A Calculus for Robot Inner Speech and Self-Awareness

A. Pipitone[1] [0000000323885887], A. Chella[1,2] [000000028625708X]

[1] Dept. of Engineering, University of Palermo, Italy
[2] C.N.R., Institute for High-Performance Computing and Networking (ICAR), Palermo, Italy
{arianna.pipitone, antonio.chella}@unipa.it

Abstract. The inner speech is the common mental experience the humans have when they dialogue with themselves. It is widely acknowledged that inner speech is related to awareness and self-awareness. The inner speech reproduces and expands in the mind social and physical sources of awareness. In this preliminary work, a calculus based on a first-order modal logic to automate inner speech is presented. It attempts to make the existing inner speech theories suitable for robot. By making robot able to talk to itself, it is possible to analyze the role of inner speech in robot awareness and self-awareness, opening new interesting research scenarios not yet investigated.

1 Introduction

The dialogue with the self plays a fundamental role in human's consciousness [2] [1] [3], and is linked to self-awareness: by talking to himself, someone accesses to self-information, or extends existing self-knowledge [4].

Morin [4] stated three main causal directions between inner speech and self-awareness: (i) inner speech always precedes (causes) self-awareness, (ii) inner speech accompanies a state of awareness, and (iii) inner speech is triggered by self-focus, that is the attention on the self.

When analyzing the robot awareness and self-awareness, it seems desirable to provide the robot with a kind of self dialogue. Few works investigate this scenario. The authors proposed a cognitive architecture [5] for inner speech, that is modeled as the rehearsing process between a working-memory and a motor module wich produces language. Same authors suggested to integrate their architecture into the IDyOT system [6]. Other works [14] demonstrated that the language re-entrance, that is a form of inner speech, allows to refine the emergent grammar shared by a population of agents.

We attempt to automate the Morin's causal directions by defining a calculus couched in first-order modal logic. The **Deontic Cognitive Event Calculus** ($\mathcal{DCEC}$) [7] underlies the proposed one which includes some of the $\mathcal{DCEC}$'s elements while adding new propositions and terms for formalizing inner speech. $\mathcal{DCEC}$ subsumes in turn the Event Calculus (EC) [8], and it was proposed to formalize thorny situations, such as *akrasia* [10] and the *Doctrine of Double Effect* ($\mathcal{DDE}$) [9].

We show that our formalization may affect the execution by robot of a simple task when a self-perception stimuli occurs and triggers inner speech. The idea we propose lays the groundwork for new investigations related to the robot's inner speech ability, and for testing its effects on robot awareness and self-awareness.

The paper is organized as follow: a brief overview about $\mathcal{DCEC}$ is presented at section 2. Then, how we think to encode inner speech and self-awareness by EC is explained at section 3. The syntax of our calculus (with sorts, functions and axioms definitions) is detailed at section 4. Inference schemata and formal conditions, that allow to make deduction and to reason on the syntax, are discussed at section 5. The section 6 shows a simple simulation for reasoning by the proposed calculus. Future works and implications are discussed at 7.

2 The Deontic Cognitive Event Calculus ($\mathcal{DCEC}$)

The common Event Calculus (EC) [8] is a sorted logic analogous to a typed programming language. It allows to formalize intuitive reasoning about actions and the changes which occur in the universe after doing these actions. The notion of 'action affecting' arises, meaning the **causal** influences of actions on the states of the universe.

$\mathcal{DCEC}$ [7] is a calculus that subsumes EC, by adding new operators and functions to enable **intensional** reasoning processes. The intensional property is necessary when modeling the typical artificial agent's states (such as knowledge, belief and intention), which are intensional (i.e. it is not possible to declare and to know all the values of these states). The models based on extensional property, which attempt to list all values, was demonstrated [15] generate inconsistencies for these states.

The $\mathcal{DCEC}$'s intensional processes was successfully used to automate the false-belief task [13], the *akrasia* situation [10] (which represents the temptation to violate moral principle) and the $\mathcal{DDE}$ situation [9] (which arises when a dilemma with positive or negative effects has to be solved by an autonomous agent). The intensional property of the model is achieved by the inclusion in the formalization of intensional modal operators. For example, the intensional modal operator for modeling the knowledge of a fact has the typical form $\mathbf{K}(a, t, \phi)$, which means that agent a knows the proposition ϕ at time t. An instance of this operator allows to model an intentional knowledge state of the agent.

Involving same kinds of states, the calculus for inner speech and self-awareness is an intensional model, and we inspired to $\mathcal{DCEC}$ for our work.

3 Encoding Inner Speech and Self-Awareness

As shown, Morin's theory [4] claims the causal directions between inner speech and self-awareness. Because of this causality property, EC is suitable to automate such directions.

Since a fluent represents a state (or fact) of the world, we can define *inner* and *external* fluents. An inner fluent represents an internal agent state, eg: an emotion, its intensity value, the status of an its physical component (on, off, functioning,...). An external fluent is related to the external context, and represents a fact of the environment, that is the typical definition of EC's fluent. Thus, how a typical fluent is initiated or terminated by an action which changes the environmental set (named *reactive action*), similarly an internal fluent is initiated or terminated by an internal action (named *inner action*, i.e. a self-regulation action, the perception of an entity, the rehearse of inner voice,...) which changes the internal set.

We think that *awareness and self-awareness of an artificial agent could be computationally modeled as all the fluents (both inner and external) the agent knows to be active at a time.* A fluent is active at time t when it holds at t. The active fluents that are not in the knowledge of the agent are not in the set modeling awareness and self-awareness: the agent does not know them, and it is not aware of them.

Now, let suppose that the object x at the instant t_1 is at the location l_1. The native EC function

$$holds(location(x, l_1), t_1)$$

models the state of the world related to that fact, that is the active state of the fluent $location(x, l_1)$ at the instant t_1.

The action $a = move\ x\ to\ l_2$ at the instant t_2 with $t_1 < t_2$, causes the end of the previous fluent

$$terminates(a, location(x, l_1), t_2)$$

and the beginning of the new fluent $location(a, l_2)$ by

$$initiates(a, location(a, l_2), t_2)$$

and for the inferential logic of EC, the creation of a new function *holds* related to such a new fluent.

When such an action has to be performed by an autonomous agent, it may involve inner fluents and statements beyond the previously seen ones. The agent will use inner speech as a cognitive tool [11] to accomplish the task $a = move\ x\ to\ l_2$. For example, it has to evaluate its abilities to solve the problem, the position of the object in respect to its position, the form of the object, the feasibility of the action, the state of its physical components for making the action. To answer to the questions it makes to itself, the robot has to retrieve useful self-information and information from the environment.

Kendall et. al [12] proposed four categories of self-questions emerging during the problem-solving process: (1) the questions for a clear formulation of the problem ('What's the problem?', 'What's I'm supposing to do?'), (2) the questions for proposing a possible solution ('I have to find a strategy'), (3) the questions for focusing on relevant aspects for the solution ('It's important', 'It's not important, I discard it'), (4) the statements for praising oneself when the solution is reached ('Good! I find the solution!) or for readjusting the approach when someone fails ('Oh, no! It's no ok').

All these evaluations will modify the set of active fluents, and hence the state of awareness and self-awareness of the agent.

The calculus we propose attempt to model such evaluations. It could be extensible by adding fluents representing further internal or external facts.

4 The Proposed Calculus

Commonly used functions, sorts and relation symbols from EC and some ones from $\mathcal{DCEC}$ are included in our formalization. The resulted calculus has a unique syntax and a proof calculus for reasoning and theorem proving. In particular, the *natural deduction* by Gentzen [16] is used as prover, as shown at section 6.

4.1 Modal Fragment

The modal fragment specifies the modal operators of the calculus.

While $\mathcal{DCEC}$ includes the standard modal operators for *common-sense knowledge* $\mathbf{C}$ and *domain knowledge* $\mathbf{K}_d$, in our formalization we include other two types of knowledge by adding two new modal operators, representing:

- the *self-knowledge* $\mathbf{K}_{self}$, that is the knowledge the robot owns about itself, (i.e. *inner world*). Then, $\mathbf{K}_{self}(a, t, \phi)$ means that the agent a knows the proposition ϕ representing an inner fact at time t;
- the *contextual-knowledge* $\mathbf{K}_{cx}$, that is the knowledge about the physical environment in which the robot is plunged, (i.e. *external world*). The contextual-knowledge we considered is not equal to the general one $\mathbf{K}$, because it may include some new objects the robot does not know (it has never seen them). Moreover, the contextual knowledge may not include some concepts of the general one because these concepts could not be in the environment at a time. $\mathbf{K}_{cx}(a, t, \phi)$ means that agent a knows the proposition ϕ representing an external fact at time t.

The *general knowledge* we refer to becomes

$$\mathbf{K} = \mathbf{K}_{self} \vee \mathbf{K}_d \vee \mathbf{K}_{cx}$$

The standard intensional operators for belief $\mathbf{B}$, desire $\mathbf{D}$, intention $\mathbf{I}$ are included too, and have the same semantics from $\mathcal{DCEC}$.

Other $\mathcal{DCEC}$ modals operators we include are $\mathbf{P}$ for perceiving a state, and $\mathbf{S}$ for agent-to-agent communication or public announcement. But the $\mathbf{S}$ operator we define crucially differs from those of $\mathcal{DCEC}$ because we consider the possibility to have the same agent (the robot) in its agent-to-agent arguments, leading to the inner speech formalization. The single argument means public announcement as for $\mathcal{DCEC}$. Formally:

- $\mathbf{S}(a, b, t, \phi)$ means that the agent a sends the message related to proposition ϕ to the agent b at time t;
- $\mathbf{S}(a, a, t, \phi)$ means that the agent a rehearses the message related to proposition ϕ it sends to itself at time t;

- $\mathbf{S}(a, t, \phi)$ means that the agent a makes a public announcement related to proposition ϕ at time t.

Finally, we add the modal operator $\mathbf{M}$ for producing a message related to a specific proposition . Then, $\mathbf{M}(a, t, \phi)$ means that the agent a produces the message related to the proposition ϕ at time t. We have to specify that $\mathbf{M}$ regards the action "to produce a message" which can be in turn sent to another agent or rehearsed according to $\mathbf{S}$.

4.2 Sorts Specification

We define new sorts upon the native EC and $\mathcal{DCEC}$ ones. Moreover, we changed the typical `Agent` and `Entity` sorts definitions because we consider the agent as a part of the universe, hence the `Agent` sort becomes a sub-type of the `Entity` sort. The *abstract* sorts are instantiated at particular times by an actors.

The following table shows all the sorts we defined. The highlighted sorts are the new ones introduced or modified by our calculus.

Sort	**Description**
`Entity`	An entity in the universe, including agent.
`Object`	A subtype of `Entity`, representing a physical object in the environment that is not an actor.
`Agent`	A subtype of `Entity`, representing human and artificial actors.
`Percept`	A perception from the environment; it can be a concrete `Entity` (an `Agent` or an `Object`), or a generic concept meaning an event.
`Moment`	A time in the domain.
`Event`	An event in the domain.
`ActionType`	An abstract *reactive action*, i.e. an action that affects the state of the external environment.
`SelfActionType`	An abstract *inner action*, i.e. an action that affects the inner state of the agent. Examples: keep calm, evaluate, feel.
`Action`	A subtype of `Event` that occurs when an agent performs a concrete `ActionType` action.
`SelfAction`	A subtype of `Event` that occurs when an agent performs a `SelfActionType` action.

`Fluent`	A state of the universe, that can be inner or external.
`Aware`	A subtype of `Fluent`, representing a fluent the agent knows. The set of active `Aware` fluents forms the agent awareness and/or self-awareness.
`Message`	A message of an agent-to-agent, agent-to-itself, agent-to-public communication.

4.3 Defining the message m

EC and $\mathcal{DCEC}$ have not any operators or functions to define the message of a communication. We take a modal approach for modeling such a situation without defining specific axioms. We assume that a message is related to a specific fact, represented by a proposition. Given a proposition ϕ, we define by the operator $\odot$ the set of all constants and names of not native function expressions in ϕ. For example:

$$\odot(happens(action(Bob, loves(Mary))), t) = \{Bob, loves, Mary\}$$
$$\odot(holds(located(apple, table)), t) = \{located, apple, table\}$$

A message is the set returned by the $\odot$ operator, that is:

$$\odot : \phi \to m$$

So, given ϕ the corresponding message m will be $m = \odot(\phi)$.
A **content** of the message is an element in the $\odot$ set. So, the content $p_i \in m$ is the i-th element in m.

4.4 The Syntax

The whole syntax of the calculus is formalized by the formulas in the following table, where S represents the sorts, f represents the functions, the *term* represents the possible variables and finally ϕ are the propositions.

$$S ::= \begin{array}{l} \texttt{Entity}\ |\texttt{Agent} \sqsubseteq \texttt{Entity} \mid \texttt{Object} \sqsubseteq \texttt{Entity} \mid \texttt{Percept} \sqsubseteq \texttt{Entity} \mid \\ \texttt{Percept} = \texttt{Agent} \sqcup \texttt{Object} \sqcup \texttt{Action} \sqcup \texttt{SelfAction} \mid \texttt{Fluent} \mid \\ \texttt{ActionType} \mid \texttt{SelfActionType} \mid \texttt{Event} \mid \texttt{SelfAction} \sqsubseteq \texttt{Event} \mid \\ \texttt{Action} \sqsubseteq \texttt{Event} \mid \texttt{Boolean} \mid \texttt{Moment} \mid \texttt{Message} \end{array}$$

$$f ::= \begin{cases} \mathit{action} : \texttt{Agent} \times \texttt{ActionType} \to \texttt{Action} \\ \mathit{initially} : \texttt{Fluent} \to \texttt{Boolean} \\ \mathit{holds} : \texttt{Fluent} \times \texttt{Moment} \to \texttt{Boolean} \\ \mathit{happens} : \texttt{Event} \times \texttt{Moment} \to \texttt{Boolean} \\ \mathit{clipped} : \texttt{Moment} \times \texttt{Fluent} \times \texttt{Moment} \to \texttt{Boolean} \\ \mathit{initiates} : \texttt{Event} \times \texttt{Fluent} \times \texttt{Moment} \to \texttt{Boolean} \\ \mathit{terminates} : \texttt{Event} \times \texttt{Fluent} \times \texttt{Moment} \to \texttt{Boolean} \\ \textbf{\textit{selfaction}} : \texttt{Agent} \times \texttt{SelfActionType} \to \texttt{SelfAction} \\ \textbf{\textit{focuses}} : \texttt{Percept} \to \texttt{SelfActionType} \\ \textbf{\textit{aware}} : \ \texttt{Agent} \times \texttt{Percept} \to \texttt{Fluent} \\ \textbf{\textit{content}}_i : \texttt{Message} \to \texttt{Percept} \\ \textbf{\textit{comprehends}} : \texttt{Message} \to \texttt{SelfActionType} \\ \textbf{\textit{produces}} : \texttt{Message} \to \texttt{SelfActionType} \\ \textbf{\textit{innerspeaks}} : \texttt{Message} \to \texttt{SelfActionType} \end{cases}$$

$$\mathit{term} ::= \mathit{var} : S \mid \mathit{const} : S \mid f(\mathit{term}_1, \mathit{term}_2, ...\mathit{term}_n)$$

$$\phi ::= \begin{cases} \mathit{term}\text{:}\texttt{Boolean} \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \mathbf{P}(a,t,\phi) \mid \mathbf{K}(a,t,\phi) \mid \\ \mathbf{C}(t,\phi) \mid \mathbf{B}(a,t,\phi) \mid \mathbf{D}(a,t,\phi) \mid \mathbf{I}(a,t,\phi) \\ \\ \mathbf{S}(a,b,t,\phi) \mid \mathbf{S}(a,a,t,\phi) \mid \mathbf{S}(a,t,\phi) \\ \\ \mathbf{M}(a,t,\phi) \end{cases}$$

The functions in bold are purposely defined for formalizing inner speech, awareness and self-awareness, while the others are native from EC.

These new functions are :

- *selfaction(a, a_t)*$\to$ a_s: that returns the concrete self action a_s of type a_t the agent a performs; it is the self-version of native *action* function of EC;
- *focuses(p)*$\to$ a_t: that represents the selfaction type the agent a performs when it focuses on the percept p. Intuitively, $selfaction(a, focuses(p))$ means the action by a of focusing on the percept p;
- *aware(p)*$\to$ f: that states the awareness about the percept p. The activation of that fluent by $holds(aware(p), t)$ means that an agent is aware of the percepts p at t. Let's notice that the *aware* fluents may be a sub-set of the all active fluents at a time. They are not the whole robot awareness and self-awareness, but are the robot awareness about the percepts.
- *content$_i$(m)*$\to$ p_i: that returns the i-th percept p_i in the message m;

- *comprehends*[3] *(m)*$\rightarrow a_t$: that states the comprehension of the message. Intuitively, $selfaction(a, comprehends(m))$ means the action by a of comprehending the message m;
- *produces*[4] *(m)* $\rightarrow a_t$: that states that the agent a is producing the message m. Intuitively, $selfaction(a, produces(m))$ means the action by a of producing the message m.

The typical truth-functional connectives $\wedge$, $\vee$, $\neg$, $\rightarrow$ are applied to propositions and they have the standard first-order semantics.

4.5 Axioms

As for $\mathcal{DCEC}$, the standard axioms of EC are considered as common-knowledge, that is:

$[A_1]$ $\mathbf{C}(\forall\ f,\ t\ .\ initially(f) \wedge \neg clipped(0, f, t) \Rightarrow holds(f, t))$
$[A_2]$ $\mathbf{C}(\forall\ e,\ f,\ t_1,\ t_2\ .\ happens(e, t_1)\ \wedge\ initiates(e, f, t_1)\ \wedge\ t_1 < t_2\ \wedge \neg clipped(t_1, f, t_2) \Rightarrow holds(f, t_2))$
$[A_3]$ $\mathbf{C}(\forall\ f,\ t_1,\ t_2\ .\ clipped(t_1, f, t_2) \iff [\ \exists\ e,\ t\ .\ happens(e, t) \wedge t_1 < t < t_2 \wedge terminates(e, f, t)])$
$[A_4]$ $\mathbf{C}(\forall\ a,\ d,\ t\ .\ happens(action(a, d), t) \Rightarrow \mathbf{K}(a, happens(action(a, d), t), t))$

The axioms from $[A_1]$ to $[A_3]$, that are native of EC, state general and innate understanding human capacity about the causality of events. $[A_4]$ is native of $\mathcal{DCEC}$ and postulates that an agent knows the action it performs, that is it means the event which occurs by doing such an action.

We postulate one more axiom pertains the inner speech triggering. It expresses that when an agent rehearses a message it produces, the inner speech happens. That is:

$[A_5]$ $\mathbf{C}(\forall\ a,\ m,\ t_1,\ t_2\ .\ happens(selfaction(a, produces(m)), t_1) \wedge$
$happens(selfaction(a, comprehends(m)), t_2) \wedge\ t_1 < t_2$
$\Leftrightarrow happens(selfaction(a, innerspeaks(m)), t_2))$

We define three more axioms that, according to Morin's theory, state the three main causal directions between inner speech and self-awareness. Considering that inner speech is an unconscious activity [14] despite it triggers self-awareness, we do not model these axioms as common-knowledge.

As result:

$[A_6]$ $\forall\ a,\ \phi,\ m,\ (\forall\ \ p_i \in m),\ t\ .\ initiates(selfaction(a, innerspeaks(m)), aware(a, p_i), t)$

[3] The term *comprehends* is the name of the function defined in the Fluid Construction Grammar engine [17] for parsing an input utterance. Such a function returns the meaning of the sentence.

[4] The term *produces* is the name of the function defined in the Fluid Construction Grammar engine [17] for producing an output sentence given a set of meanings. Such a function returns the syntactic form of the conjunctions of the meanings.

$[A_7]\ \forall a,\ t_1,\ t_2,\ m,\ p_i \in m\ .\ clipped(t_1, aware(a,p_i), t_2)\ \Rightarrow$
$\exists\ t \in [t_1,t_2]\ ,\ happens(selfaction(a, innerspeaks(m)), t)$
$[A_8]\ \forall a,\ t,\ m,\ p_i \in m\ .\ happens(selfaction(a, focuses(p_i)), t) \Rightarrow$
$happens(selfaction(a, innerspeaks(m)), t)$

$[A_6]$ postulates that inner speech precedes (causes) awareness. The formula means that when the self-action related to the comprehension of a message produced by itself (i.e. rehearsed) has taken place at time t (*happens(e,t)*), the corresponding event has the aware fluent about the content of the message as an effect. More specifically, $[A_6]$ formalizes that 1 precedes 2 in the following:

1. the agent a produces a message m:
   produces(m)
   and it rehearses such a message:
   comprehends(m)
   generating the event by axiom $[A_5]$:
   selfaction(a, innerspeaks(m))
   which then has taken place at t:
   happens(selfaction(a, innerspeaks(m)), t)
2. the agent becomes aware of each content of the message, generating the corresponding fluents of awareness:
   aware(a, p_i)
   hence the previous event has each of this fluent as an effect:
   initiates(selfaction(a, innerspeaks(m)), aware(a, p_i), t)

$[A_7]$ postulates that inner speech accompanies a state of awareness.
That is:

1. If the robot is aware of the percept p_i, and if the corresponding fluent:
   aware(a, p_i)
   has not been made false in the time interval $[t_1, t_2]$:
   clipped(t_1, aware(a, p_i), t_2)
2. the inner action to rehearse itself has taken in the meanwhile:
   $\exists\ t \in [t_1, t_2]$
   happens(selfaction(a, innerspeaks(m)), t)
   with m related to p_i, i.e. $p_i \in m$.

Finally, $[A_8]$ states that inner speech is triggered by *focus*, that is a specific inner action. Then:

1. If the agent a focuses on a percept p at time t:
   happens(selfaction(a, focus(p)), t)
2. the inner speech action starts, begin m the message whose content is p:
   selfaction(a, innerspeaks(m))
   with $p \in m$

The following axiom states that is common-knowledge that when an agent perceives a fact, it is focusing almost one percept which will be related to that fact:

$[A_9]$ $\mathbf{C}(\forall a,\ t,\ \exists\ (\phi,\ p)\ .\ \mathbf{P}(a,t,\phi) \Leftrightarrow\ \mathit{happens}(\mathit{selfaction}(a, \mathit{focuses}(p))))$

Considering that the action to formulate a message generates the corresponding message m, we also postulate the axiom $[A_{10}]$:

$[A_{10}]$ $\mathbf{C}(\forall a,\ t,\ \exists\ (\phi,\ m)\ .\mathbf{M}(a,t,\phi) \Rightarrow\ \mathit{happens}(\mathit{selfaction}(a, \mathit{produces}(m))))$.

Finally, the awareness about itself, leads to the self-knowledge of the proposition triggering it, so:

$[A_{11}]$ $\mathbf{C}(\forall a,\ t\ .\ \mathit{holds}(\mathit{aware}(a),t) \Leftrightarrow [\exists\ \phi\ .\ \mathbf{K}_{self}(a,t,\phi)])$.

5 Inference Schemata

Some of the inference rules we define for reasoning by the calculus are:

$[R_1]$ $\frac{\mathbf{M}(a,t,\phi)}{\mathbf{S}(a,a,t,\phi)}$

The generation of a message **M** about ϕ leads to rehearse it.

$[R_2]$ $\frac{\mathbf{P}(a,t,\phi)}{\mathbf{M}(a,t,\phi)}$

The perception of ϕ leads to a message about ϕ.

$[R_3]$ $\frac{\mathbf{K}_{self}(a,t,\phi)}{\mathbf{K}_{self}(a,t,\ \mathbf{K}_{self}(a,t,\phi))}$

It captures an essential property of self-awareness: the knowledge of self about ϕ leads to the knowledge about such a knowledge. That is "I know to know".

$[R_4]$ $\frac{\mathbf{K}(a,t,\phi)}{\phi}$

The knowledge of ϕ implies ϕ.

6 Simulation

A concrete example allows us to clarify and to show the application of the calculus we propose. The scenario we consider is inspired to those described at [11]: inner speech is conceived as a cognitive tool the individual uses for self-reflection. In this case, the self is the object of the questions ('Who am I?', 'What am I doing?'), and the self-knowledge and information from the environment are the answers. A person engaged in a task might self-reflect by this kind of inner dialogue, and it was demonstrated it facilitates the task execution.

Therefore, we describe the method for reasoning by the proposed calculus about self-reflection during task execution. The knowledge of the robot changes by self-reflection in respect to the knowledge of the typical reactive behavior, putting the robot in the conditions to make more inferences than the case without self-reflection.

At this step, our goal is to demonstrate how the robot awareness and self-awareness grow by modeling inner speech by the proposed calculus implementing self-reflection. How the questions emerge, and the answers are produced are out of the scope of this paper, and regard important future works.

6.1 Encoding self-reflection

We suppose that the robot is engaged in a simple task, that is to remove an object from its locations and to put it in a different location. In the reactive behavior, the robot runs a set of routines which allow: to identify the object o, to move to the location l, to grasp the object o from the location l, to place the object o in the location l.

Let suppose that at a certain time during task execution, the robot perceives itself by a mirror, and it sees itself to perform one of the described actions. In particular, it sees itself to grasp the object.

It will be engaged in the following soliloquy: “What am I doing? Did I perform yet this task?”.

We introduce the sort `Location`, and the following function symbols for reasoning about this situation:

$$grasp : \texttt{Object} \rightarrow \texttt{ActionType}$$

$$remember : \texttt{ActionType} \times \texttt{Moment} \rightarrow \texttt{Boolean}$$

The set of questions will generate a set of corresponding propositions:

“What am I doing?” $\Rightarrow$ “I grasp the object o”
$\Rightarrow$ ***happens(action(a, grasp(o)),t)***
“Did I perform this task?” $\Rightarrow$ “Yes, I do!”
$\Rightarrow$ ***remember(grasp(o), t)***

We consider two temporal lines which allow to specify the narrative of self-reflection, that are the times before inner speech and the time after inner speech. Begin t_1 the limit point, for all times before t_1 (i.e. $t < t_1$), the robot performs the task in reactive way, i.e. without self-reflection. At $t = t_1$, the inner speech triggers. For all times after t_1 (i.e. $t > t_1$), the robot is provided with self-reflection ability by inner speech. Therefore, the robot upon t_1 will know that until this moment it was not aware of the task: then it will be conscious that it has carried out the specific action.

6.2 Reasoning

Now, let suppose that an event triggers inner speech. The robot perceives itself by a mirror.

1. The starting premise is that the robot a sees itself by mirror to perform the action α at time t:

$$\mathbf{P}(a, t, happens(action(a, \alpha), t))$$

2. As consequence, from A_9 it focuses on itself:

$$happens(selfaction(a, focuses(a), t))$$

   Other kinds of percepts may activate the focus, as the object to take.
3. Given A_9, from A_8 the inner speech starts:

$$happens(selfaction(a, innerspeaks(\odot(happens(action(a, \alpha)))), t)$$

4. Hence, from A_5 the following events take place:

$$e_1 = happens(selfaction(a, produces(\odot(happens(action(a, \alpha))))))$$

$$e_2 = happens(selfaction(a, comprehends(\odot(happens(action(a, \alpha))))))$$

5. From A_6, the fluents of awareness start:

$$\forall p_i \in \odot(happens(action(a, \alpha))) \rightarrow initiates(e, aware(p_i))), t)$$

6. Form A_{10} the previous fluents extend the knowledge about the self because ϕ involves a:

$$\mathbf{K}_{self}(a, t, happens(action(a, \alpha)))$$

7. From R_4:

$$\mathbf{K}_{self}(a, t, \mathbf{K}_{self}(a, t, happens(action(a, \alpha))))$$

By invoking the process with α related to the first question, that is

$$\alpha = happens(action(a, grasp(o), t_1),$$

the message becomes $\odot = \{a, grasp, o\}$. The final conclusion of an iteration of the reasoning process is the following: the robot talks to itself upon t_1, and it becomes aware of itself, of the object o, and that it is performing the action *grasp*. Moreover, it knows that it knows it. All these knowledge are not considered under t_1.

If in the meanwhile the agent focuses on the new proposition

$$(remember(grasp(o), t_1))$$

the reasoning process iterates by starting from the new message $\odot= \{$*remember, grasp, o*$\}$ and new fluents extent the agent awareness and self-awareness set.

The proposed method is general-purpose, and allows to reason on inner speech and self-awareness generically: it can be reused in any context that requires reasoning by self-talking, not only for self-reflection.

7 Conclusions

In this paper, a preliminary version of a sorted-calculus to automate inner speech and its links to awareness and self-awareness is proposed. The idea is to consider the robot awareness and self-awareness as the set of fluents the robot knows as active at a time. The calculus allows to reason by natural deduction, and extents the robot general knowledge by including fluents representing internal and external conditions. The inner speech allows to activates such fluents. A simple scenario is described for demonstrating how the calculus works. Many other aspects have to be considered: the way the inner queries emerge, and the ability to formulate answers are fundamentals. The processes underling the composition of queries and answers may be automated by the same calculus by adding further axioms, functions and sorts, or by considering models of question-answering systems. The work opens new challenges and research scenario not yet investigated for robot awareness and self-awareness.

References

1. Fernando Martinez-Manrique, Agustin Vicente The activity view of inner speech Front Psychol. 2015; 6: 232. Published online 2015 Mar 9. doi: 10.3389/fpsyg.2015.00232
2. Clowes, Robert. (2007). A Self-Regulation Model of Inner Speech and Its Role in the Organisation Of Human Conscious Experience. Journal of Consciousness Studies. 14. 59-71.
3. Mirolli, Marco (2011). Towards a Vygotskyan Cognitive Robotics: The Role of Language as a Cognitive Tool. New Ideas in Psychology 29:298-311.
4. Morin, A. (2005). Possible Links Between Self-Awareness and Inner Speech: Theoretical Background, Underlying Mechanism, and Empirical Evidence. Journal of Consciousness Studies, 12(4-5), 115-134.
5. Pipitone, A.; Lanza, F.; Seidita, V.; Chella, A. (2019). Inner speech for a self-conscious robot. In CEUR Workshop Proceedings. CEUR-WS.
6. Chella A.; Pipitone A. (2019) The inner speech of the IDyOT: Comment on "Creativity, information, and consciousness: The information dynamics of thinking" by Geraint A. Wiggins. Phys Life Rev.
7. Govindarajulu N.S.; Bringsjord S. (2017). On Automating the Doctrine of Double Effect. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. (IJCAI-17)
8. Shanahan, M. (2000). The Event Calculus Explained. In Artificial Intelligence LNAI. 1600. 10.1007/3-540-48317-9_17
9. McIntyre, A. Relevance Logic. In Edward Zalta, editor, The Standford Encyclopedia of Philosophy. September, 2014 edition, 2014.
10. S. Bringsjord, N. Sundar G., D. Thero and M. Si, ‘Akratic robots and the computational logic thereof’ 2014 IEEE International Symposium on Ethics in Science, Technology and Engineering, Chicago, IL, 2014, pp. 1-8. doi: 10.1109/ETHICS.2014.6893436
11. Morin, Alain (1995). Characteristics of an effective internal dialogue in the acquisition of self-information. Imagination, Cognition and Personality 15 (1):45-58.

12. Kendall, P. C., and Hollon, S. D. (1981). Assessing self-referent speech: Methods in the measurement of self-statements. In P. C. Kendall and S. D. Hollon (Eds.),Assessment strategies for cognitive-behavioral interventions. New York: Academic Press.
13. Arkoudas, K.; Bringsjord, S. (2008a), Toward Formalizing Common-Sense Psychology: An Analysis of the False-Belief Task, in T.-B. Ho & Z.-H. Zhou, eds, 'Proceedings of the Tenth Pacific Rim International Conference on Artificial Intelligence (PRICAI 2008)', number 5351 in 'Lecture Notes in Artificial Intelligence (LNAI)', Springer-Verlag, pp. 1729.
14. Steels, L.: Language Re-Entrance and the 'Inner Voice.' Journal of Consciousness Studies 10(4-5), 173185 (2003)
15. Bringsjord, S; Govindarajulu, N.S. (2012) Given the Web, What is Intelligence, Really? Metaphilosophy, 43(4):361532.
16. Gentzen, G. (1969) Investigations into Logical Deduction in M. E. Szabo (ed.), The Collected Works of Gerhard Gentzen, Amsterdam.
17. Steels, Luc. Basics of fluid construction grammar Journal Article Constructions and frames, 9 (2), pp. 178-255, 2017, ISBN: 1876-1933.