Journal of Artificial Intelligence and Consciousness

Consciousness as a Trigger to Adaptation

Massimo Cossentino
National Research Council of Italy (CNR), Via Ugo La Malfa, 253
Palermo, 90146, Italy
massimo.cossentino@icar.cnr.it

Luca Sabatucci
National Research Council of Italy (CNR), Via Ugo La Malfa, 253
Palermo, 90146, Italy
luca.sabatucci@icar.cnr.it

John Mylopoulos
University of Toronto
40 St George St, Toronto, ON M5S 2E4, Canada
jm@cs.toronto.edu

Received 22 November 2022
Accepted 8 December 2022
Published 30 January 2023

Consciousness refers to mental states of a cognitive agent that make it aware of elements of its environment and its own state.Phenomena of consciousness have been studied in Philosophy, Psychology and Cognitive Science. With the advent of Artificial Intelligence (AI) the study of consciousness has broadened to include questions such as "When is an AI system conscious?", "What role does consciousness play in the cognitive architecture of an AI system?". This paper proposes a cognitive architecture that attempts to answer such questions and is an extension of proposals for practical reasoning, founded on the notion of epistemic goals, goals where the desired state-of-affairs is knowledge of something in the environment or within the agent itself. The proposal is illustrated with an example involving an autonomous vehicle.

Keywords: Consciousness; Self-Consciousness; Goal; Goal-Oriented; Practical Reasoning; Agent.

1. Introduction

The first occurrence of the term "artificial consciousness" according to Chella [2022] appeared in the book *Cybernetic Machines* by T. N. Nemes, published in Hungary in 1962, and later translated into English in 1970. Several works from that time can be found in the Robotics literature, e.g., Igor Aleksander (1992, 2015), Holland (2003a), Clowes et al. (2007), Chella and Manzotti (2007b), and Chella et al. (2019), also the Neuroscience, e.g., Rees et al. 2002; Tononi and Koch 2008; Koch et al.

(2016). Despite the great bibliography on the matter, the definition of artificial consciousness is still an open challenge. In this paper, we adopt a perspective inspired by research on Requirements Engineering (RE), e.g. Dardenne *et al.* [1993]; Chung *et al.* [2000]; Giorgini *et al.* [2002]; Sebastiani *et al.* [2004]; Nguyen *et al.* [2018], and founded on an agent- and goal-oriented paradigms, hoping to contribute to the debate.

Many definitions exist of agents, some addressing the intelligent/ autonomous features of a virtual entity, others referring to the embodiment of a software system, or the collective (swarm) behavior of some software system. In this manuscript we will address the agents as software systems that exhibit properties such as autonomy, social ability and pro-activity, among others (Wooldridge & Jennings [1995]). The great benefit of conceiving a software system as an agent of this kind is the level of abstraction in terms of which the system is conceived and designed (Newell [1982]), where beliefs replace variables and machine states, plans substitute for routines and procedures, and goals are attitudes over states-of-affairs that drive the agent's behavior and define software requirements.

Many attempts exist in the literature to provide agents with some kind of intelligent/rational reasoning method. Likely, the most known one is Practical Reasoning (PR) concerned with the use of reasoning by an agent to determine how to act. PR has had a great impact on many research disciplines (Philosophy, Artificial Intelligence (AI), Robotics,. . .) from its initial formulation by Bratman [1979, 1987]. An interesting perspective on PR has been provided by Hughes [2009]: "Engineering is to Science as PR is to theoretical reasoning": where the ultimate aim of Science (and theoretical reasoning) is to build evidence-based beliefs, whereas the goal of Engineering (and PR) is to determine what an agent ought to do next to satisfy its needs and desires represented as goals on its agenda.

In the last 30 years, researchers put effort on how agents use PR to select which goals they should pursue, and to choose a plan (or even a single action) from a set of available alternatives the agent knows are available in its given circumstances. When the agent selects the best or most practical means of achieving the goal in the given situation, it tries to go ahead and carry out this plan. In goal-based practical reasoning if an agent finds that the state of the world or its goals have changed or a different (and better) plan is now available, it can change its previous deliberation and adopt a new strategy (see Walton [2015]). In Bratman's theory, plans have some interesting features, they are stable, incomplete, and may have sub-plans (see also Hughes [2009]). In Computer Science, PR is often referred to as means-ends reasoning, or goal-directed reasoning (Russell [2010]), or means-ends analysis (Simon [2019]).

We consider consciousness as a fundamental cognitive function associated with PR: an agent cannot reason about what to do next without taking into account the status of the environment and that of your current goals and mental/bodily state. We adopt a framework of consciousness proposed in **?**Souza *et al.* [2012] for self-adaptive systems, which, in turn, adopts ideas from Rosenthal [2019]. We consider an agent conscious when it is aware of its environment, of its own state, and it has the capability to report how it got these beliefs. We also argue a relevant role in agent's reasoning is played by self-consciousness that critically evaluates conscious decisions applying an introspective perspective.

Finally, we propose some considerations about the impact that the proposed theoretical framework would have on the classical PR architecture proposed by Bratman.

The paper is organized as follows: Sect. 2 defines our concept of goal, and introduces a formalization grammar for that. Sect. 3 starts defining awareness, and then introduces consciousness also deducing some interesting features an agent could exploit as a consequence of being conscious. Sect. 4 defines self-consciousness and some interesting consequences for an agent that supports that. Finally, Sect. 5 considers the classical Bratmans's architecture for PR, and extends that with the support for the features descending from consciousness, and self-consciousness introduced in the previous sections.

2. Goals

A goal is a desired state-of-affairs for an agent. In RE, goals describe requirements for a system-to-be are largely used in requirement engineering, see Yu [1993]; Yu & Mylopoulos [1994], to depict why, what and how a piece of software is developed Bresciani *et al.* [2004].

More recently, goals have been adopted in the research area of self-adaptive systems Sawyer *et al.* [2010] (and in particular by the run-time models community Szvetits & Zdun [2016]). In order for a system to be able to adapt to unknown circumstances, it needs a representation of its goals which is causally connected to its adaptation mechanism, thereby ensuring proactivity, self-adaptation and dependability.

Goals have an associated modality that indicates whether the state-of-affairs is desired, i.e., is to be achieved (A) or denied (D) at some point in the future; alternatively the state-of-affairs may have to be maintained (M) or needs to be prevented (P) for some time period [Dardenne *et al.* [1993]]. For example,

$$\mathrm{g}_0 := \text{“Drive to Trento by tonight”},$$

is an A goal, while

$$\mathrm{g}_1 := \text{“Achieve } \mathrm{g}_0 \text{ safely”},$$

is an M goal.

In our proposal, we also distinguish between Operational, Quality and Epistemic goals. The former are goals that an agent is able to operationalize through its actions. They represent the most classical notion of goal, widely discussed in the literature (see, for instance, Bresciani *et al.* [2004]; Letier & Van Lamsweerde

[2004]; Sawyer *et al.* [2010]; Sebastiani *et al.* [2004]) and closely connected to AI planning Cardoso *et al.* [2018]; Höller *et al.* [2018]. An operational goal describes the final state of affair that and agent wants to address or to maintain through its functions (decision-making, actuators and social interactions). g_0 is an operational goal for an autonomous vehicle (hereafter AV), where the AV performs driving actions towards reaching its destination.

Quality goals impose quality constraints on actions or other goals. They have been largely analyzed and discussed in literature as instruments to depict non-functional requirements (such as usability, maintainability, security and performance) Feng-Lin *et al.* [2014]; Chung *et al.* [2000]; Mylopoulos *et al.* [1992]. g_1 is a quality goal that might be operationalized by imposing a constraint on the driving done until g_0 is achieved that consists in driving at a maximum speed of 70km/hr.

Epistemic goals are goals where the desired state-of-affairs is an epistemic state, such as:

$$g_2 := \text{“There is a rock in the middle of the road”}.$$

Epistemic goals were proposed in Souza *et al.* [2011, 2012] as awareness requirements for implementing a MAPE (monitor-analyze-plan-execute) loop for a self-adaptive system. Epistemic goals are achieved through monitoring an agent's environment for obstacles/opportunities, that may lead to corrective action if an obstacle is encountered or an opportunity arises.

Monitoring for epistemic goals may be accomplished through perceptual functions the agent has, such as a visual system for the AV that can recognize rocks on its path. It can also be accomplished through communication with other agents, such as a driver driving ahead of the AV who issues warnings if it sees anything dangerous on the road.

Goal reasoning with goals such as g_0, g_1, g_2 is only possible if we can formalize them in some sort of logic. For our purposes, we propose to use typed linear temporal first-order logic with bounded quanification (bLTFOL) where quantification is only done with respect to finite domains Dardenne *et al.* [1993]. bLTFOL is inductively defined by the grammar:

$$\begin{aligned} \Phi \equiv \mathrm{p} \ \mid \top \mid \neg\Phi \mid \Phi_1 \vee \Phi_2 \mid \Phi_1 \wedge \Phi_2 \mid \\ \Phi_1 U_I \ \Phi_2, \mid G_I \ \Phi \mid F_I \ \Phi \end{aligned} \tag{2.1}$$

State formulas assume objects exist and change their properties over time; towards this aim, we use first-order predicates and universal quantifiers. Path formulas pose constraints on variability over state evolution. To talk about time, we adopt temporal operators like Until, Finally and Globally. These operators may be defined within intervals indicating the temporal validity of the formula. Since quantification is allowed only over finite domains, bLTFOL formulas may be rewritten as Propostional Logic (PL) formulas, thereby simplifying goal reasoning. We now introduce the formalization according to bLTFOL of the goals proposed above:

$$g_0 := F_{[0,Tonight]}\ drive\_to(AV, Trento)$$
$$g_1 := G_{[0,Tonight]}\ safely(AV)$$
$$g_2 := known(AV, rock\_on\_the\_road)\ U_{[0,Tonight]}\ no\_alternative\_routes$$

As we can see the first goal, g_0, includes a Final condition over the time interval [0,Tonight] for the *drive_to (AV, Trento)* predicate thus meaning that the goal will be satisfied if the AV will drive to Trento before nighttime. The safety constrain on the previous goal represented by goal g_1 is expressed by a Globally condition in the same interval that prescribed the AV always be safe. Formalization of the goal g_2 presents an interesting approach, since it prescribes that the AV should know if there are rocks on the road until there are no alternative routes available (since discovering rocks on the road when the route cannot be changed is useless).

In PR, an agent is acting on several goals concurrently and such goals may be conflicting. In such a setting, the agent must reason on possible tradeoffs, thus arriving to a choice in which goal satisfaction is often partial. Dealing with these situations is a central problem for self-adaptive systems acting on the real world. To the purpose, measuring partial goal satisfaction Souza *et al.* [2011]; Angelopoulos *et al.* [2014] allows for an agent to overcome the classical all-or-nothing approach to goal satisfaction Zhou *et al.* [2008]. For example, g_0 and g_1 may be conflicting and therefore cannot both be fulfilled. Measuring partial fulfillment is a recent open issue that may be faced in different ways.

Probabilistic analysis is one of the possible approaches Horkoff & Yu [2013]; Dell'Anna *et al.* [2019]; Giorgini *et al.* [2002]. The key idea is to adopt an approach where the evidence of satisfiability is represented as the a posteriori probability that a goal is satisfied. For example, fulfilling g_0 if we fulfill g_1, e.g., 1 out of 10 drivers can fulfill g_0 if they drive at a speed that is $\leq$ 70 Km/hr, also 2 out of 10 drivers can respect the speed limit while fulfilling g_0.

Another approach to measure partial satisfaction is to measure how much more effort is needed to fulfill a goal Van Riemsdijk & Yorke-Smith [2010]; Lahijanian *et al.* [2015]. For instance, the effort to satisfaction of g_0 may be measured in terms of extra hours driven, if g_1 is fulfilled, or extra gasoline needed if g_0 is fulfilled.

Another interesting research trend is the goal change technique Van Riemsdijk & Yorke-Smith [2010] consisting in the pursuit of runtime modification of a goal Lesser *et al.* [2004]; Shapiro *et al.* [2005]; Zhou *et al.* [2008]) that cannot be fully satisfied. For instance, Jureta Jureta *et al.* [2014] proposes relaxation of requirements by weakening the original state when it cannot be achieved. Another approach is to reformulate failed goals into new problems that can be solved using a planner Vukovic & Robinson [2005]. Goal adaptation requires the agent owns a deep understanding of the domain and of which impact goal achievement has on that.

The agent ability to know and reason about the progress towards achieving their own goals (sometimes called progress appraisal Van Riemsdijk & Yorke-Smith [2010]; Feltovich *et al.* [2007]) can constitute the basis for building the self-awareness property.

3. Consciousness as a Collection of Epistemic Goals

In this section, we introduce our concept of consciousness. Briefly, we will start from the definition of awareness as a state of mind of an agent who wants to know things about its goals. Starting from considerations of Rosenthal (Rosenthal [2019]) and Smithies (Smithies [2019]), we then introduce consciousness as awareness with reportability where the agent can document how it came to this awareness.

The first step of our process consists in a definition of what we address as the entity having awareness and consciousness. The weak-agent concept, as presented by Jennings and Wooldrige's in their seminal article Wooldridge & Jennings [1995], seems like a good starting point. A weak agent has properties like: autonomy, proactivity, sociality. Our agents are weak agents with two additional properties: (a) Rationality that allows an agent to choose its actions in order to satisfy its goals and has a PR mechanism to choose its next action; (b) Sensoriality that allows an agent to interact with its physical and cyber environment. Rationality allows an agent to reason about its own goals (operational, quality and epistemic) and its next actions. Sensoriality complements sociality and they both allow an agent to interact with its environment to support the satisfaction of its goals.

In the following, when referring to a generic agent, or just 'agent' when there is no ambiguity, we are referring to weak agents with the rationality and sensoriality property.

3.1. *Awareness*

Awareness is the first agent property we will discuss. We consider awareness an agent's state-of-mind about goals on its agenda:

Awareness of a goal on an agent's agenda is a state-of-mind where the agent knows something about the satisfaction of that goal.
Awareness requires the capability of an agent to fulfill epistemic goals.

This means the agent knows something about the world and/or its own state-of-mind. For instance, the AV is aware there is an obstacle on the road, or it is aware that goal g_2 has been satisfied three times in the past ten minutes.

It is important to emphasize here that awareness may concern 'inner knowledge' whose origin is lost in the past, such as knowing that a rock was encountered on the AV's path three times in ten minutes in a previous travel long time ago. Note that we are not remarking any information about the origin of this news (who found the rocks in a previous travel? When? Where?). In fact, awareness doesn't require reportability, the agent does not know the origins of its beliefs. This means that the agent cannot answer "why do you believe" questions.

For instance, an agent is aware that the weather is rainy, but it cannot justify that referring to a sensing act (specific and positioned in time), or a communication from another (trusted) agent stating that it is raining. This may be, for instance, the consequence of an implementation platform that does not relate the agent's knowledge to the sensing acts performed in the world.

As indicated earlier, an agent may acquire its awareness about the satisfaction of its goals by using its own perception capabilities, but also by means of an interaction with other agents (as a consequence of its social propensity). It is beyond the scope of this paper to discuss the case when other agents are malicious or not and, therefore, whether the information they communicate should be trusted or not. This depends on the specific application context and implementation architecture, and does not influence our proposal.

Besides, the concept of goal satisfaction we refer to, is not only addressing the full satisfaction of some goal (the proposition formalizing the goal holds true), but also a partial satisfaction condition addressed in every possible way described in literature (see, for instance, Van Riemsdijk & Yorke-Smith [2010]; Letier & Van Lamsweerde [2004]; Cossentino *et al.* [2018]).

In the next subsection we introduce the definition of consciousness that we base on awareness with the additional property of reportability.

3.2. *From Awareness to Consciousness*

As we discussed in the previous subsection, our concept of awareness is based on some kind of knowledge, as Williamson [2000] says: "knowing is a mental state", and consequently we could groundedly regard awareness as a mental state. On the concept of awareness we also base our definition of consciousness that is:

Consciousness of a goal on an agent's agenda is a state-of-mind where the agent knows something about the satisfaction of that goal. Like awareness, consciousness requires the capability of an agent to fulfill epistemic goals. Unlike awareness, however, consciousness also requires the agent to know the origins of its knowledge about the satisfaction of its epistemic goals.

In other words, consciousness is traceable awareness about an agent's epistemic goals. We now detail how we arrive to conceive this formulation. Making a step towards consciousness requires further concepts to be added to the mere awareness. A relevant research on the matter is the *mentalism* proposed in Smithies [2019]. In this work, Smithies states that "epistemic justification depends solely on your mental states, rather than the reliability of their connections to the external world". This further consolidates our concept of awareness as an individual belief of the agent about its goals that has no reliable connection to the external perceived world. Indeed, we here enter the limits of the awareness concept: this connection is of paramount importance when the belief has to be relied upon for strategically relevant choices.

Let us consider an example: an AV is 'aware' that the road ahead is closed because of a landslide, and it decides to change its route; this is a correct choice if the slide is still in place and forbids safe passage. But, conversely, it will negatively affect the quality of the AV performance if the rock slide has been removed and the road is now passable. Consequences may have an effect on quality goals, such as taking too much time, or consuming too much fuel for reaching the destination. This may happen, for instance, if awareness of the rock slide comes from an old belief that is not marked with any timestamp, and the agent does not evaluate it as likely to be old.

Starting from these considerations, and going towards the definition of consciousness, we consider the relevance of relating what the agent knows about the status of its epistemic goals (agent's awareness) with the way the agent achieved its beliefs. On the matter, Smithies [2019] combines two main elements in his theory of epistemic justification, which he calls *Phenomenal Accessibilism*, they are:

- The epistemic justification of a mental state comes from possibly independent facts about the agent's mental states;
- The epistemic justification has to be accessible, in the sense that the agent knows which beliefs he has epistemic justification to hold.

A similar concept is introduced in Rosenthal [2019], when the author says that "there is a crucial connection between reportability and consciousness". This creates an immediate link with the epistemic justification of a mental state underlined by Smithies. Rosenthal makes another interesting consideration about the importance of awareness, when he states that if an individual is not aware of her psychological state, then the unique explanation for that is that the individual is not conscious of this state. In some way we may see awareness as a 'necessary' condition for consciousness, but it (awareness) does not require reportability. Consequently, we think that the passage from awareness to consciousness is in the agent capability to provide an epistemic justification of its mental state.

Suppose, again, that our AV is driving along the road. It makes sense to imagine that it has checked weather conditions before deciding between alternative roads. For instance, to select between a shorter path through a mountain pass (where it may be snowing) or a longer route going around the same mountain at a lower altitude.

In case of a long travel, it is reasonable to expect that the agent does that at the beginning of the travel, early in the morning, but it also looks for an updated weather report after a few hours.

Let us detail a little bit the mental process we imagine happening in this agent: the agent has an operational goal (arrive to Trento before nighttime), and some quality goals constraining that (make a safe trip, save fuel,...). The plan used to pursue the goal g_0, also includes the evaluation of weather conditions for selecting the best route. About that, the agent has a relevant belief (*weather_is_snowy*) holding if there is significant snow along some route. If the belief holds, the route is to be considered dangerous (it clashes with the quality goal about making a safe trip) and this motivates to relax the fulfillment of the goal about saving fuel and taking a longer (safer) route. Incidentally, we here note that we are explicitly introducing the consideration of partial goal satisfaction in the agent's reasoning process.

Now, the point is that the agent needs to know if it is snowing in order to pursue its operational goals. Consequently, the same existence of the belief *weather_is_snowy*, creates a new (epistemic) goal on the agent's agenda: acquiring knowledge whether it is snowing along the route.

In this case, we say that the agent will be conscious whether it is snowy if it knows that it is (or it is not) snowing, and it also can relate this belief with the source of information (for instance some weather bulletin) and the time of acquisition of the belief. This linking provides the epistemic goal with a traceable property that becomes of paramount importance when the information is used in a mission-critical decision. Moreover, we can expect that, according to the specific epistemic goal, the agent will have a different concern in pursuing the goal satisfaction (that is achieving/updating its knowledge), thus giving place to epistemic goals that can have different priority with regard to the operational goals they are connected to, as well as epistemic goals that have a kind of 'refresh time' urgency.

For instance, let us suppose the AV is provided with many sensors useful for improving safety, among them a rain sensor. This agent may know if the weather is raining by accessing the readings of that rain sensor; in this way, the agent may adjust travelling speed to road surface conditions. Therefore, if the sensor is available, the agent may deliberate to fulfil the epistemic goal to discover the state of the *weather_is_rainy* belief.

Indeed, there is another interesting consideration about the deliberation phase the agent performs: Bratman [1987] states that it is not rational for an agent to bring about some state of affairs, while at the same time having the belief that this state of affairs cannot be achieved.

Therefore, acting under the Principle of Rationality as proposed by Newell [1982]; Wooldridge [2001], no agent will commit to pursue an epistemic goal that it knows cannot be achieved. For instance, resuming the example about rain belief, the agent will not even attempt to activate the rain sensor if it knows that its hardware memory is full, and no other piece of information can be written in it, or that activating the rain sensor requires energy from the driving lights (and the car is electric, running in night conditions with a low battery level). Again we are dealing with a tradeoff situation that may require the agent to accept partial goal satisfaction according to environment conditions or the consciousness of its own state (self-consciousness, see next section).

As already said, consciousness consists of a set of epistemic goals the agent's reasons about. These goals deal with the desired states of affairs the agents wants to achieve, and therefore it wants to know something about them. It is worth noting that when saying 'epistemic goals' we are indeed referring to a specific subset of such goals. Consciousness as we defined above is not concerned with the acquisition of whatever belief the agent may meet in its life, but, rather, in the beliefs related to its operational goals, in terms of their degree of fulfillment as well as all the factors (opportunities, obstacles) that could influence that.

We should also consider that being conscious requires the agent has some additional capabilities: it should be able to add to its agenda epistemic goals related to its operational goals, and reason about how to fulfil them. Fulfilling an epistemic goal means acquiring some information that could make true or false some belief, this implies the agent will use it perception or communication capabilities to investigate the status of the world and achieve the desired state of mind. The accomplishment of epistemic goals has a direct impact on the agent's decisions about which goal to pursue and the way to do that.

We can, now, wonder what is the rationale for providing agents with consciousness. For our proposal, the rationale is that consciousness enhances an agent's practical reasoning capabilities by improving its effectiveness and giving the agent with more chances to catch its goals by avoiding risks and tacking opportunities.

In this direction, consciousness plays an essential role in the agent's life, it becomes the inner mechanism that motivates the agent to pursue its (operational) goals, and when these goals cannot be achieved, consciousness triggers an adaptation mechanism that leads to goals and plans being discarded, weakened/strengthned or new goals plans added to an agent's agenda.

Consciousness triggers adaptation.

Many reasons exist that can impede the planned achievement of some goal: some depend on exogenous factor, like changes occurring in the environment because of uncontrollable events, others emerge from the agent itself, for instance the failing of some of its own capabilities in providing the expected result or a wrong plan. In such a case, consciousness triggers a search for alternatives. We may consider two main ways to do that: adopting another plan for achieving the same goal or modifying that. If alternative plans are available the obvious choice consists in selecting the best among them but this may not be the case so consciousness of that forces the agent towards a more complex activity:

Consciousness is a mechanism for enabling the revision of goals.

Goal revision is a very sensitive task, it may consists in different approaches, for instance the agent may choose to aim at only achieving a partial satisfaction of its goals (Letier & Van Lamsweerde [2004]; Van Riemsdijk & Yorke-Smith [2010]; Cossentino *et al.* [2018]), alternatively it can decide to modify them in order to remove the conditions that made them unachievable. The correct and detailed implementation of this goal revision process is still an open issue in literature, despite its great importance. After all, no designer would like to run an agent that, at the first difficulty in executing its plan, aborts its mission.

Above we implied that the agent was forbidden to achieve its goal, and would need to start some adaptation process because of the occurrence of some obstacles that invalidated the current plan.

Indeed, there is another totally different, yet equally important motivation that is the emergence of some opportunity. Suppose the AV plans some route according to the current information, that the better alternative one is closed because of an accident. Now, let us suppose that the news arrives that the route has reopened. An unconscious agent would not consider this news (unless it dumbly repeats the route planning process from time to time). Conversely, a conscious agent should immediately notice that this news changes the belief that motivated the adoption of the current route and therefore a new planning phase is to be triggered.

We can, therefore, summarize that consciousness allows identifying obstacles to be removed and opportunities that should be exploited to fulfill the agent goal.

The role of obstacles and opportunities in agents' reasoning has been initially discussed by Van Lamsweerde & Letier [2000]. According to the authors, there are three categories of obstacles:

- Non-satisfaction obstacles are obstacles that obstruct the satisfaction of agent requests;
- Non-information obstacles are obstacles that obstruct the generic goal of making agents informed about object states;
- Inaccuracy obstacles are obstacles that obstruct the consistency between the state of objects in the environment and the state of their representation in the software.

An agent, while pursuing its goals with some plan, should naturally try to take advantage of the opportunities raising around while, at the same time, it should avoid obstacles. This deeply impacts the agent's reasoning process: an agent that is conscious (i.e. it has traceable awareness of its epistemic goals) will consider obstacles and opportunities in its deliberation and planning activities. This is a survival need, since it allows a less risky reasoning about what goal to pursue and how to do that.

Agents, just like living beings, should be provided with a survival instinct, and we think consciousness is the realization of this instinct, we may say that:

Consciousness is a mechanism for reducing risks and taking advantage of opportunities.

Nowadays, a popular research topic in artificial intelligence (AI) is explainability. Since many years, AI applications have been working in ways that could not be exactly explained. Results were correct, but the way the system arrived to them was not traceable. A notable example of that is the use of neural networks in many of their applications, for instance classification or image recognition.

In this sense:

Consciousness contributes to explainability.

If an agent is conscious of an epistemic goal, it can report why (and when) the corresponding belief had been acquired and the goal satisfied. This belief will likely be used by the agent to perform some decision; having a clear traceability (we could use the word ‘reportability’ as done by Rosenthal in the previously cited paper) makes the use of consciously managed goals a contribution to the explainability of the agent behavior.

4. Self-Consciousness as Epistemic Goals about other Goals

In this section we introduce self-consciousness building on the concepts introduced on previous sections.

Self-consciousness for an agent is as a kind of second order consciousness: consciousness upon consciousness. A similar position (also about high-order awareness) has been already discussed by other authors, for instance, Rosenthal [2012]. Accepting this interpretation of the selfishness in consciousness, we may say that:

A self-conscious agent is aware (in a traceable way) of something about its own conscious state of mind (i.e. its beliefs plus their reportability).

This means the agent has some reportable belief about beliefs dealing with the status of an epistemic goal, or it has some reportable belief about its reportability.

Fig. 1 shows an example of that: suppose our AV has the operational goal g_0 (“Drive to Trento by tonight”), this is constrained by the quality goal g_1 that requires goal g_0 be fulfilled safely. In order to successfully accomplish these goals, the agent introduces an epistemic goal g_2:

$$g_2 := \text{“There are rocks along the road”}.$$

The outcome of this goal feeds the belief *Rocks_on_the_road*. If this belief is true, the agent has found an obstacle to the achievement of goal g_0, and this will trigger its adaptation mechanism.

The belief *Rocks_on_the_roa*d is complemented by some other belief about the reportability of this state of mind; this may include, for instance, information about the sensor device that has perceived the information, and the time when that had happened. In Fig. 1 we see the agent has got the information from some camera at a specific time. This constitutes the conscious state of mind of the agent: it is aware that rocks have been found on its route, and this information originates from its front camera at a specific time.

Soon after, the agent could develop a self-consciousness state of mind about its belief *Rocks_on_the_road*, which is now deemed unreliable because the camera used has been found to be defective just seconds before it perceived rocks on the road.

This self-conscious state of mind acts as a supervisory level upon the conscious one. Starting some adaptation mechanism for achieving goal g_0 (as it was suggested by the conscious belief that there are rocks on the road) perhaps is not the right choice. The agent would prefer to verify the reliability of the belief *Rocks_on_the_road* before acting differently. This may involve starting some testing procedure for the camera and will look for alternative information on traffic conditions in the area. This information may confirm or confute the current conscious state of mind, and

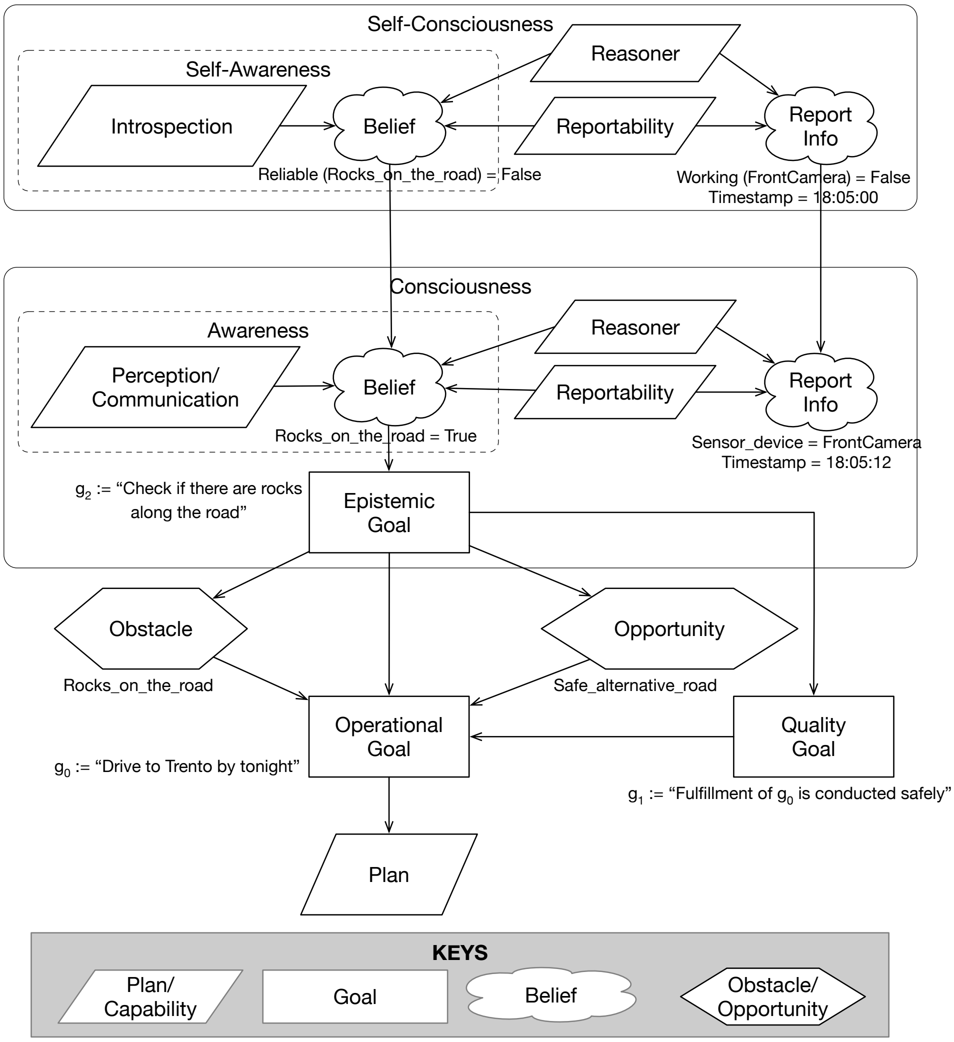

Fig. 1. Consciousness vs Self-Consciousness.

resolve the doubt posed by the self-conscious consideration that perhaps the camera was not properly working when it perceived rocks on the road. Therefore, we may declare a significant property of self-consciousness:

Self-consciousness triggers the override of consciously taken decisions

In the next section we discuss how this latter property and others included in our proposal relate to the practical reasoning architecture presented in Bratman *et al.* [1988].

5. Consciousness, Self-Consciousness and Practical Reasoning

In the previous sections we proposed consciousness and self-consciousness definitions that are inspired by a goal-oriented perspective, they refer to agent's operational and epistemic goals. In this section, we relate these concepts to the most classical practical reasoning architecture. More specifically, we discuss their link to practical reasoning Bratman [1979], and propose some future research directions.

In the 80s, Bratman proposed the belief-desire-intention model (BDI) (see Bratman *et al.* [1988]) bridging Philosophy of Mind and AI. The work was seminal and numerous approaches adopted his theories either from a purely formal perspective or from an implementation perspective. The original Bratman's theory is qualified as a planning theory of intention, where the role of intentions is to intermediate between goals, plans and actions.

Indeed, before going into details, it may be useful to disambiguate the meaning of some terms addressing the idea of agent/system objective as they are used in the respective communities of requirements engineering and artificial intelligence (practical reasoning), namely: operational/epistemic/quality goal vs belief, desire, intention. As a general consideration, we can say that they come from different perspectives: the requirements engineering (RE) community is concerned about the specification of the requirements of some system and therefore considers goals as something that is 'imposed from outside' to the system. Usually, the analyst/designer builds a partially ordered goal graph with stakeholder requirements as roots, intermediate goals derived through refinements from root goals, and actions as leafs of the graph that define the main components of a system that satisfies stakeholder requirements. This is an imperative approach to system design in that it is an external agent, the designer, who decides how stakeholder goals are fulfilled and this is done at design- rather than run-time. The system designed here is not even a weak agent because it doesn't have autonomy. In AI, specifically Multi-Agent Systems and PR, an agent agrees to pursue some objectives which are added to its list of desires, but being autonomous, it is free to decide when it is ready to actively pursue each desire by promoting it to an intention. We can therefore say that in RE, goals are externally provided and design amounts to choosing how they will be fulfilled. In AI, on the other hand, desires and intentions concern agent mental attitudes: an agent 'desires' to achieve some state of the world or 'commits' to achieve that state. This implies an internal perspective on the specification of the agent's behavior. Trying to compare the scope of these different concepts, on the left side of in Fig. 2 we represent the three types of goals: operational, epistemic, quality. Their definitions may not be mapped one-to-one to any of the practical reasoning main concepts (belief, desire, intention), rather, we may identify some coverage areas that help in illustrating their correspondence.

An operational goal describes a desired state-of-the-world and therefore covers both desires and intentions. Similarly, an epistemic goal requires a system to act to acquire the information needed to determine the validity of some belief. In this sense, an epistemic goal has a relationship with the belief concept but at the same time it is a goal and therefore it covers the scope of a desire and an intention. Finally, quality goals impose constraints on other goals, operational, quality or epistemic.

We can now proceed to the illustration of the impact that the proposed approach to consciousness and self-consciousness has on a practical reasoning architecture.

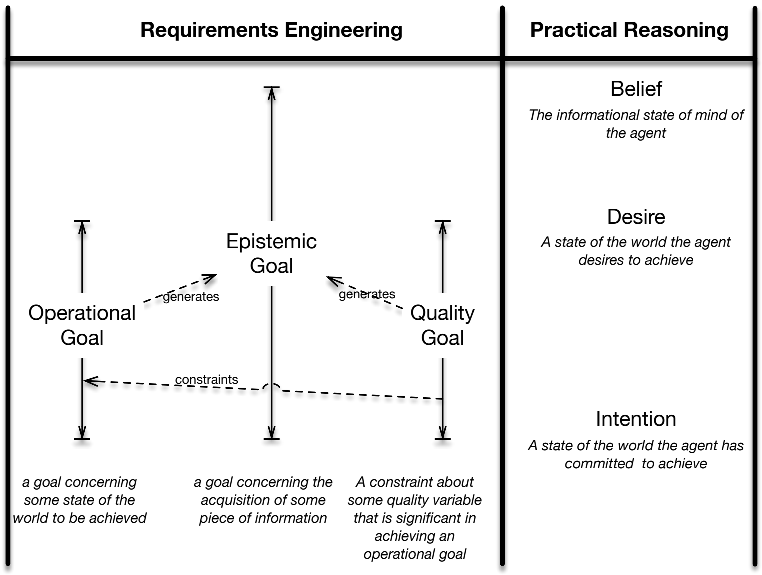

Fig. 2. An informal comparison of significant concepts in requirements engineering and artificial intelligence (practical reasoning).

Fig. 3 revises the classical abstract architecture proposed in Bratman *et al.* [1988] for practical reasoning with the aim to focus on consciousness and self-consciosness properties.

Whereas the essential nature of the original architecture is maintained, Fig. 3 introduces two new processes (namely, introspection and reportability mechanisms) and it also highlights the various functions.

The architecture is clearly grounded on beliefs, desires, intentions and plans (executable refinements of intentions). Ovals in the figure are information stores whereas processes are denoted by rectangles. A specific notation with compartments allows us to cluster each process function in the four areas (awareness, self-awareness, consciousness and self-consciousness). For the sake of clarity we omitted to put relationship from belief to all the processes (as it was in the original figure). It is understood that all these relationships must exist given the very central nature of the belief element.

The ***Means-End Reasoner*** has the responsibility to refine agent's intentions

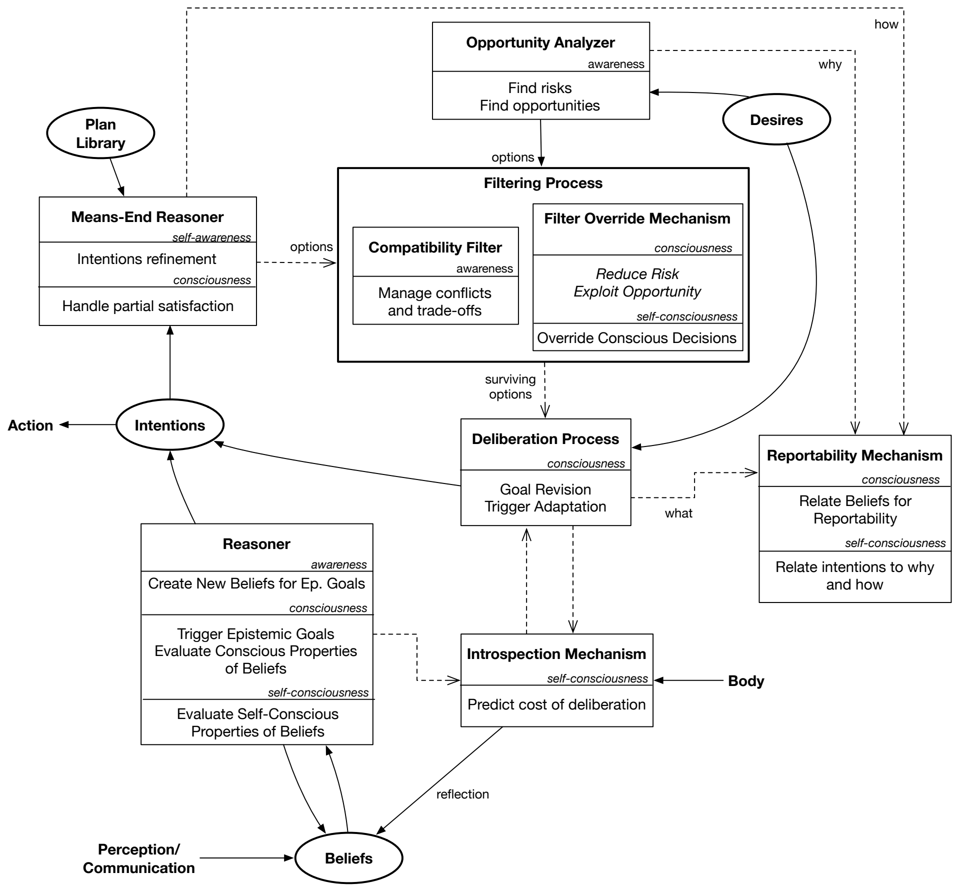

Fig. 3. Bratman's practical reasoning architecture annotated with significant consciousness and self-consciousness properties. It is worth noting that, even if not explicitly represented, the belief element is connected to all the processes of the figure.

into lower-level intention until arriving to atomic plans. These become options for the deliberation process that turn them into intentions. This capability to reason on an a particular end is a controversial property in the Bratman's architecture. It was considered of paramount importance by the author but, so far, all the mainstream BDI frameworks (Bordini *et al.* [2007]; Dastani *et al.* [2005]) remain silent about this concept (Herzig *et al.* [2017]). One of the reason is that decomposing an intention into sub-intentions requires the agent owns a structured representation of its intentions, that is founded on self-awareness, i.e. beliefs about its desires and intentions. An additional ingredient that may enrich the architecture is the capability to deal with plans that do not ensure the full satisfaction of the agent's goals. Relying on self-adaptation could help to this purpose. Indeed, measuring partial goal satisfaction is one of the current challenges of self-adaptive systems Souza *et al.*

[2011]; Angelopoulos *et al.* [2014]

Not all the options for the deliberation process come from a means-end reasoning. Some of them could derive from changes in the agent's environment. The ***Opportunity Analizer*** was a great intuition by Bratman, who conceived it as an active component that encapsulates agent's autonomy. The Opportunity Analyzer explores the current state of the world to evaluate new opportunities for the agent as well as possible obstacles. Opportunities represent situations in which the agent may improve its performances, whereas obstacles are undesired situations that may prevent goal fulfillment. The component exploits agent's awareness to provide options (i.e. proposal for new intentions) to the filtering process.

The ***Compatibility Filter*** has the objective to narrow the scope of deliberation by filtering options (from the means-end reasoner or from the opportunity analyzer) on the base of their consistence with the already adopted plans.

In opposition with the principle that a plan is adopted only if the agent is sure to be able to entirely achieve it, we retain this component should also be able to handle with partial goal satisfaction. This would allow for an agent to overcome the all-or-nothing approach to goal satisfaction and goal conflict Zhou *et al.* [2008]. This is a frequent situation that occurs when acting on the real world. For instance, the component could consider pursuing the partial satisfaction of some goal as a way for making them feasible, avoiding conflicts, and so on.

The ***Filter Override Mechanism*** embodies the agent's sensitivity to obstacles and opportunities defining when one of the current plans is to be suspended and considered against other options.

Considering risks and evaluating new opportunities is part of the consciousness of an agent. When the current plan becomes incoherent with the changing world, or when a better plan is discovered, this component selects an alternative one to deliver to the deliberation component.

A final consideration about this component regards its relationship with self-consciousness. This is the place where the consequences of self-conscious considerations about the status of the agent will impact on its decisions. Suppose the agent has undertaken some plan of action, but the intermediate states of the world are monitored by a not reliable sensor. Likely, if an opportunity occurs to change the flow of actions the agent will soon take that thus resolving the issue.

The ***Deliberation Process*** is the core of the agent's consciousness. It reasons on options that survived from the filtering process and decides which one will be transformed into intentions. It basically drives agent's commitment towards the various operative goals to be addressed. This component is responsible of triggering the run-time change of the agent's behavior. A simpler formulation may be done as adaptation at the plan level (Salehie & Tahvildari [2009]; Sabatucci *et al.* [2018]). It means that when a plan is considered obsolete for some reasons, the component is responsible to operate a transition towards an alternative plan. Here, a more complex formulation of adaptation could be adopted. In facts, the component may also include an extensive adaptation process that consider the revision of agent's goals rather than agent's plans (Shapiro *et al.* [2005]).

The deliberation process and the means-end reasoning are generally considered high demanding tasks. Adapting the behavior according to risks and opportunities requires time and resources, that may be a problem for a resource bounded agent. The novel component ***Introspection Mechanism*** (it was not present in the original architecture) takes into account self-consciousness of agent's internal processes. In Bratman *et al.* [1988], the author discusses if and when the agent should perform an extra deliberation process according to the cost in computational terms. This kind of component is devoted to analyze and direct the computational effort according to the kind of expected result. For this purpose, the component is directly connected to the agent's embodiment (body in figure). Available approaches may include the consideration of the agent's laziness as a way for devoting the right effort to the cases that are worth to be treated.

The second novel component of the architecture is the ***Reportability Mechanism***. As already discussed in Sect. 3, consciousness requires a relationship between beliefs and the way the agent knew them. For this reason, we think a 'conscious' reportability mechanism is required in the agent architecture, and it should support this crucial property: beliefs must be connected to the process the agent used to acquire them. The mechanism should to be able to work on beliefs and create relationships between them. Besides, we think another 'self-conscious' level of the reportability mechanism is very relevant, this should be specifically conceived to connect the following three dimensions: a) what – the result of the deliberation process (i.e. the agent's intentions), b) why – the motivation that brought to the adoption of the current intention(s) (for instance, opportunities or risks that turned a desire into an intention) and c) how – the way the agent is actively pursuing its active intention(s). This component is responsible to enrich the agent's belief in order to connect these elements and produce all the necessary information to trace knowledge and behavior.

Concluding, whereas beliefs come, as usual, from perception and communications, the proposed architecture underlines the roles of the ***Reasoner*** is not only to make deductions, but also to sustain consciousness and self-consciousness properties. Indeed, we assign to the Reasoner the following functionalities, someone derived from classical logical reasoner, whereas other ones must be specifically devised:

- deduct new beliefs from existing ones;

- beliefs revision;
- check for belief base inconsistencies;
- trigger epistemic goals as consequence of operational goals;
- evaluate properties of existing beliefs related to their reportability (conscious properties), like expiration date/time, ...
- evaluate properties of existing beliefs related to self-conscious considerations (self-conscious properties), for instance weak (or expired) motivations, reliability, trust in decoding/understanding the received information,...

Notably, the first three of these functionalities are already part of all the existing implementation of Bratman's architecture. For instance, the arrival of a new perception may trigger the creation of a new belief as well as the update of an existing one; each time the belief-base is updated, a check is mandatory to ensure its consistency.

Moreover, as discussed before, an agent should be able to trigger new epistemic goals (intentions in the proposed architecture) for evaluating the achievement of its operational goal. This means that new beliefs will be generated, and the agent commits to acquire (or update) them. This contributes to improve the chances that the means-end reasoner finds good options for pursuing the agent's desires, and provides the deliberation process with more updated information about the results of the agent's actions in the world that can be used to improve the performance of this process. Reportability also allows the agent to reason about some properties of its own beliefs, for instance their obsolescence, the precaution level to be adopted in its actions (the number of rocks found on the road is increasing in the last few minutes, not really enough to justify the change of the selected route, but that suggests moderating the speed).

Finally, self-consciousness requires even higher-level functionalities: the reasoner needs to be able to evaluate some properties of the beliefs representing the agent's conscious knowledge. For instance, is the motivation that triggered some intention still valid? Is the content of a belief reliable? Issues related with this question include considerations like: may the agent rely on its own sensors for that? Can the agent rely on its decoding/understanding capacity in deducing some information from a sensor, message, etc.?

6. Conclusions

We have sketched a framework for treating agent consciousness as a state-of-mind that an agent arrives at by satisfying its epistemic goals, goals where the desired state-of-affairs is an epistemic state of the form "I believe that X is true". Our framework distinguishes between awareness where the epistemic goal is satisfied through interaction with the agent's environment and consciousness which is reportable awareness. Moreover, self-consciousness is treated as consciousness about the satisfaction status of the agent's goals. However, our framework is just a proposal that needs to be validated through an implementation using an extension of an off-the-self practical reasoner, experimentation and empirical evaluation through case studies.

In particular, PR needs to be extended to perform three tasks: (a) Given a set of operational goals, derive relevant epistemic goals that need to be added to the agent's agenda, (b) Given that current goals can't be satisfied with the currently pursued plan, search for alternative plans to be considered; (c) If there are no alternative plans, search for alternative goals to considered. We propose to address these issues in our future work.

References

Angelopoulos, K., Souza, V. E. S. and Mylopoulos, J. [2014] "Dealing with multiple failures in zanshin: a control-theoretic approach," in *Proceedings of the 9th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, pp. 165–174.

Bordini, R., Hübner, J. and Wooldridge, M. [2007] *Programming multi-agent systems in AgentSpeak using Jason* (Wiley-Interscience).

Bratman, M. [1979] Practical reasoning and weakness of the will, *Nous* **13**(2), 157–171.

Bratman, M. [1987] *Intention, plans, and practical reason* (Harvard University Press).

Bratman, M. E., Israel, D. J. and Pollack, M. E. [1988] Plans and resource-bounded practical reasoning, *Computational intelligence* **4**(3), 349–355.

Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F. and Mylopoulos, J. [2004] Tropos: An agent-oriented software development methodology, *Autonomous Agents and Multi-Agent Systems* **8**(3), 203–236.

Cardoso, E. C. S., Horkoff, J., Sebastiani, R. and Mylopoulos, J. [2018] "Planning with strategic goals," in *2018 IEEE 22nd International Enterprise Distributed Object Computing Conference (EDOC)* (IEEE), pp. 44–50.

Chella, A. [2022] *Cognitive robotics*, chap. Robots and Machine Consciousness (MIT Press).

Chung, L., Nixon, B., Yu, E. and Mylopoulos, J. [2000] *Non-Functional Requirements in Software Engineering* (Kluwer Publishing).

Cossentino, M., Sabatucci, L. and Lopes, S. [2018] "Partial and full goal satisfaction in the musa middleware," in *Multi-Agent Systems. EUMAS 2018. Lecture Notes in Computer Science* (Springer), pp. 15–29.

Dardenne, A., Van Lamsweerde, A. and Fickas, S. [1993] Goal-directed requirements acquisition, *Science of computer programming* **20**(1-2), 3–50.

Dastani, M., Birna Riemsdijk, M. v. and Meyer, J.-J. C. [2005] Programming multi-agent systems in 3apl, in *Multi-agent programming* (Springer), pp. 39–67.

Dell'Anna, D., Dalpiaz, F. and Dastani, M. [2019] Requirements-driven evolution of sociotechnical systems via probabilistic reasoning and hill climbing, *Automated Software Engineering* **26**(3), 513–557.

Feltovich, P. J., Bradshaw, J. M., Clancey, W. J., Johnson, M. and Bunch, L. [2007] "Progress appraisal as a challenging element of coordination in human and machine joint activity," in *International Workshop on Engineering Societies in the Agents World* (Springer), pp. 124–141.

Feng-Lin, L., Horkoff, J., Mylopoulos, J., Borgida, A., Guizzardi, G., Guizzardi, R. and Lin, L. [2014] "Non-functional requirements as qualities, with a spice of ontology," in *22nd IEEE International Requirements Engineering Conference (RE'14)*.

Giorgini, P., Mylopoulos, J., Nicchiarelli, E. and Sebastiani, R. [2002] "Reasoning with goal models," in *International Conference on Conceptual Modeling* (Springer), pp. 167–181.

Herzig, A., Lorini, E., Perrussel, L. and Xiao, Z. [2017] Bdi logics for bdi architectures: old problems, new perspectives, *KI-Künstliche Intelligenz* **31**(1), 73–83.

Höller, D., Behnke, G., Bercher, P. and Biundo, S. [2018] "Plan and goal recognition as htn planning," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)* (IEEE), pp. 466–473.

Horkoff, J. and Yu, E. [2013] Comparison and evaluation of goal-oriented satisfaction analysis techniques, *Requirements Engineering* **18**(3), 199–222.

Hughes, J. [2009] in A. Meijers (ed.), Practical reasoning and engineering, , *Philosophy of Technology and Engineering Sciences*Handbook of the Philosophy of Science, Handbook of the Philosophy of Science (North-Holland, Amsterdam), pp. 375–402.

Jureta, I. J., Borgida, A., Ernst, N. A. and Mylopoulos, J. [2014] The requirements problem for adaptive systems, *ACM Transactions on Management Information Systems (TMIS)* **5**(3), 1–33.

Lahijanian, M., Almagor, S., Fried, D., Kavraki, L. and Vardi, M. [2015] "This time the robot settles for a cost: A quantitative approach to temporal logic planning with partial satisfaction," in *Proceedings of the AAAI Conference on Artificial Intelligence*.

Lesser, V., Decker, K., Wagner, T., Carver, N., Garvey, A., Horling, B., Neiman, D., Podorozhny, R., Prasad, M. N., Raja, A. *et al.* [2004] Evolution of the gpgp/taems domain-independent coordination framework, *Autonomous agents and multi-agent systems* **9**(1-2), 87–143.

Letier, E. and Van Lamsweerde, A. [2004] "Reasoning about partial goal satisfaction for requirements and design engineering," in *ACM SIGSOFT Software Engineering Notes* (ACM), pp. 53–62.

Mylopoulos, J., Chung, L. and Nixon, B. [1992] Representing and using non-functional requirements: A process-oriented approach, *IEEE Transactions on Software Engineering* **18(6)**, 483–497.

Newell, A. [1982] The knowledge level, *Artificial intelligence* **18**(1), 87–127.

Nguyen, C. M., Sebastiani, R., Giorgini, P. and Mylopoulos, J. [2018] Multi-objective reasoning with constrained goal models, *Requirements Engineering* **23**(2), 189–225.

Rosenthal, D. [2012] Higher-order awareness, misrepresentation and function, *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**(1594), 1424–1438.

Rosenthal, D. [2019] Consciousness and confidence, *Neuropsychologia* **128**, 255–265, neural Routes to Awareness in Vision, Emotion and Action: A tribute to Larry Weiskrantz.

Russell, S. J. [2010] *Artificial intelligence a modern approach* (Pearson Education, Inc.).

Sabatucci, L., Seidita, V. and Cossentino, M. [2018] "The four types of self-adaptive systems: a metamodel," in *International Conference on Intelligent Interactive Multimedia Systems and Services* (Springer), pp. 440–450.

Salehie, M. and Tahvildari, L. [2009] Self-adaptive software: Landscape and research challenges, *ACM transactions on autonomous and adaptive systems (TAAS)* **4**(2), 1–42.

Sawyer, P., Bencomo, N., Whittle, J., Letier, E. and Finkelstein, A. [2010] "Requirements-aware systems: A research agenda for re for self-adaptive systems," in *2010 18th IEEE International Requirements Engineering Conference* (IEEE), pp. 95–103.

Sebastiani, R., Giorgini, P. and Mylopoulos, J. [2004] "Simple and minimum-cost satisfiability for goal models," in *International Conference on Advanced Information Systems Engineering (CAiSE'04)*.

Shapiro, S., Lespérance, Y. and Levesque, H. J. [2005] "Goal change," in *IJCAI*, pp. 582–588.

Simon, H. A. [2019] *The Sciences of the Artificial* (MIT press).

Smithies, D. [2019] *The epistemic role of consciousness* (Philosophy of Mind).

Souza, V. E. S., Lapouchnian, A. and Mylopoulos, J. [2012] "(requirement) evolution requirements for adaptive systems," in *2012 7th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*, pp. 155–164.

Souza, V. E. S., Lapouchnian, A., Robinson, W. N. and Mylopoulos, J. [2011] "Awareness requirements for adaptive systems," in *Proceedings of the 6th international symposium on Software engineering for adaptive and self-managing systems*, pp. 60–69.

Szvetits, M. and Zdun, U. [2016] Systematic literature review of the objectives, techniques, kinds, and architectures of models at runtime, *Software & Systems Modeling* **15**(1), 31–69.

Van Lamsweerde, A. and Letier, E. [2000] Handling obstacles in goal-oriented requirements engineering, *IEEE Transactions on software engineering* **26**(10), 978–1005.

Van Riemsdijk, M. B. and Yorke-Smith, N. [2010] "Towards reasoning with partial goal satisfaction in intelligent agents," in *International Workshop on Programming Multi-Agent Systems* (Springer), pp. 41–59.

Vukovic, M. and Robinson, P. [2005] "Goalmorph: Partial goal satisfaction for flexible service composition," in *International Conference on Next Generation Web Services Practices (NWeSP'05)* (IEEE), pp. 6–pp.

Walton, D. [2015] *Introduction to Practical Reasoning* (Cambridge University Press), pp. 1–37.

Williamson, T. [2000] Knowledge and its limits oxford, .

Wooldridge, M. [2001] Reasoning about rational agents, .

Wooldridge, M. and Jennings, N. R. [1995] Intelligent agents: Theory and practice, *The knowledge engineering review* **10**(2), 115–152.

Yu, E. S. [1993] "Modeling organizations for information systems requirements engineering," in *[1993] Proceedings of the IEEE International Symposium on Requirements Engineering* (IEEE), pp. 34–41.

Yu, E. S. and Mylopoulos, J. [1994] "Understanding" why" in software process modelling, analysis, and design," in *Proceedings of 16th international conference on software engineering* (IEEE), pp. 159–168.

Zhou, Y., Van Der Torre, L. and Zhang, Y. [2008] "Partial goal satisfaction and goal change: weak and strong partial implication, logical properties, complexity," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1* (International Foundation for Autonomous Agents and Multiagent Systems), pp. 413–420.