms, whereas in the

[3] Code URL: https://socola.ics.forth.gr/tools/

[4] Clingo URL: https://potassco.org/

third part, the 2^{nd}-order $\mathcal{DECKT}$ module is an adaptation of the $\mathcal{DECKT}$ axioms appropriate for nested statements. For instance, the following two encodings specify how knowledge is generated when all preconditions of an effect axiom are known to the agent

```
initiates(notices(Observer, Event),
          believes(Observer, Effect), T) :-
  axiomEvent(ID, Event),
  happens(notices(Observer, Event), T),
  allPrecBelievedTrue(ID, Observer, T),
  axiomEffectPos(ID, Effect).

initiates(notices(Observer, Event),
          believesNot(Observer, Effect), T) :-
  axiomEvent(ID, Event),
  happens(notices(Observer, Event), T),
  allPrecBelievedTrue(ID, Observer, T),
  axiomEffectNeg(ID, Effect).
```

Informally, the rules state that when an observer notices the occurrence of an event that may cause a certain effect and she also believes that all preconditions for that effect hold, then she will also believe that the effect holds after the event, i.e., the observer will believe the effect to be true, for positive effect axioms (first rule) or she will believe the effect to be false, for negative effect axioms (second rule). A unique *ID* is assigned to each effect axiom, that is used for rules, such as the above, to generate domain-independent epistemic inferences (this also explains why $\mathcal{DECKT}$ is considered a meta-theory).

Similar rules specify how the mental state of agents should change when partial information about the preconditions is available. Note that these rules do not assume that the beliefs are correct; false initial beliefs or events not observed by the agents may lead to the generation of erroneous conclusions. The axiomatization only ensures sound belief inference given a specific state of mind.

As already mentioned, these rules are generic and apply to any effect axiom, regardless of the domain. The actual domain axiomatization, the part that defines the dynamics of a specific environment of interest inhabited by the agents and humans, is captured by a different module that encodes rules, such as:

```
terminates(turnsTowards(Agent, Dir),
           faces(Agent, DirInitial), T) :-
   holdsAt(faces(Agent, DirInitial), T),
   orientation(Dir),
   DirInitial != Dir,
   time(T).
```

In order for the epistemic parts to utilize such non-epistemic domain axiomatization, i.e., in order for $\mathcal{DECKT}$ to apply its meta-axiomatization approach, we developed a parser that automatically generates a set of rules for each domain axiom, which specify the constituent parts of this axiom. The parser assigns a unique identifier to each effect axiom and defines meta-predicates that capture which the preconditions are, which event triggers the axiom and which the effect is. Care needs to be taken during this decomposition process to correctly maintain the binding of variables between the different parts of the original axiom. This is one of the main contributions of this work, as it relieves the knowledge engineer from having to model complex epistemic rules. In practice, this means that non-epistemic EC theories can now be translated for epistemic reasoning, with no additional manual modeling effort. For the time being, our implementation only translates effect axioms, but we currently expand the types of axioms supported.

A final note about our methodology in building the epistemic reasoner concerns its modularity. Some of the modules are mandatory, in order for the inferences to be sound. Others though can be omitted, according to the type of reasoning one wishes to perform. For instance, $\mathcal{DEC}$ and core $\mathcal{DECKT}$ are sufficient for 1^{st}-order statement inference; the omission of HCD axioms, which are computationally intensive, do not affect soundness, but may lead to partial conclusions (fluents that could be inferred to be true or false will remain unknown). As a result, this modularity of the encoding helps support approximate reasoning. Note that such a flexibility is not easily accomplished with possible worlds-based theories, as it is not always straightforward how to decide which worlds to maintain and which to drop, in order to reduce complexity without losing soundness of inference.

3.2 Implementation

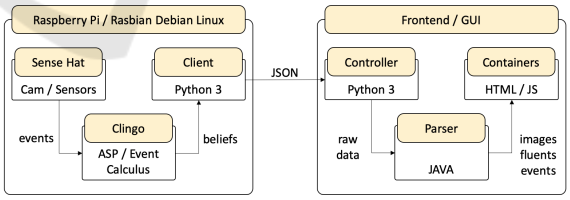

Figure 2: The system architecture.

To test our cognitive model, we are implementing a system that can be used as the basis for experimenting with diverse scenarios (Figure 2). The system comprises a Raspberry Pi computing environment (named Raspie from now on) that plays the role of a social robot operating in the environment. We used a Raspberry Pi 4 Model B 8GB, equipped with various sensors, such as camera, gyroscope and acceleromenter. We also installed the Clingo 5.5 ASP reasoner onboard, so that all epistemic inferencing needed to support ToM behavior is executed at run-time locally.

Figure 3: The fronend displays different world views: the actual world state, the humans beliefs, Raspie's beliefs and Raspie's beliefs about what the human believes.

In addition to Raspie, we assume that a human user is positioned behind the desk. Any event, such as change in the location of Raspie or the human user, will trigger the reasoner, which will generate new beliefs about where each entity is, what can be observed by each entity, which objects are known to each entity to be on the table, which their spatial relations are, etc.

The new belief states are then sent to the frontend, which groups beliefs of the same type together and displays them in dedicated panels (Figure 3). Apart from Raspie's beliefs about the environment (1^{st}-order belief statements) and about the human user (2^{nd}-order belief statements), the frontend also displays the actual world state and the human's beliefs, based on separate axiomatizations provided from a different channel. These latter world views are not directly accessible to Raspie, but help us better understand the epistemic inferences, when sense or communication actions take place.

4 DISCUSSION

In this section, we briefly discuss different scenarios that highlight both the expressive power and the impact that such a ToM-enabled robot can have in supporting complex, real-world situations. For the purpose of the current position paper, we omit most of the technical details. The goal is to showcase situations that cannot easily be implemented without a rich cognitive model or cases where ToM can provide important leverage to intelligent systems. While most of the modeling requirements described next are already known to the research community working on classical AI, the fact that the proposed framework comes with a unified solution to these phenomena, while taking into consideration how to reduce the computational complexity, is, to our opinion, a step forward.

False Beliefs: Variations of the classic "Sally and Anne test" are often being used to model the state of mind of an observer, when modeling facets of social cognition. The office desk example can offer an adaptation of such a setting: imagine that the human believes that, from her point of view, a pen is located behind the screen:

$$holdsAt(believes(human, loc(human, behindOf(pen, laptop))), 0)$$

Raspie, on the other hand, from its current position, has no knowledge about objects located there:

$$\neg holdsAt(believes(raspie, loc(raspie, leftOf(Object, laptop))), 0)$$
$$\neg holdsAt(believesNot(raspie, loc(raspie, leftOf(Object, laptop))), 0)$$

Yet, it also believes that the human does not believe there is a pen behind the screen (2^{nd}-order statement)

$$holdsAt(believes(raspie, believesNot(human, loc(human, behindOf(Object, laptop)))), 0)$$

Such a representation can capture the subjectivity of each entity, as well as the ability of agents to engage in perspective-taking, ascribing a mental state to another that they themselves believe to be false.

The fact that the actual state of the world may be such that no pen is placed there makes things even more interesting: considering the above belief states, as well as the position of all observing entities around the table, one can see that the result of a sense action may significantly differ from the result of a communication action. In general, the proper handling of ontic actions, such as move, pick up, grab, etc., and epistemic actions that only change one's perspective about the state of the world, e.g., sense, announce, ask, distract, constitute essential ingredients for any cognitive entity operating in causal domains.

Intuitive Communication and Explainability: When two humans engage in a dialogue, a lot of information is left out, because it is considered too obvious to be shared (in rhetorical syllogism, such statements are called enthymemes). This is a cognitive ability that is particularly difficult for an intelligent agent to master, as it requires both a wealth of background knowledge to be held and a good understanding of what can be considered common knowledge between the discussing parties. For social robots, deciding *when* to ask the human user for information or to provide guidance, as well as *how* to express an utterance, can make the difference between providing assistance or becoming an obstruction. A rich cognitive model, enhanced with ToM capabilities, can drive the agent to only place questions if it believes that the human may know the answer, based on her current or past activity. It can also help the agent become more elaborate ("You can use the blue pen behind the carton box on your right") or abstract ("You can use the blue pen"), based on the level of common information the two entities it is believed they share.

More importantly, the ability of AI agents to explain their actions and decision making processes is becoming more urgent lately. The transparency and provability of formal methods and the scrutiny of beliefs grounded not only on the perspective of the different observers, but also on the type of beliefs, as discussed next, can significantly impact the trustworthiness of a system interacting with non-expert users.

Revision based on Types of Beliefs: The example so far has revealed three types of belief: beliefs coming from observation (sense actions), beliefs communicated by other entities (announce actions) and beliefs inferred, based on logical inference. Additionally, theories, such as the EC, allow for defaults to be modelled, e.g., agents may typically believe that pencils can be found in a pencil box, if one is located on the desk. Defaults constitute big part of human intuition and reflect the experiences and background knowledge of humans when they operate in familiar to them environments. Apparently, an observation may invalidate such beliefs. The point is that, in certain cases, some types of knowledge or beliefs can be considered more trusted than others. This is proven helpful when the agent's beliefs contradict each other; although statistical methods try to find quantitative measures, in order to assign confidence values from contradicting sources of information, a qualitative approach that takes into consideration the type of knowledge manipulated can lead to more intuitive and efficient revision schemes. For instance, preference-based models are often used in relevant literature, and have recently been applied to action formalisms, such as the EC (Tsampanaki et al., 2021).

Action Prediction: Inferences such as the ones discusses so far constitute the first step towards accomplishing complex reasoning tasks. By relying on a rich cognitive model of human beliefs, along with past interactions with objects in a given domain, an intelligent system can go one step further and try to anticipate human needs and intentions, predict future actions and, in general, provide timely assistance, rather than just respond to commands.

Consider the following statement: "Typically, a human will a) look for an object she needs, based on her currently committed intentions, b) reach for the object that is closer to her/easier to reach, and c) choose the object that is working properly (not broken)/is clean/is fresh etc.". Template statements such as this are both generic enough to capture typical user behavior and can easily be adapted to particular domain-specific requirements (part (c) of the statement). Endowing social agents with generic human behavior prescriptions can help in interpreting scenes, predicting the human's next actions, and ultimately identifying opportunities for offering assistance ("There is a pencil behind the screen, in case you haven't noticed it") or for informing the user about false beliefs ("While your attention was on your mobile, the cat run away with the laptop mouse").

5 CONCLUSIONS

In this paper, we discussed work in progress towards developing a formal framework for intelligent agents capable of exhibiting ToM. We argued about the importance of such cognitive skills for autonomous entities operating close to the human and we further provided initial implementation directions that build on existing research in epistemic action languages.

This initial work lays the ground for both theoretical and practical advancement. For start, given that we introduced new features to $\mathcal{DECKT}$(new

epistemic actions, nested epistemic statements etc.), we also need to update the formal proofs regarding the equivalence with possible worlds-based theories. We also identified numerous ways of extending the expressive power of $\mathcal{DECKT}$, to account for more complex cases, such as revision of beliefs (recall that $\mathcal{DECKT}$ only supports knowledge, i.e., in the presence of contradicting statements, the theory collapses), potential action occurrences, beliefs of diverse types, among others.

From the practical standpoint, our main goal is to evaluate how ToM can improve typical prediction tasks that are of interest in the field of Computer Vision. Already recent studies, as by (Ji et al., 2021), try to take advantage of past human-object interactions, including where the user looked at, in order to predict future actions in videos. Datasets, such as Action Genome, that provide annotations about attentional relationships (whether a person is looking at something), in addition to spatial and contact relationships, can help build cognitive models about the mental state of users. In addition to such experiments, we also plan to evaluate the proposed formalism in terms of scalability and to further explore efficient means of implementing HCDs, the main component that introduces exponential complexity to the epistemic reasoner.

ACKNOWLEDGEMENTS

This project has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement No 188.

REFERENCES

Apperly, I. A. (2012). What is "theory of mind"? concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, 65(5):825–839.

Chen, B., Vondrick, C., and Lipson, H. (2021). Visual behavior modelling for robotic theory of mind. *Scientific Reports*, 11(1):424.

D'Asaro, F. A., Bikakis, A., Dickens, L., and Miller, R. (2020). Probabilistic reasoning about epistemic action narratives. *Artificial Intelligence*, 287:103352.

Davis, E. and Morgenstern, L. (2005). A First-order Theory of Communication and Multi-agent Plans. *Journal of Logic and Computation*, 15(5):701–749.

Ditmarsch, H. v., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*. Springer Publishing Company, Incorporated, 1st edition.

Gelfond, M. and Lifschitz, V. (1988). The stable model semantics for logic programming. In *Proc. 5th International Joint Conference and Symposium on Logic Programming, IJCSLP 1988*, pages 1070–1080.

Ji, J., Desai, R., and Niebles, J. C. (2021). Detecting human-object relationships in videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8106–8116.

Kowalski, R. and Sergot, M. (1986). A logic-based calculus of events. newgeneration computing 4.

Liu, Y. and Levesque, H. (2005). Tractable reasoning with incomplete first-order knowledge in dynamic systems with context-dependent actions. In *IJCAI-05*, pages 522–527.

Ma, J., Miller, R., Morgenstern, L., and Patkos, T. (2013). An epistemic event calculus for asp-based reasoning about knowledge of the past, present and future. In *LPAR 2013, 19th International Conference on Logic for Programming*, volume 26, pages 75–87.

Marcus, G. and Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon Books, USA.

Marek, V. W. and Truszczynski, M. (1999). Stable models and an alternative logic programming paradigm. In *The Logic Programming Paradigm: A 25-Year Perspective*, pages 375–398. Springer Berlin Heidelberg.

Miller, R. and Shanahan, M. (2002). Some alternative formulations of the event calculus. In *Computational logic: logic programming and beyond*, pages 452–490. Springer.

Mueller, E. (2015). *Commonsense Reasoning: An Event Calculus Based Approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2^{nd} edition.

Patkos, T. and Plexousakis, D. (2009). Reasoning with knowledge, action and time in dynamic and uncertain domains. In *IJCAI-09*.

Patkos, T., Plexousakis, D., Chibani, A., and Amirat, Y. (2016). An event calculus production rule system for reasoning in dynamic and uncertain domains. *Theory Pract. Log. Program.*, 16(3):325–352.

Scherl, R. (2003). Reasoning about the interaction of knowlege, time and concurrent actions in the situation calculus. In *IJCAI-03*, pages 1091–1096.

Schwering, C., Lakemeyer, G., and Pagnucco, M. (2015). Belief revision and progression of knowledge bases in the epistemic situation calculus. In *IJCAI-15*.

Shapiro, S., Pagnucco, M., Lespérance, Y., and Levesque, H. (2011). Iterated belief change in the situation calculus. *Artificial Intelligence*, 175(1):165–192.

Suchan, J., Bhatt, M., Wałega, P., and Schultz, C. (2018). Visual explanation by high-level abduction: On answer-set programming driven reasoning about moving objects. In *AAAI Conference on Artificial Intelligence*, pages 1965–1972.

Tsampanaki, N., Patkos, T., Flouris, G., and Plexousakis, D. (2021). Revising event calculus theories to recover from unexpected observations. *Annals of Mathematics and Artificial Intelligence*, 89(1-2):209–236.