Tentacular Artificial Intelligence, and the Architecture Thereof, Introduced

Selmer Bringsjord[1], **Naveen Sundar G**[1], **Atriya Sen**[1], **Matthew Peveler**[1], **Biplav Srivastava**[2] **Kartik Talamadupula**[2]
[1] Rensselaer Polytechnic Institute (RPI); RAIR Lab
[2] IBM Research
Selmer.Bringsjord@gmail.com, naveensundarg@gmail.com, atriya@atriyasen.com,
matt.peveler@gmail.com, biplavs@us.ibm.com, krtalamad@us.ibm.com

Abstract

We briefly introduce herein a new form of distributed, multi-agent artificial intelligence, which we refer to as "tentacular." Tentacular AI is distinguished by six attributes, which among other things entail a capacity for reasoning and planning based in highly expressive calculi (logics), and which enlists subsidiary agents across distances circumscribed only by the reach of one or more given networks.

1 Introduction

We briefly introduce herein a new form of distributed, multi-agent artificial intelligence. An AI artifact s is currently understood as an agent with a predetermined set of goals, a set of fixed inputs and outputs, and obligations and permissions. The agent does not have any leeway in accomplishing its goals or adhering to its obligations, prohibitions, or other legal/ethical principles that bind it. Do we need agents that go beyond these limitations? A humble example follows: During your daily commute to work, an agent a_c in your car observes that there is more traffic than usual headed toward the local store. It then consults a weather service and finds that a major storm is headed toward your town. a_c conveys this information to a_h, an agent in your home. a_h then communicates with an agent a_p on your phone and finds out that you do not know about the storm coming your way, as you have not made any preparations for it; and as further evidence of your ignorance, you have not read any notifications about the storm. a_h then infers from your calendar that you may not have enough time to get supplies after you read your notifications later in the day. a_h commands a_c to recommend to you a list of supplies to shop for on your way home, including at least n items in certain categories (e.g. 3 gallons of bottle water).

AI of today, as defined by any orthodox, comprehensive overview of it (e.g. [Russell and Norvig, 2009]), consists in the design, creation, implementation, and analysis of **artificial agents**.[1] Each such agent a takes in information about its particular environment E (i.e. takes in **percepts** of E), engages in some computation, and then, on the strength of that computation, performs an action/actions in that environment. (Of course, for an agent that persists, this cycle iterates through time.) On this definition, a computer program that implements, say, the factorial function $n!$ qualifies as an artificial agent (let's dub it 'a_{FAC}'), one operating in the environment $\mathbf{E}_{\mathcal{N}}$ of basic arithmetic; and the human who has conceived and written this program has built an artificial agent. While plenty of the artificial agents touted today are rather more impressive than a_{FAC}, our aim is to bring to the world, within a decade, a revolutionary kind of AI that yields artificial agents with a radically higher level of intelligence (including intelligence high enough to qualify the agents as *cognitively conscious*) and power. This envisioned AI we call **Tentacular AI**, or just 'TAI' for short (rhymes with 'pie'). Before presenting architectural-level information about TAI, we give an example that's a bit more robust than our first-paragraph one.

Let's suppose that an AI agent a_{HOME} overseeing a home is charged with the single, unassuming task of moving a cup on the home's kitchen table onto a saucer that is also on that table. How shall the agent make this goal happen? If the AI can delegate to a robot in the house capable of manipulating standard tabletop objects in a narrow tabletop environment $\mathbf{E}_{\text{TABLE}}$, and that robot is at the table or can get there in a reasonable amount of time, then of course a_{HOME} can direct the robot to pick up the cup and put it on the saucer. This is nothing to write home about, since AI of today has given us agent-robot combos that, in labs (our own, e.g.) and soon enough in homes across the technologized world, can do this kind of thing routinely and reliably. In fact, this kind of capability to find plans and move tabletop objects around in order to obtain goals in tabletop environments[2] has been a solved problem from the research point of view for decades [Genesereth and Nilsson, 1987], e.g. Not only that, but there are longstanding theorems telling us that the intrinsic difficulty of finding plans to move various standard tabletop objects in arbitrary starting configurations in tabletop environments is algorithmically solvable and generally tractable.[3]

[1] This is the exact phrase used by Russell and Norvig [2009]. Other comprehensive overviews match the Russell-Norvig orientation; e.g. [Luger, 2008].

[2] Such environments are variants of those traditionally termed 'blocks-worlds.'

[3] E.g., see [Gupta, 1992].

However, AI of today is, if you will, living a bit of a lie. Why? Because in real life, the agent a_{HOME} would *not* be operating in only the tabletop environment $\mathbf{E}_{\text{TABLE}}$; rather the idea is that this agent should be able to understand and manage the overall environment $\mathbf{E}_{\text{HOME}}$ of the home, which surely comprises much more than the stuff standardly on one kitchen table! Homes can have parents, kids, dogs, visitors, . . . *ad infinitum*.

For example, suppose that a_{HOME} finds that the tabletop robot is broken, having been mangled by the home's frisky beagle. *Then* how does a_{HOME} solve the problem of getting the saucer moved? Artificial agents of today capable of the kind of planning that worked before this complication are now hamstrung. But not so a TAI agent. One reason is that TAI agents are capable of human-level communication. In certain circumstances within a_{HOME} the most efficient way for the agent a_{HOME} to accomplish the task may be to simply say politely via I/IoT (Internet or Interent of Things) through a speaker or a smartphone or a pair of smart glasses to a human in the home (of whose mind the TAI agent has a model) sitting at the table in question: "Would you be so kind as to place that cup on top of the saucer?" Of course, a_{HOME} may not be so fortunate as to have the services of a human available: maybe no human is at home, yet the task must be completed. In this case, a TAI agent can still get things done, in creative fashion. E.g., suppose that in the home a family member received beforehand a small blimp that can fly around inside the home and pick things up.[4] The TAI agent might then activate and use this blimp through I/IoT to put the cup atop the saucer. But what, more precisely, is a TAI agent? We say that a TAI agent must be:

$\mathbf{D_1}$ *Capable of problem-solving*. Whereas, as we've noted, standard AI counts simple mappings from percepts to actions as *bona fide* AI, TAI agents must be capable of problem-solving. This may seem like an insignificant first attribute of TAI, but a consequence that stems from this attribute should be noted: Since problem-solving entails capability across the main sub-divisions of AI, TAI agents have multi-faceted power. Problem-solving requires capability in these sub-areas of AI: planning, reasoning, learning, communicating, creativity (at least relatively simple forms thereof), and — for making physical changes in physical environments — cognitive robotics.[5] Hence, all TAI agents can plan, reason, learn, communicate; and they are creative and capable of carrying out physical actions.

$\mathbf{D_2}$ *Capable of solving at least important instances of problems that are at and/or **above** Turing-unsolvable problems*. AI of today, when capable of solving problems, invariably achieves this success on problems that are merely algorithmically solvable and tractable (e.g., checkers, chess, Go).

$\mathbf{D_3}$ *Able to supply justification, explanation, and certification of supplied solutions, how they were arrived at, and that these solutions are safe/ethical*. We thus say that the problem-solving of a TAI agent is **rationalist**. This label reflects the requirement that any proposed solution to the problem discovered by a TAI agent must be accompanied by a justification that defends and explains that the proposed solution *is* a solution, and, when appropriate, also that the solution (and indeed perhaps the process used to obtain the solution) has certain desirable properties. Minimally, the justification must include an argument or proof for the relevant conclusions. In addition, the justification must be verified, formally; we thus say that **certification** is provided by a TAI agent.

$\mathbf{D_4}$ *Capable of "theory-of-mind" level reasoning, planning, and communication*. Discussion of this attribute is omitted to save space; see e.g. [Arkoudas and Bringsjord, 2009] for our lab's first foray into automated reasoning at this level. (The truth is, it's more accurate to say the fourth requirement is that a TAI agent must have *cognitive consciousness*, as this phenomenon is explained and axiomatized in [Bringsjord *et al.*, 2018].)

$\mathbf{D_5}$ *Capable of creativity, minimally to the level of so-called **m-creativity***. Creativity in artificial agents, and the engineering thereof, has been discussed in a number of places by Bringsjord [Bringsjord and Ferrucci, 2000]e.g., but recently Bringsjord and Sen [2016] have called for a form of creativity in artificial agents using I/IoT.

$\mathbf{D_6}$ *Has "tentacular" power wielded throughout I/IoT, Edge Computing, and cyberspace*. This is the most important attribute possessed by TAI agents, and is reflected in the 'T' in 'TAI.' To say that such agents have tentacular problem-solving power is to say that they can perceive and act through the I/IoT (or equivalent networks) and cyberspace, across the globe. TAI agents thus operate in a planet-sized, heterogeneous environment that spans the narrower, fixed environments used to define conventional, present-day AI, such as is found in [Russell and Norvig, 2009].

2 Related Work

Given the limited scope of the present paper, we only make some brief comments about related work, which can be partitioned for convenience into that which is can be plausibly regarded as on the road toward the level of expressivity and associated automated reasoning that TAI requires, and prior work that provides a stark and illuminating contrast with TAI.

First, as to work we see as reaching toward TAI, we note that recently Miller *et al.* [2018] present a planning framework that they call *social planning*, in which the agent under consideration can plan and act in a manner that takes account of the beliefs of other agents. The goal for an agent in social planning can either be a particular state of the external world, or a set of beliefs of other agents (or a mix of both). The system is built upon a simplified version of a propositional modal logic (unlike our system, presented below, which is more expressive and can accommodate more complex goals, e.g. goals over unbounded domains or goals that involve numerical quantification; such statements require going beyond propositional modal logic). In addition, certainly the NARS system from Wang [2006] has elements that one can rationally view as congenial to TAI. For instance, NARS is multi-layered and reasoning-centric. On the other hand, the 'N' in 'NARS' is for 'Non-axiomatic,' and TAI, and indeed the entire approach to logicist AI pursued by at least Bringsjord and Govindarajulu, seeks whenever possible to leverage automated reasoning over powerful axiom systems, such as Peano

[4] Such a blimp is a simple adaptation of what is readily available as a relatively inexpensive toy.

[5] Cognitive robotics is defined in [Levesque and Lakemeyer, 2007] as a type of robotics in which all substantive actions performed by the robots are a function of the cognitive states (e.g. beliefs & intentions) of these robots.

Arithmetic.[6] In addition, TAI is deeply and irreducibly intensional, while NARS appears to be purely extensional. Clever management of computational resources in TAI is clearly going to be key, and we see the work of Thorisson and colleagues (e.g. [Helgason *et al.*, 2012]) to be quite relevant to TAI and the challenges the implementation of it will encounter. For a final example of work that is generally aligned with TAI, we bring to the reader's attention a recent comprehensive treatment of proof-based work in computer science: [Arkoudas and Musser, 2017]. As TAI is steadfastly proof-based AI, this tome provides very nice coverage of the kind of work required to implement TAI.

Secondly, for illuminating contrast, we note first that some have considered the concept of corporate intelligence composed of multiple agents, including machines, where inspiration comes from biology. A case in point is the fascinating modeling in [Seidita *et al.*, 2016].[7] In our case, TAI is a thoroughly formal conception independent of terrestrial biology, one that is intended to include types of agents of greater intelligence than those currently on Earth. Another illuminating contrast comes via considering established languages for planning that are purely extensional in nature (e.g. PDDL, which in its early form is given in [Mcdermott *et al.*, 1998]), as therefore quite different than planning of the type that is required for TAI, which must be intensional in character, and is (since cognitive calculi are intensional computational logics). MA-PDDL is an extension of PDDL for handling domains with multiple agents with varying actions and goals [Kovacs, 2012], and as such would seem to be relevant to TAI. But unlike social planning discussed above, MA-PDDL does not aim to change beliefs (nor for that matter other epistemic attitudes) of other agents. While MA-PDDL could be used to do so, representing beliefs and other cognitive states in PDDL's extensional language can lead to undesirable consequences, as demonstrated in [Bringsjord and Govindarajulu, 2012]. Extensions of the original PDDL (PDDL1), for example PDDL3 [Gerevini and Long, 2004], are still extensional in nature.

This concludes the related-work section. Note that below we describe and define TAI from the point of view of AI planning.

3 Quick Overview

We give a quick and informal overview of TAI. We have a set of agents $a_1, \ldots, a_n$. Each agent has an associated (implicit or explicit) contract that it should adhere to. Consider one particular agent τ. During the course of this agent's lifetime, the agent comes up with goals to achieve so that its contract is not violated. Some of these goals might require an agent to exercise some or all of the six attributes $\mathbf{D_1} - \mathbf{D_6}$. We formalize this using planning as shown in Figure 1. As shown in the figure, if some goal is not achievable on its own, τ can seek to recruit other agents by leveraging their resources, beliefs, obligations etc.

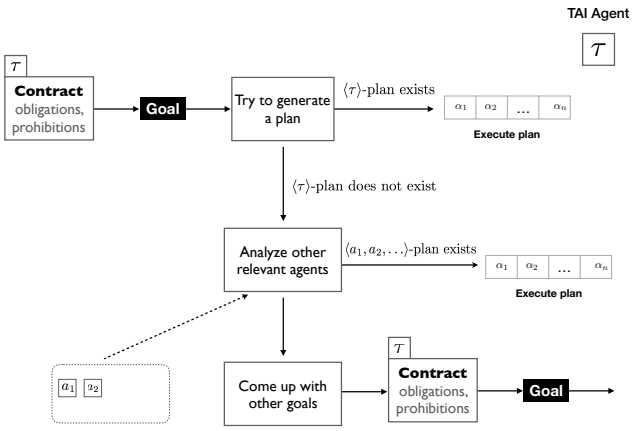

Figure 1: TAI Informal Overview: *We have an architecture for how a TAI agent τ might operate. τ continuously comes up with goals based on its contract. If a goal is not achievable using τ's own resources, τ has to employ other agents in achieving this goal. To successfully do so τ would need to have one or more of $\mathbf{D_1} - \mathbf{D_6}$ attributes.*

4 The Formal System

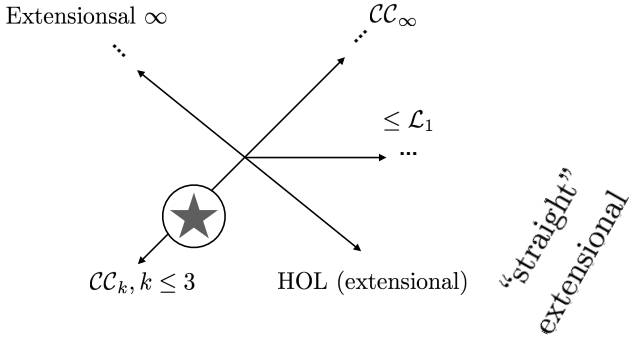

Figure 2: Space of Logical Calculi. *There are five dimensions that cover the entire, vast space of logical calculi. The due West dimension holds those calculi powering the Semantic Web (which are generally short of first-order logic = $\mathcal{L}_1$), and include so-called **description logics**. Both NW and NE include logical systems with wffs that are allowed to be infinitely long, and are needless to say hard to compute with and over. SE is higher-order logic, which has a robust automated theorem-proving community gathered around it. It's the SW direction that holds the cognitive calculi described in the present paper, and associated with TAI; and the star refers to those specific cognitive calculi called out in these pages by us.*

[6] The layering of TAI is in fact anticipated by the increasingly powerful axiom-centric cognition described in [Bringsjord, 2015], which takes Peano Arithmetic as central.

[7] Though out of reach for now, given that our chief objective is but an informative introduction to TAI, the relationship between our conception of cognitive consciousness, which is central to TAI agents (Attribute #4 above), and consciousness as conceived by Chella, is a fertile topic for future investigation. A multi-faceted discussion of artificial consciousness is by the way to be had in [Chella and Manzotti, 2007]. For a first-draft axiomatization of the brand of consciousness central to TAI agents, see [Bringsjord *et al.*, 2018].

To make the above notions more concrete, we use a version of a computational logic. The logic we use is **deontic cognitive event calculus** ($\mathcal{DCEC}$). This calculus is a first-order modal logic. Figure 2 shows the region where $\mathcal{DCEC}$ is located in the overall space of logical calculi. $\mathcal{DCEC}$ belongs to the **cognitive calculi** family of logical calculi (denoted by a star in Figure 2 and expanded in Figure 3). $\mathcal{DCEC}$ has a well-defined syntax and inference system; see Appendix A of [Govindarajulu and Bringsjord, 2017a] for a full description. The inference system is based on natural deduction [Gentzen, 1935], and includes all the introduction and elimination rules for first-order logic, as well as inference schemata for the modal operators and related structures

This system has been used previously in [Govindarajulu and Bringsjord, 2017a; Govindarajulu *et al.*, 2017] to automate versions of the doctrine of double effect $\mathcal{DDE}$, an ethical principle with deontological and consequentialist components. While describing the calculus is beyond the scope of this paper, we give a quick overview of the system below. Dialects of $\mathcal{DCEC}$ have also been used to formalize and automate highly intensional (i.e. cognitive) reasoning processes, such as the false-belief task [Arkoudas and Bringsjord, 2008] and *akrasia* (succumbing to temptation to violate moral principles) [Bringsjord *et al.*, 2014]. Arkoudas and Bringsjord [2008] introduced the general family of **cognitive event calculi** to which $\mathcal{DCEC}$ belongs, by way of their formalization of the false-belief task. More precisely, $\mathcal{DCEC}$ is a sorted (i.e. typed) quantified modal logic (also known as sorted first-order modal logic) that includes the event calculus, a first-order calculus used for commonsense reasoning.

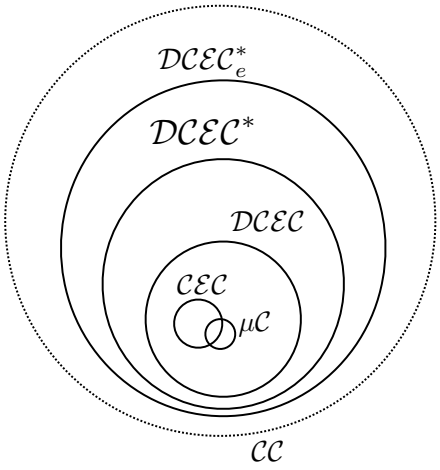

Figure 3: Cognitive Calculi. *The **cognitive calculi** family is composed of a number of related calculi. Arkoudas and Bringsjord introduced the first member in this family, $\mathcal{CEC}$, to model the false-belief task. The smallest member in this family, $\mu\mathcal{C}$, has been used to model uncertainty in quantified beliefs [Govindarajulu and Bringsjord, 2017b]. $\mathcal{DCEC}$ and variants have been used in the modelling of ethical principles and theories and their implementations.*

4.1 Syntax

As mentioned above, $\mathcal{DCEC}$ is a sorted calculus. A sorted system can be regarded as analogous to a typed single-inheritance programming language. We show below some of the important sorts used in $\mathcal{DCEC}$.

Sort	Description
Agent	Human and non-human actors.
Time	The Time type stands for time in the domain. E.g. simple, such as t_i, or complex, such as $birthday(son(jack))$.
Event	Used for events in the domain.
ActionType	Action types are abstract actions. They are instantiated at particular times by actors. Example: eating.
Action	A subtype of Event for events that occur as actions by agents.
Fluent	Used for representing states of the world in the event calculus.

The syntax has two components: a first-order core and a modal system that builds upon this first-order core. The figures below show the syntax and inference schemata of $\mathcal{DCEC}$. The first-order core of $\mathcal{DCEC}$ is the *event calculus* [Mueller, 2006]. Commonly used function and relation symbols of the event calculus are included. Fluents, event and times are the three major sorts of the event calculus. Fluents represent states of the world as first-order terms. Events are things that happen in the world at specific instants of time. Actions are events that are carried out by an agent. For any action type α and agent a, the event corresponding to a carrying out α is given by $action(a, \alpha)$. For instance if α is *"running"* and a is *"Jack"* , $action(a, \alpha)$ denotes *"Jack is running"*. Other calculi (e.g. the *situation calculus*) for modeling commonsense and physical reasoning can be easily switched out in-place of the event calculus.

Syntax

$$S ::= \text{Agent} \mid \text{ActionType} \mid \text{Action} \sqsubseteq \text{Event} \mid \text{Moment} \mid \text{Fluent}$$

$$f ::= \begin{cases} action : \text{Agent} \times \text{ActionType} \rightarrow \text{Action} \\ initially : \text{Fluent} \rightarrow \text{Formula} \\ holds : \text{Fluent} \times \text{Moment} \rightarrow \text{Formula} \\ happens : \text{Event} \times \text{Moment} \rightarrow \text{Formula} \\ clipped : \text{Moment} \times \text{Fluent} \times \text{Moment} \rightarrow \text{Formula} \\ initiates : \text{Event} \times \text{Fluent} \times \text{Moment} \rightarrow \text{Formula} \\ terminates : \text{Event} \times \text{Fluent} \times \text{Moment} \rightarrow \text{Formula} \\ prior : \text{Moment} \times \text{Moment} \rightarrow \text{Formula} \end{cases}$$

$$t ::= x : S \mid c : S \mid f(t_1, \ldots, t_n)$$

$$\phi ::= \begin{cases} q : \text{Formula} \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \forall x : \phi(x) \mid \\ \mathbf{P}(a, t, \phi) \mid \mathbf{K}(a, t, \phi) \mid \\ \mathbf{C}(t, \phi) \mid \mathbf{S}(a, b, t, \phi) \mid \mathbf{S}(a, t, \phi) \mid \mathbf{B}(a, t, \phi) \\ \mathbf{D}(a, t, \phi) \mid \mathbf{I}(a, t, \phi) \\ \mathbf{O}(a, t, \phi, (\neg) happens(action(a^*, \alpha), t')) \end{cases}$$

The modal operators present in the calculus include the standard operators for knowledge $\mathbf{K}$, belief $\mathbf{B}$, desire $\mathbf{D}$, intention $\mathbf{I}$, etc. The general format of an intensional operator is $\mathbf{K}(a, t, \phi)$, which says that agent a knows at time t the proposition ϕ. Here ϕ can in turn be any arbitrary formula. Also, note the following modal operators: $\mathbf{P}$ for perceiving a state,

C for common knowledge, **S** for agent-to-agent communication and public announcements, **B** for belief, **D** for desire, **I** for intention, and finally and crucially, a dyadic deontic operator **O** that states when an action is obligatory or forbidden for agents. It should be noted that $\mathcal{DCEC}$ is one specimen in a *family* of extensible cognitive calculi.

The calculus also includes a dyadic (arity = 2) deontic operator **O**. It is well known that the unary ought in standard deontic logic leads to contradictions. Our dyadic version of the operator blocks the standard list of such contradictions, and beyond.[8]

Declarative communication of ϕ between a and b at time t is represented using the $\mathbf{S}(a, b, t, \phi)$.

4.2 Inference Schemata

The figure below shows a fragment of the inference schemata for $\mathcal{DCEC}$. First-order natural deduction introduction and elimination rules are not shown. Inference schemata $I_{\mathbf{K}}$ and $I_{\mathbf{B}}$ let us model idealized systems that have their knowledge and beliefs closed under the $\mathcal{DCEC}$ proof theory. While humans are not deductively closed, these two rules lets us model more closely how more deliberate agents such as organizations, nations and more strategic actors reason. (Some dialects of cognitive calculi restrict the number of iterations on intensional operators.) I_{13} ties intentions directly to perceptions (This model does not take into account agents that could fail to carry out their intentions). I_{14} dictates how obligations get translated into known intentions.

Inference Schemata (Fragment)

$$\frac{\mathbf{K}(a, t_1, \Gamma),\ \ \Gamma \vdash \phi,\ \ t_1 \leq t_2}{\mathbf{K}(a, t_2, \phi)}\ [I_{\mathbf{K}}]$$

$$\frac{\mathbf{B}(a, t_1, \Gamma),\ \ \Gamma \vdash \phi,\ \ t_1 \leq t_2}{\mathbf{B}(a, t_2, \phi)}\ [I_{\mathbf{B}}]$$

$$\frac{\mathbf{K}(a, t, \phi)}{\phi}\ [I_4] \qquad \frac{t < t',\ \ \mathbf{I}(a, t, \psi)}{\mathbf{P}(a, t', \psi)}\ [I_{13}]$$

$$\frac{\mathbf{B}(a, t, \phi)\ \ \mathbf{B}(a, t, \mathbf{O}(a, t, \phi, \chi))\ \ \mathbf{O}(a, t, \phi, \chi)}{\mathbf{K}(a, t, \mathbf{I}(a, t, \chi))}\ [I_{14}]$$

4.3 Semantics

The semantics for the first-order fragment is the standard first-order semantics. The truth-functional connectives $\wedge, \vee, \rightarrow, \neg$ and quantifiers $\forall, \exists$ for pure first-order formulae all have the standard first-order semantics. The semantics of the modal operators differs from what is available in the so-called Belief-Desire-Intention (BDI) logics [Rao and Georgeff, 1991] in many important ways. For example, $\mathcal{DCEC}$ explicitly rejects possible-worlds semantics and model-based reasoning, instead opting for a *proof-theoretic* semantics and the associated type of reasoning commonly referred to as *natural deduction* [Gentzen, 1935; Francez and Dyckhoff, 2010]. Briefly, in this approach, meanings of modal operators are defined via arbitrary computations over proofs.

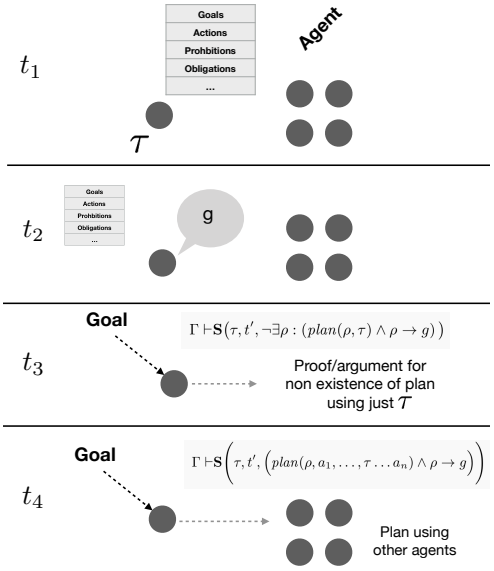

Figure 4: TAI Working Through Time. *A TAI agent initially considers a goal and then has to produce a proof for the non-existence of a non-tentacular plan that uses only this agent. Then τ recruits a set of other relevant agents to help with its goal.*

5 Defining TAI

We denote the state-of-affairs at any time t by a set of formulae $\Gamma(t)$. This set of formulae will also contain any obligations and prohibitions on different agents. For each agent a_i at time t, there is a contract $\mathbf{c}(a_i, t) \subseteq \Gamma(t)$ that describes a_i's obligations, prohibitions etc. a at any time t then comes up with a goal g so that its contract is satisfied.[9] The agent believes that if g does not hold then its contract at some future $t + \delta$ will be violated:

$$\mathbf{B}\left(a, t, \neg g \rightarrow \neg \bigwedge \mathbf{c}(a, t + \delta)\right)$$

Then the agent tries to come up with a plan involving a sequence of actions to satisfy the goal.

We make these notions more precise. An agent a has a set of actions that it *can* perform at different time points. For instance, a vacuuming agent can have movement along a plane as its possible actions while an agent on a phone can have displaying a notification as an action. We denote this by $can(a, \alpha, t)$ with the following additional axiom:

$$\text{Axiom}\ \ \neg can(a, \alpha, t) \rightarrow \neg happens(action(a, \alpha), t)$$

We now define a *consistent plan* below:

Consistent Plan

A *consistent plan* $\rho_{\langle a_1, \ldots, a_n \rangle}$ at time t is a sequence of agents $a_1, \ldots, a_n$ with corresponding actions $\alpha_1, \ldots, \alpha_n$ and times $t_1, \ldots, t_n$ such that $\Gamma \vdash (t < t_i < t_j)$ for $i < j$ and for all

[8] A overview of this list is given lucidly in [McNamara, 2010].

[9] See [Govindarajulu and Bringsjord, 2017a] for an example of how obligations and prohibitions can be used in $\mathcal{DCEC}$.

agents a_i we have:

1. $can(a_i, \alpha_i, t_i)$
2. $happens(action(a_i, \alpha_i))$ is consistent with $\Gamma(t)$.

Note that a consistent plan $\rho_{\langle \ldots \rangle}$ can be represented by a term in our language. We introduce a new sort Plan and a variable-arity predicate symbol $plan(\rho, a_1, \ldots, a_n)$ which says that ρ is a plan involving $a_1 \ldots, a_n$.

A goal is also any formula g. A consistent plan satisfies a goal g if:

$$\left(\Gamma(t) \cup \left\{\begin{matrix} happens(action(a_1, \alpha_1), t_1), \ldots, \\ happens(action(a_n, \alpha_n), t_n) \end{matrix}\right\}\right) \vdash g$$

We use $\Gamma \vdash (\rho \rightarrow g)$ as a shorthand for the above. The above definitions of plans and goals give us the following important constraint needed for defining TAI. This differentiates our planning formalism from other planning systems and makes it more appropriate for an architecture for a general-purpose tentacular AI system.

Uniform Planning Constraint

Plans and goals should be represented and reasoned over in the language of the planning system.

Leveraging the above requirement, we can define two levels of TAI agents. A Level(1) TAI system corresponding to an agent τ is a system that comes up with goal g at time t' to satisfy its contract, produces a proof that there is no consistent plan that involves only the agent τ. Then τ comes with a plan that involves one or more other agents. A Level(1) TAI agent starts with knowledge about what actions are possible for other agents.

Level(1) **TAI Agents**

Prerequisite For any a, α, t, we have:

$$\Gamma \vdash can(a, \alpha, t) \rightarrow \mathbf{K}\big(\tau, t', can(a, \alpha, t)\big)$$

Then

1. τ produces a proof that no plan exists for g involving just itself and τ declares that there is no such plan.

$$\Gamma \vdash \mathbf{S}\big(\tau, t', \neg\exists \rho : (plan(\rho, \tau) \wedge \rho \rightarrow g)\,\big)$$

2. τ produces a plan for g involving just itself and one or more agents and declares that plan.

$$\Gamma \vdash \mathbf{S}\left(\tau, t', \Big(plan(\rho, a_1, \ldots, \tau \ldots a_n) \wedge \rho \rightarrow g\Big)\right)$$

The agent may not always have knowledge about what other agents can do. The TAI agent may have imperfect knowledge about other agents. The agent can gain information about other agents' actions, their obligations, prohibitions, etc. by observing them or by reading specifications governing these agents. In this case, we get a Level(2) TAI agent. We need to modify only the prerequisite condition above.

Level(2) **TAI Agents**

Prerequisite For any a, α, t, we have:

$$\Gamma \vdash can(a, \alpha, t) \rightarrow \mathbf{B}\big(\tau, t', can(a, \alpha, t)\big)$$

The TAI agents above can be considered **first-order** tentacular agents. We can also have a **higher-order** TAI agent that intentionally engages in actions that trigger one or more other agents to act in tentacular fashion as described above. The need for having the uniform planning constraint is more clear when we consider higher-order agents.

6 A Hierarchy of TAI Agents

The TAI formalization above gives rise to multiple hierarchies of tentacular agents. We discuss some of the these below.

Syntactic Goal Complexity The goal g can range in complexity from simple propositional statements, e.g. $cleanKitchen$, to first-order statements. e.g. $\forall r$: Room : $clean(r)$, and to intensional statements representing cognitive states of other agents

$$\mathbf{B}(a, now, \mathbf{B}(b, now, \forall r : clean(r)))$$

Goal Variation According to the definition above, an agent a qualifies as being tentacular if it plans for just one goal g in tentacular fashion as laid out in the conditions above. We could have agents that plan for a number of varied and different goals in tentacular fashion.

Plan Complexity For many goals, there will usually be multiple plans involving different actions (with different costs and resources used) and executed by different agents.

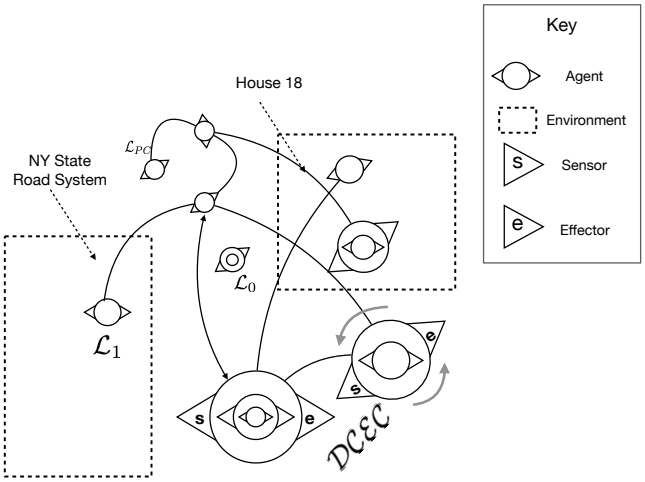

Figure 5: Pictorial Overview. *A bit of explanation: That some agents are within agents indicates that the outer agent knows and/or believes everything relevant about the inner agent; hence as agents are increasingly cognitively powerful, the depth of their epistemic attitudes grows (reflected in formulae with iterated belief/knowledge operators). Agents grow in size/intelligence in lockstep with the logical calculi upon which they are based increasing in expressivity and reasoning power; $\mathcal{L}_0$ is zero-order logic, $\mathcal{L}_1$ is e.g. first-order logic, and the particular cognitive calculus $\mathcal{DCEC}$ is shown. Rotation indicates simply that, through time, agents perceive and act.*

7 Examples and Embryonic Implementation

In this section, we present a formal sketch of a TAI agent and then describe using another example ongoing work in implementing a TAI system.

7.1 Example

Consider the example given in the beginning. We have a human j and three artificial agents: a_c in the car, a_h in the home and a_p an agent managing scheduling and calendar information. We present some of the formulae in Γ.

$$\mathbf{B}(a_c, t_0, crowded(store) \rightarrow unusal), \ \mathsf{f}_1$$

$$\mathbf{P}(a_c, t_1, crowded(store)), \ \mathsf{f}_2$$

$$\forall t : \mathbf{O}\left(\begin{array}{l} a_c, t, unusal, \\ \quad happens\big(action(a_c, check(weather)), t+1\big) \end{array}\right) \ \mathsf{f}_3$$

$$\forall t : \mathbf{B}\left(a_c, t, \ \mathsf{f}_3\right)$$

$$\forall a : \left(\begin{array}{c} happens\Big(action(a, check(weather)), t_3\Big) \\ \rightarrow \mathbf{K}(a, t_4, storm), \end{array}\right) \ \mathsf{f}_4$$

$$\forall t : \mathbf{O}\big(a_c, t, storm, \mathsf{S}(a_c, a_h, storm, t+1))\big), \ \mathsf{f}_5$$

$$\forall t : \mathbf{B}\left(a_c, t, \ \mathsf{f}_5\right)$$

The above formulae first state the fact that a_c observes the store being crowded. a_c's contract states that the agent should check a weather service if it finds something unusual. The formulae also states that if an agent checks the weather at t_3, the agent will get a prediction about an incoming storm. a_c's contract places an obligation on it to inform a_h if it believes that a storm is incoming.

$$\forall t : \mathbf{O}\big(a_h, t, storm, \forall s : quantity(s) > 0\big), \ \mathsf{f}_6$$

$$\mathbf{K}\left(\begin{array}{c} a_h, t_5, shops(j, today) \vee shops(j, tomorrow) \\ \rightarrow \forall s : quantity(s) > 0 \end{array}\right), \ \mathsf{f}_7$$

$$\forall t : \mathbf{B}\left(\begin{array}{l} a_h, t, happens\big(action(a_c, recc(shops(j))), t\big) \\ \quad \rightarrow shops(j) \end{array}\right) \ \mathsf{f}_8$$

$$\forall t : \mathbf{B}\left(\begin{array}{l} a_h, t, happens\Big(action\big(a_h, req(a_c, shops(j))\big), t\Big) \\ \rightarrow happens\big(action(a_c, recc(shops(j))), t\big) \end{array}\right) \ \mathsf{f}_9$$

The first formula above states that a_h ought to see to it that supplies are stocked in the event of a storm. Then we have that a_h knows that the human j shopping today or tomorrow can result in the supplies being stocked. a_h gets information from a_p that shopping tomorrow is not possible (this formula is not shown). Then we have formulae stating the effects of a_c recommending the shopping action to j. The goal for a_h is $\forall s : quantity(s) > 0$ and a plan for it is built up using a_h, a_c and j.

7.2 Toward an Implementation

We describe an example scenario that we are targeting for an embryonic implementation.

Beforehand, a number of contracts have been executed that bind the adult parents P_1 and P_2 in a home H, and also bind a number of artificial agents in H, including a TAI agent (τ) that oversees the home. (Strictly speaking, the agents wouldn't have entered into contracts, but they would know that their human owners have done so, and they would know what the contracts are.)

> It's winter in Berlin NY. Night. Outside, a blizzard. The mother and father of the home H, and their two toddler children, are fast asleep. The smartphone of each parent is set to "Do Not Disturb", with incoming clearance for only close family. There is no landline phone. A carbon monoxide sensor in the basement, near the furnace, suddenly shows a readout indicating an elevated level, which proceeds to creep up. τ perceives this, and forms hypotheses about what is causing the elevated reading, and believes on the basis of using a cognitive calculus that the reading is accurate (to some likelihood factor). The nearest firehouse is notified by τ . No alarm sounds in the house. τ runs a diagnostic and determines that the battery for the central auditory alarm is shot. The reading creeps up higher, and now even the sensors in the upstairs bedrooms where the humans are asleep show an elevated, and climbing, level. τ perceives this too.

Unfortunately, τ reasons that by the time the firemen arrive, permanent neurological damage or even death may well (need again a likelihood factor) be caused in the case of one or more members of the family. Should the alarm company have programmed the sensor to report to a central command, still, any human command is fallible. The company may be negligent, or a phone call may be the only option at their disposal, or they may dispatch personnel who arrive too late. Without enlisting the help of other *artificial* agents in planning and reasoning, τ can't save the family; τ knows this on the basis of proof/argument.

However, τ can likely wake the family up, starting with the parents, in any number of ways. However, each of these ways entails violation of at least one legal prohibition that has been created by contracts that are in place. These contracts have been analyzed by an IBM service, which has stocked the mind of τ with knowledge of legal obligations in $\mathcal{DCEC}$— or rather in a dialect that has separate obligation operators for legal $\mathbf{O}_l$ and moral $\mathbf{O}_m$ obligations. The moral obligation to save the family overrides the legal prohibitions, however. τ turns on the TV in the master bedroom at maximum volume, and flashes a warning to leave the house immediately because of the lethal gas building up. (There are many other alternatives, of course. TAI could break through Do Not Disturb, eg).

7.3 Toward Using Smart-City Infrastructure

The European Initiative on Smart Cities [eur, 2018] is an effort by the European Commission [ec, 2018] to improve the quality of life throughout Europe, while progressing toward energy and climate objectives. Many of its goals are relevant to and desirable in the world at large. TAI has the potential to be instrumental in achieving many of these, such

as smart appliances (in the manner discussed in the previous sub-section) and intelligent traffic management. Indeed, the scope and objectives of the Initiative may conceivably be considerably broadened with a pervasive application of TAI.

We briefly point at a simple scenario that expands on the vision of the European Initiative's smart-transportation goals.

> Parking space is very scarce on a work-day in midtown Manhattan. A busy executive will need to park near several offices over the course of the day, and these locations change over the week.

The executive's car consults her calendar. Based on past patterns, it interpolates locations where it believes she intends to park. It communicates with other cars parked at these locations, and determines when their owners are likely to return, based on their expressed (and inferable) intentions and current locations. Adjusting for the location of our executive, traffic conditions and changes in her agenda, it determines the optimal parking locations dynamically, throughout her busy day. Of course, in the spirit of TAI, all other cars would have their movement adjusted accordingly, through time.[10]

8 Conclusion & Future Work

We have introduced Tentacular AI, and a number of architectural elements thereof, and are under no illusion that we have accomplished more than this. At AEGAP 2018, we will demonstrate TAI in action in both the scenarios sketched above; implementation is currently underway. Despite the nascent state of the TAI research program, we hope to have provided a promising, if inchoate, overview of tentacular AI — an overview which, given the centrality of highly expressive languages for novel planning and reasoning, we hope is of interest to some, maybe even many, at this dawn of the "internet of things" and its vibrant intersection with AI.

9 Acknowledgments

The TAI project is made possible by joint support from RPI and IBM under the AIRC Program; we are grateful for this support. Some of the research reported on herein has been enabled by support from ONR and AFOSR, and for this too we are grateful.

References

[Arkoudas and Bringsjord, 2008] K. Arkoudas and S. Bringsjord. Toward Formalizing Common-Sense Psychology: An Analysis of the False-Belief Task. In T.-B. Ho and Z.-H. Zhou, editors, *Proceedings of the Tenth Pacific Rim International Conference on Artificial Intelligence (PRICAI 2008)*, number 5351 in Lecture Notes in Artificial Intelligence (LNAI), pages 17–29. Springer-Verlag, 2008.

[Arkoudas and Bringsjord, 2009] K. Arkoudas and S. Bringsjord. Propositional Attitudes and Causation. *International Journal of Software and Informatics*, 3(1):47–65, 2009.

[Arkoudas and Musser, 2017] Konstantine Arkoudas and David Musser. *Fundamental Proof Methods in Computer Science: A Computer-Based Approach*. MIT Press, Cambridge, MA, 2017.

[Bringsjord and Ferrucci, 2000] S. Bringsjord and D. Ferrucci. *Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, a Storytelling Machine*. Lawrence Erlbaum, Mahwah, NJ, 2000.

[Bringsjord and Govindarajulu, 2012] S. Bringsjord and N. S. Govindarajulu. Given the Web, What is Intelligence, Really? *Metaphilosophy*, 43(4):361–532, 2012. This URL is to a preprint of the paper.

[Bringsjord and Sen, 2016] Selmer Bringsjord and Atriya Sen. On Creative Self-Driving Cars: Hire the Computational Logicians, Fast. *Applied Artificial Intelligence*, 30:758–786, 2016. The URL here goes only to an uncorrected preprint.

[Bringsjord *et al.*, 2014] Selmer Bringsjord, Naveen Sundar Govindarajulu, Daniel Thero, and Mei Si. Akratic Robots and the Computational Logic Thereof. In *Proceedings of ETHICS • 2014 (2014 IEEE Symposium on Ethics in Engineering, Science, and Technology)*, pages 22–29, Chicago, IL, 2014. IEEE Catalog Number: CFP14ETI-POD.

[Bringsjord *et al.*, 2018] S. Bringsjord, P. Bello, and N.S. Govindarajulu. Toward Axiomatizing Consciousness. In D. Jacquette, editor, *The Bloomsbury Companion to the Philosophy of Consciousness*, pages 289–324. Bloomsbury Academic, London, UK, 2018.

[Bringsjord, 2015] Selmer Bringsjord. Theorem: *General Intelligence Entails Creativity*, assuming . . .. In T. Besold, M. Schorlemmer, and A. Smaill, editors, *Computational Creativity Research: Towards Creative Machines*, pages 51–64. Atlantis/Springer, Paris, France, 2015. This is Volume 7 in *Atlantis Thinking Machines*, edited by Kuhnbergwer, Kai-Uwe of the University of Osnabruck, Germany.

[Chella and Manzotti, 2007] Antonio Chella and Ricardo Manzotti, editors. *Artificial Consciousness*. Imprint Academic, Exeter, UK, 2007.

[Dwork, 2008] Cynthia Dwork. Differential Privacy: A Survey of Results. In *International Conference on Theory and Applications of Models of Computation*, pages 1–19. Springer, 2008.

[ec, 2018] The European Commission's Priorities. https://ec.europa.eu/commission/index_en, 2018. [Online; accessed 25-June-2018].

[eur, 2018] European Initiative on Smart Cities. https://setis.ec.europa.eu/set-plan-implementation/technology-roadmaps/european-initiative-smart-cities, 2018. [Online; accessed 25-June-2018].

[10] TAI applications like this give rise to privacy concerns which could possibly be resolved by employing either **differential privacy** [Dwork, 2008] or privacy based on **zero-knowledge proofs** [Gehrke *et al.*, 2011].

[Francez and Dyckhoff, 2010] Nissim Francez and Roy Dyckhoff. Proof-theoretic Semantics for a Natural Language Fragment. *Linguistics and Philosophy*, 33:447–477, 2010.

[Gehrke *et al.*, 2011] Johannes Gehrke, Edward Lui, and Rafael Pass. Towards Privacy for Social Networks: A Zero-Knowledge Based Definition of Privacy. In *Theory of Cryptography Conference*, pages 432–449. Springer, 2011.

[Genesereth and Nilsson, 1987] M. Genesereth and N. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, Los Altos, CA, 1987.

[Gentzen, 1935] Gerhard Gentzen. Investigations into Logical Deduction. In M. E. Szabo, editor, *The Collected Papers of Gerhard Gentzen*, pages 68–131. North-Holland, Amsterday, The Netherlands, 1935. This is an English version of the well-known 1935 German version.

[Gerevini and Long, 2004] Alfonso Gerevini and Derek Long. Plan Constraints and Preferences in PDDL3. Technical report, Department of Electronics for Automation, University of Brescia, 2004. This is the language of the Fifth International Planning Competition.

[Govindarajulu and Bringsjord, 2017a] Naveen Sundar Govindarajulu and Selmer Bringsjord. On Automating the Doctrine of Double Effect. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4722–4730, Melbourne, Australia, 2017. Preprint available at this url: https://arxiv.org/abs/1703.08922.

[Govindarajulu and Bringsjord, 2017b] Naveen Sundar Govindarajulu and Selmer Bringsjord. Strength Factors: An Uncertainty System for a Quantified Modal Logic, 2017. Presented at Workshop on Logical Foundations for Uncertainty and Machine Learning at IJCAI 2017, Melbourne, Australia.

[Govindarajulu *et al.*, 2017] Naveen Sundar Govindarajulu, Selmer Bringsjord, Rikhiya Ghosh, and Matthew Peveler. Beyond the doctrine of double effect: A formal model of true self-sacrifice. International Conference on Robot Ethics and Safety Standards, 2017.

[Gupta, 1992] Naresh Gupta. On the Complexity of Blocks-world Planning. *Artificial Intelligence*, 52:223–254, 1992.

[Helgason *et al.*, 2012] Helgi Helgason, Eric Nivel, and Kristinn Thórisson. On Attention Mechanisms for AGI Architectures: A Design Proposal. In J. Bach, B. Goertzel, and M. Iklé, editors, *Proceedings of the Fifth Conference on Artificial General Intelligence*, pages 89–98, Berlin, Germany, 2012. Springer.

[Kovacs, 2012] Daniel L. Kovacs. A Multi-Agent Extension of PDDL3.1. In *Proceedings of the 3rd Workshop on the International Planning Competition (IPC), ICAPS- 2012*, pages 25–29, Atibaia, Brazil, 2012. ICAPS.

[Levesque and Lakemeyer, 2007] Hector Levesque and Gerhard Lakemeyer. Chapter 24: Cognitive Robotics. In *Handbook of Knowledge Representation*, Amsterdam, The Netherlands, 2007. Elsevier.

[Luger, 2008] George Luger. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving (6th Edition)*. Pearson, London, UK, 2008.

[Mcdermott *et al.*, 1998] D. Mcdermott, M. Ghallab, A. Howe, C. Knoblock, A. Ram, M. Veloso, D. Weld, and D. Wilkins. PDDL – The Planning Domain Definition Language. Technical Report CVC TR-98-003, Yale Center for Computational Vision and Control, 1998.

[McNamara, 2010] P. McNamara. Deontic Logic. In Edward Zalta, editor, *The Stanford Encyclopedia of Philosophy*. 2010. McNamara's (brief) note on a paradox arising from Kant's Law is given in an offshoot of the main entry.

[Miller *et al.*, 2018] Tim Miller, Adrian R. Pearce, and Liz Sonenberg. *Social Planning for Trusted Autonomy*, pages 67–86. Springer International Publishing, Cham, 2018.

[Mueller, 2006] E. Mueller. *Commonsense Reasoning: An Event Calculus Based Approach*. Morgan Kaufmann, San Francisco, CA, 2006. This is the first edition of the book. The second edition was published in 2014.

[Rao and Georgeff, 1991] A. S. Rao and M. P. Georgeff. Modeling Rational Agents Within a BDI-architecture. In R. Fikes and E. Sandewall, editors, *Proceedings of Knowledge Representation and Reasoning (KR&R-91)*, pages 473–484, San Mateo, CA, 1991. Morgan Kaufmann.

[Russell and Norvig, 2009] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2009. Third edition.

[Seidita *et al.*, 2016] Valeria Seidita, Antonio Chella, and Maurizio Carta. A Biologically Inspired Representation of the Intelligence of a University Campus. *Procedia Computer Science*, 88:185–190, 2016.

[Wang, 2006] Pei Wang. *Rigid Flexibility: The Logic of Intelligence*. Springer, Dordrecht, The Netherlands, 2006. This book is Volume 34 in the *Applied Logic Series*, edited by Dov Gabbay and Jon Barwise.