Establishing a Roadmap and Metrics for Conscious Machines Development

Raul Arrabales

Agapito Ledezma

Araceli Sanchis

Computer Science Department
Carlos III University of Madrid,
Leganés, SPAIN
E-mail: rarrabal@inf.uc3m.es

Computer Science Department
Carlos III University of Madrid,
Leganés, SPAIN
E-mail: ledezma@inf.uc3m.es

Computer Science Department
Carlos III University of Madrid,
Leganés, SPAIN
E-mail: masm@inf.uc3m.es

Abstract—From the point of view of Cognitive Informatics, consciousness can be considered as a grand integration of a number of cognitive processes. Intuitive definitions of consciousness generally involve perception, emotions, attention, self-recognition, theory of mind, volition, etc. Due to this compositional definition of the term consciousness it is usually difficult to define both what is exactly a conscious being and how consciousness could be implemented in artificial machines. When we look into the most evolved biological examples of conscious beings, like great apes or humans, the vast complexity of observed cognitive interactions in conjunction with the lack of comprehensive understanding of low level neural mechanisms makes the reverse engineering task virtually unreachable.

With the aim to effectively address the problem of modeling consciousness at a cognitive level, in this work we propose a concrete developmental path in which key stages in the progressive process of building conscious machines are identified and characterized. Furthermore, a method for calculating a quantitative measure of artificial consciousness is presented. The application of the proposed framework is illustrated with the comparative study of different software agents designed to compete in a first-person shooter video game.

Index Terms—Artificial intelligence, cognitive science, intelligent robots.

I. INTRODUCTION

DEALING with a complex phenomenon like consciousness involves taking into account many interrelated dimensions. Even considering a purely functional point of view, setting aside the philosophical implications of phenomenology, the complexity of multiple cognitive processes and their associated combined dynamics makes it difficult to effectively model and design artificial conscious systems.

Whether or not human-like conscious machines are possible remains an open question. However, many different cognitive skills, which usually are identified in humans, are the inspiration of nowadays artificial implementations. This is the

case of artificial visual attention mechanisms (e.g. [1]), face emotion retrieval methods [2], or even imagination in robots [3]. Also, a number of complex cognitive architectures have been proposed with the aim to combine several cognitive functionalities into a single integrated system. Classical examples are ACT-R [4], SOAR [5], BDI agent architectures [6], and the like. Most recent efforts seem to follow the line of embodiment and physically grounded meanings in an attempt to provide artificial implementations with the required world meaningful knowledge. Relevant examples of this trend are LIDA [7], Haikonen's cognitive architecture [8], CRONOS project [9], and ICARUS [10].

Current research in artificial cognitive systems seems to conclude that a number of cognitive skills are to be successfully integrated into a single situated agent in order to build an artificial human-like mind. Furthermore, consciousness (at least from a functional point of view) might emerge as a result of the synergy developed during the interaction of multiple concurrent cognitive processes. But some of the pending design questions are: what is the right combination of cognitive skills? How exactly should they be integrated? What cognitive capabilities are completely required for consciousness? Which others are expendable? Which ones should be implemented first?

With the aim to shed some light into the resolution of the former questions we propose a bio-inspired roadmap for conscious machines development. By looking into how consciousness has evolved in biological organisms we have tried to identify a set of key developmental stages or consciousness levels. The characterization of each level is twofold, architectural and behavioral. Architecture refers to the basic components of the system, while behavior is the outcome of the interaction of existing cognitive skills. Applying the proposed framework, a particular implementation can be evaluated and assigned a level of artificial consciousness. In fact, our present proposal is based on our former proposed scale for consciousness called *ConsScale* [11, 12]. The scale is not oriented to any particular type of agents, but intended to be applicable to both physically-situated agents and purely software agents like web agents. In the latter case, activities like sending a message are

considered at the same level as a physical action in terms of behavioral analysis.

Using agent's architecture analysis and implemented cognitive skills, *ConsScale* defines an ordered list of levels of artificial consciousness. Although the scale was originally conceived to simply provide a qualitative measure of consciousness, a mechanism for calculating an accurate quantitative measure has been also defined (see Fig. 1 for a schematic description of the generic process followed to assess the level of artificial consciousness of an intelligent agent). In the following we argue that these reference levels can be used as a plausible guideline or roadmap in the challenge of building conscious machines.

In the next section we briefly introduce the conceptual levels of consciousness defined in *ConsScale* and their conception as a design roadmap is also discussed. Section 3 introduces the *ConsScale* Quantitative Score (CQS) as a tool for the comparative study of machine consciousness implementations. The meaning of the score is also thoroughly discussed. A particular case study based on video games autonomous agents is described in Section 4, where particular implementations are analyzed and evaluated. Finally, we offer some conclusions based on the experimentation results and expected future developments.

II. A MACHINE CONSCIOUSNESS ROADMAP PROPOSAL

In this work we propose to use *ConsScale*, a bio-inspired scale for measuring consciousness in artificial agents, as a roadmap for the design of advanced implementations of Machine Consciousness (MC).

A. ConsScale Levels of Consciousness

ConsScale has been defined to serve as a reference MC scale that can be applied to any artificial agent. Therefore, it allows comparative study, providing a qualitative measure represented by a level of consciousness. Levels are defined as an incremental progression, i.e. each higher level subsumes lower ones. An agent is said to comply with a given level *if and only if* it complies with that particular level and also with all lower levels. *ConsScale* reference levels can be briefly introduced as follows (see [12] for a thorough description):

Level -1 (Disembodied). This is the initial reference level which corresponds to very simple implementations lacking defined boundaries. In other words, implementations that cannot be regarded as agents, and can be easily confounded with the rest of the environment. There are no characteristic cognitive skills defined for this level. A biological analogy for this level could be an amino acid as part of a protein.

Level 0 (Isolated). Although there is an obvious distinction between body of implementation and environment, there is a total lack of autonomous processing and no sensorimotor machinery is present. There are no characteristic cognitive skills defined at this stage. An isolated chromosome could be regarded as a valid biological analogy for this level.
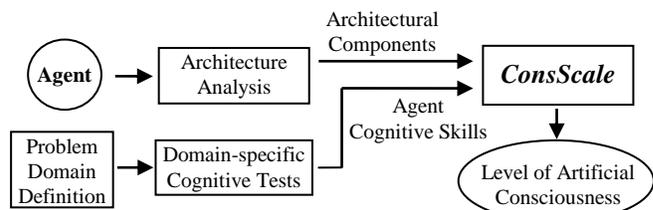
Fig. 1. In order to assess the level of artificial consciousness of an agent using *ConsScale*, its architectural components have to be identified and its cognitive skills tested. Using this information as input, the scale can be used to obtain both a qualitative and a quantitative measure of consciousness.

Level 1 (Decontrolled). This level refers to those implementations where sensors and actuators are present, but there is no working relation between them. As sensing and action are inexistent or unrelated, still no cognitive skills can be defined. Dead bacteria could be an exemplifying analogy from the biological world.

Level 2 (Reactive). Both sensing and action machinery is functional and related by a predefined function. Fixed reactive responses are produced as an invariable function of the input acquired by senses. Primitive situatedness based on reflexes is the only characteristic cognitive skill of this level. A biological analogy for this level could be a virus.

Level 3 (Adaptive). At this level agent's action are a dynamic function of both memory and current information acquired by sensors. Characteristic cognitive skills are basic ability to learn new reflexes and the use of propioceptive sensing for orientation and positioning behaviors. The earthworm could be an illustrative biological analogy for this level.

Level 4 (Attentional). At this level agent's behavior is biased by attention. The attention mechanism selects specific contents out of the total repertory of contents available from sensors and memory. Selected contents are also evaluated positively or negatively, thus constituting the seed for emotions. The cognitive capability of level 4 agents permits the production of attack and escape behaviors. Fish could be a plausible biological analogy for this level.

Level 5 (Executive). Agents at this level are able to interleave multiple goals as different working sets are represented in memory. Characteristic cognitive skills are set shifting and basic emotional learning. Multiple goals can be achieved and most emotionally rewarding tasks are assigned more time and effort. Quadruped mammals are a suitable biological analogy for this level.

Level 6 (Emotional). This level is characterized by the support of Theory of Mind (ToM) stage 1, "I know". ToM is the ability to attribute mental states to oneself and others [13]. From a human developmental standpoint, Lewis suggests four stages in the acquisition of ToM: (1) "I know", (2) "I know I know", (3) "I know you know", and finally (4) "I know you know I know" [14]. At level 6, feelings appear as representations of organism changes due to an emotion [15]. As the effects of emotions in the organism are mapped, a sense of "I know" appears in the agent. Characteristic cognitive skill of level 6 is the ability to develop complex emotional learning.

The agent generalizes the learned lessons to its general behavior, furthermore, emotions are also assigned to the self and self-status monitoring, producing a self-evaluation that gives place to a sense of "I know". Monkeys are the biological analogy of level 6.

Level 7 (Self-Conscious). Self-consciousness is acquired as ToM stage 2, "I know I know", is supported by the agent. The presence of a model of self in the agent makes self-recognition possible. Furthermore, learning mechanisms can operate now in the realm of anticipated future. The agent can plan about itself (as the proper agent is part of the plan), and later learn if the plan was efficient or not. Learning to use tools is a cognitive skill featured in this level, since being a character in the plan is a key factor for learning tool usage [16]. 18 month old human babies are the biologic analogy for level 7.

Level 8 (Empathic). Intersubjectivity is the main characteristic of this level, where the agent is endowed with an enhanced internal model which includes the self but also includes the ability to model others as selves. The support of ToM stage 3, "I know you know" allows social behaviors. Chimpanzees are an illustrative biological analogy for this level.

Level 9 (Social). At this level, the internal model of other selves is enhanced with a full support of ToM, "I know you know I know". This means that characteristic behavior of this level is defined by sophisticated Machiavellian strategies (or social intelligence) involving social behaviors like lying, cunning, and leadership. Also, accurate report and linguistic capabilities are the cognitive features of level 9. Agents at this level are able to develop a culture. 4 year old humans are the biological analogy for this level.

Level 10 (Human-Like). As the name of this level suggests, the corresponding biological analogy is the adult human. The formation of a complex culture and the capability of accurate verbal report are features of this level. This also implies the usage of external complex tools for learning. Fluidity between social and technical intelligence permits the extension of knowledge using external media (like written communication) and technological advances are also possible. Level 10 agents are able to profoundly modify their environment.

Level 11 (Super-Conscious). This last level is characterized by the ability to synchronize and coordinate several streams of consciousness in one physical self. There are no exemplifying examples in the biological world that could be used as a valid analogy for level 11.

B. Applying ConsScale

ConsScale levels are characterized by abstract architectural components and the cognitive skills they generate. Therefore, in order to determine the level of artificial consciousness of an implementation, the presence of a set of concrete cognitive skills (*CS*) has to be tested. Table I contains the list of *CS* defined in *ConsScale* V2.

In order to apply the scale to a particular MC implementation, concrete tests have to be defined to check

TABLE I
CONSSCALE V2 COGNITIVE SKILLS

Level	Cognitive Skills ($CS_{i,j}$)
2	$CS_{2,1}$: Fixed reactive responses ("reflexes").
3	$CS_{3,1}$: Autonomous acquisition of new adaptive reactive responses. $CS_{3,2}$: Usage of proprioceptive sensing for embodied adaptive responses.
4	$CS_{4,1}$: Selection of relevant sensory information. $CS_{4,2}$: Selection of relevant motor information. $CS_{4,3}$: Selection of relevant memory information. $CS_{4,4}$: Evaluation (positive or negative) of selected objects or events. $CS_{4,5}$: Selection of what needs to be stored in memory. $CS_{4,6}$: Trial and error learning. Re-evaluation of selected objects or events. $CS_{4,7}$: Directed behavior toward specific targets like following or escape. $CS_{4,8}$: Evaluation of the performance in the achievement of a single goal. $CS_{4,9}$: Basic planning capability: calculation of next n sequential actions. $CS_{4,10}$: Depictive representations of percepts [17].
5	$CS_{5,1}$: Ability to move back and forth between multiple tasks. $CS_{5,2}$: Seeking of multiple goals. $CS_{5,3}$: Evaluation of the performance in the achievement of multiple goals. $CS_{5,4}$: Autonomous reinforcement learning (emotional learning). $CS_{5,5}$: Advanced planning capability considering all active goals.
6	$CS_{6,1}$: Self-status assessment (background emotions). $CS_{6,2}$: Background emotions cause effects in agent's body. $CS_{6,3}$: Representation of the effect of emotions in organism (feelings). $CS_{6,4}$: Ability to hold a precise and updated map of body schema. $CS_{6,5}$: Abstract learning (learned lessons generalization).
7	$CS_{7,1}$: Representation of the relation between self and perception. $CS_{7,2}$: Representation of the relation between self and action. $CS_{7,3}$: Representation of the relation between self and feelings. $CS_{7,4}$: Self-recognition capability. $CS_{7,5}$: Advance planning including the self as an actor in the plans. $CS_{7,6}$: Use of *imaginational* states in planning [17]. $CS_{7,7}$: Learning of tool usage.
8	$CS_{8,1}$: Ability to model others as subjective selves. $CS_{8,2}$: Learning by imitation of a counterpart. $CS_{8,3}$: Ability to collaborate with others in the pursuit of a common goal. $CS_{8,4}$: Social planning (planning with socially aware plans). $CS_{8,5}$: Ability to make new tools.
9	$CS_{9,1}$: Ability to develop Machiavellian strategies like lying and cunning. $CS_{9,2}$: Social learning (learning of new Machiavellian strategies). $CS_{9,3}$: Advanced communication skills (accurate report of mental content). $CS_{9,4}$: Groups are able to develop a culture.
10	$CS_{10,1}$: Accurate verbal report. Advanced linguistic capabilities. $CS_{10,2}$: Ability to pass the Turing test. $CS_{10,3}$: Ability to modify and adapt the environment to agent's needs. $CS_{10,4}$: Groups are able to develop a civilization and advance culture and technology.
11	$CS_{11,1}$: Ability to manage several streams of consciousness.

whether or not the agent fulfills every $CS_{i,j}$. In other words, while the scale is defined as a generic tool, the application is domain specific, as concrete domain tests have to be specified in order to evaluate the agent.

Examples of *ConsScale* evaluation in the domain of video game agents are given below.

C. ConsScale as a Roadmap

We believe that defining a scale for artificial consciousness is not only valuable as a tool for MC implementations comparative study, but also for establishing a possible engineering roadmap to be followed in the quest for conscious machines. Addressing the challenge of consciousness from the perspective of engineering involves defining a concrete plan. Most of current efforts are focused in particular aspects of consciousness; however, the way in which existing research projects might converge is not clear. Taking biological organisms as inspiration has been an obvious account for decades. Nevertheless, this inspiration usually takes the form of concrete functions or mechanisms, like attention or emotions, neglecting the whole picture. In this work, we propose to consider consciousness as the integrator that puts a mind together. Considering specific cognitive skills as individual or independent components will not help in the design of human-like machines. The idea is to re-define consciousness (at least, from the engineer's point of view) as a grand function which is composed of integrated cognitive capabilities. In this scenario, *ConsScale* suggests a path for the progressive development of artificial conscious agents. Other roadmaps could be defined. However, looking at how consciousness has evolved in nature seems to be a good approach. It does make sense to build *empathic* (level 8) agents in the first place, in order to be able to develop *social* (level 9) agents in the second place.

Whether or not the roadmap proposed in *ConsScale* is the best bet is an open question (possible enhancements of the scale could be thoroughly discussed elsewhere). Nonetheless, it certainly provides a plausible account for the incremental addition of integrated cognitive capabilities into an artificial system. Additionally, the definition and evolution of artificial consciousness metrics will help to determine the current state of the art, and to define the concrete objectives of the MC research field.

III. THE CONSSCALE QUANTITATIVE SCORE

In addition to the qualitative measure provided by *ConsScale* in the form of the 13 conceptual levels of artificial consciousness, a related quantitative score (the *CQS*) can be also calculated.

Having a quantitative score associated with the scale is a benefit from the following points of view: on one hand, as implementations can be evaluated concretely, individual and comparative studies are possible; on the other, specific fitness functions could be defined in order to apply evolutionary computing approaches.

The *CQS* is designed to cover all the possibilities that could exist in a MC implementation, i.e. all possible combinations of $CS_{i,j}$. At the same time, *CQS* also maintains the original constraints of the conceptual scale. In other words, the *CQS* is able to provide a specific numerical value for each conceptual level; additionally, non-canonical implementations which do not comply exactly with a concrete level (because they also show a partial amount of features from other levels) can be quantitatively evaluated and assigned the corresponding value between the two canonical levels where they are situated.

A. Calculating the CQS

The *CQS* can be calculated in three steps. First of all, the quantitative score for each particular level in the scale (L_i) has to be determined. Basically, L_i represents to what extend the implementation under evaluation complies with a given level i. Provided that the implementation fulfills the corresponding architectural requirements, the number of cognitive skills ($CS_{i,j}$) present in the agent is taken into account. Even though the total number of cognitive skills considered for each level is different, the calculation of L_i has to be pondered equally for all levels; in other words, L_i is to provide a measure of the degree of accomplishment of level i, independently of the total number of cognitive features defined for each level. As mentioned in the definition of the scale, the first three levels (-1, 0, and 1) are of reference value, but they do not have any associated cognitive skills, therefore L_{-1}, L_0, and L_1 are always 0 and there is no need to perform any calculation. For the rest of levels (2 to 11) the equation (1) is to be resolved.

$$L_i = \begin{cases} 0 & \textit{if ncsf is } 0 \\ \dfrac{\left(ncsf + \left(J - J_i\right)\right)}{10^3} & \textit{otherwise} \end{cases} \quad (1)$$

Where *ncsf* (number of cognitive skills fulfilled) is the number of cognitive skills ($CS_{i,j}$) that the agent entirely fulfills, J is the maximum number of cognitive skills considered for any level (10 in *ConsScale V2*), and J_i is the total number of cognitive skills defined in level i. L_i is designed as an exponential curve in order to provide a meaningful value correlated with the existing synergy between the various cognitive skills that can be present at a certain level (see Fig. 2). Minimum possible value of L_i is 0 and maximum value is 1. The latter means that level i is fully accomplished by the implementation.

Once the value of L_i has been calculated for each level, a Cumulative Level Score (*CLS*) can be obtained as indicated in (2).

$$CLS = \sum_{i=2}^{11} \left(\frac{L_i}{i-1} \right)^2 \quad (2)$$
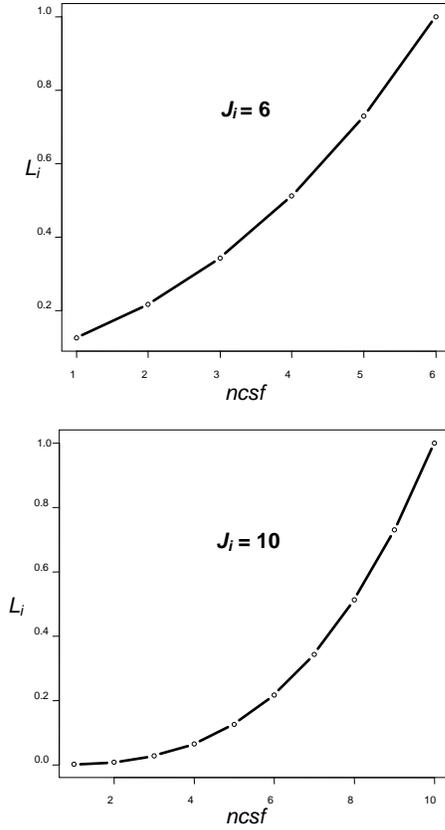
Fig. 2. The graphic in the top represents possible values of L_i for a level with 6 defined cognitive skills; analogously bottom graphic corresponds to a level with 10 cognitive skills. In both representations, X-axis is the number of cognitive skills fulfilled by the implementation (*ncsf*), and Y-axis is the value of L_i which ranges from 0 to 1. L_i values follow an exponential curve to provide the meaning of cognitive function synergy.

CLS combines all particular level scores into a single measure that follows a logarithmic progression (see Fig. 3).

The arrangement of the cumulative measure prevents the overall scale meaning to be distorted by the combined effect of poor L_i scores in lower levels with high scores in higher levels. In other words, *CLS* enforces the quantitative measure to follow the same guideline as indicated in the qualitative measure. Additionally, an implementation can only be ranked as a particular level *i* if and only if it also fulfills all lower levels. For instance, if we have an implementation called *A* that does well at levels 2, 3, 5, 6, and 7, but does not fully comply with level 4, its nominal level is 3. In contrast, another implementation *B* which entirely fulfils levels 2, 3 and 4, although scoring nothing in higher levels, would be ranked as level 4 according to our definition of *ConsScale* qualitative levels. Finally, the *CQS* score is obtained evaluating the exponential function defined in (3).

$$CQS = \frac{e^{\left(CLS^5/K\right)} + a}{10} \qquad (3)$$
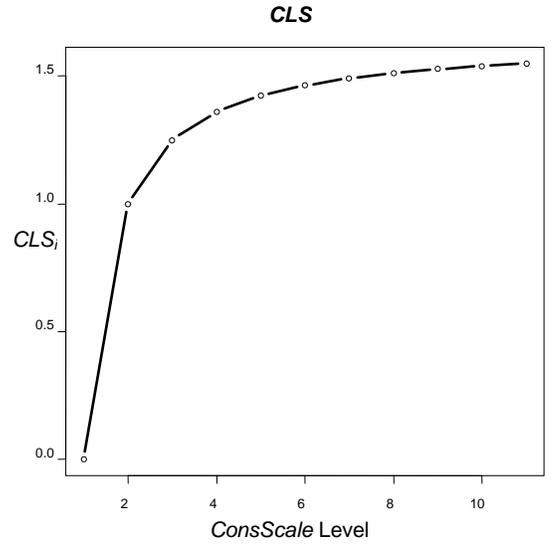
Fig. 3. This curve depicts the values of *CLS* for reference machine consciousness implementations qualifying for levels 1 to 11. X-axis represents *ConsScale* levels and Y-axis the corresponding cumulative level score. *CLS* possible values range from 0 to ~1.55.

Where *K* and *a* are constant numbers (see below) specifically defined to normalize the score possible values from 0 to 1000. Consequently, the minimum score is 0 (which corresponds to *ConsScale* levels -1, 0, and 1) and maximum level of artificial consciousness is represented by a *CQS* of 1000 (for level 11). Fig. 4 shows a graphical representation of the values that *CQS* takes for agents that fall exactly at levels 1 to 11. The rest of agents will obtain a determined score in the *CQS* exponential curve between two consecutive reference levels. The constants *K* and *a* (*0.97062765* and *-1* respectively) are calculated by resolving the simultaneous equations specified in (4).

$$\begin{cases} \dfrac{e^{0/K} + a}{10} = 0 \\[3em] \dfrac{e^{c^5/K} + a}{10} = 1000 \end{cases} \qquad (4)$$

Where *c* is the maximum value that *CLS* can take (*1.549768*).

B. CQS Meaning

The *CQS* curve depicted in Fig. 4 spans from 0 to 1000 in an exponential fashion, representing the cumulative synergy produced by the addition and integration of cognitive features level after level. As the perceived overall cognitive performance increases, the score also increases exponentially, providing significant values from levels 4 and 5 upwards (see Table II).
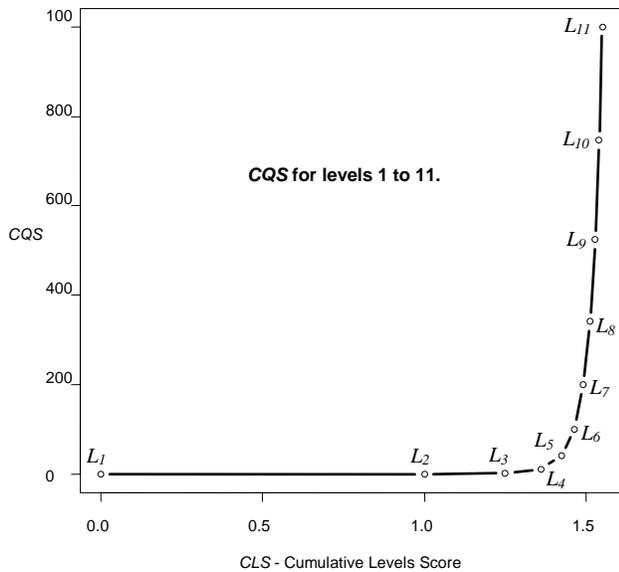
Fig. 4. Representation of *CQS* values for agents qualifying for levels 1 to 11. X-axis represent the *CLS* value, and Y-axis the resulting *CQS*. A *CQS* calculator is available online at http://www.conscious-robots.com/consscale/.

When levels are considered as binary properties, the qualitative level of an implementation can be deduced from its *CQS*. For example, a *CQS* of 2.83 indicates that the implementation is classified as level 3 (because it falls between the minimum *CQS* for level 3, which is 2.22 and the minimum for level 4, which is 12.21); however, since its score is higher than level 3 canonical *CQS*, we also know it fulfills some other additional features from other higher levels. In sum, although all fulfilled cognitive features count, *CQS* appreciably rewards those implementations which follow the path or roadmap defined in *ConsScale*. This is to enforce the bio-inspired roadmap in opposition of other possible developmental paths. We argue that the characterization of the score by the particular roadmap proposed in *ConsScale* is valuable, as we are not only considering an arbitrary combination of different cognitive skills, but an evolutionary integration of incremental synergies.

We consider both composition and ordering from a number of perspectives. From the point of view of perception and action, the following compositional order is considered: 1. Sensing, 2. Adaptation, 3. Attention, 4. Set Shifting, 5. Planning, 6. Imagination. From the point of view of emotions: 1. Emotion, 2. Feeling, 3. Feeling of a feeling [15], 4. Ability to fake emotions intentionally. From the point of view of ToM: 1. "I know", 2. "I know I know", 3. "I know you know", 4. "I know you know I know". *CQS* is consistent with *ConsScale's* conception of levels and rewards this particular ordering. For instance, for a situated agent it is better to have sensing plus adaptation plus attention capabilities rather than just having set shifting and planning capabilities without the former (if somehow possible).

TABLE II
CONSSCALE V2 *CQS* VALUES OF QUALITATIVE REFERENCE LEVELS

Level	Description	CQS
1	*Decontrolled*	0.00
2	*Reactive*	0.18
3	*Adaptive*	2.22
4	*Attentional*	12.21
5	*Executive*	41.23
6	*Emotional*	101.08
7	*Self-Conscious*	200.03
8	*Empathic*	341.45
9	*Social*	524.54
10	*Human-Like*	745.74
11	*Super-Conscious*	1000.00

IV. ASSESSING THE LEVEL OF ARTIFICIAL CONSCIOUSNESS IN VIDEO GAME BOTS

In order to illustrate the application of the proposed scale we have configured an experimentation environment based on the first-person shooter video game Unreal Tournament 3 (UT3) developed by Epic Games, Inc. [18]. Using the *.NET UT3 Bots* custom server modification (referred to as "server mutator") and *UT3RemoteBot* API [19], video game custom bots can be controlled by the CERA (Conscious and Emotional Reasoning Architecture) cognitive architecture [20, 21] (see Fig. 5). UT3 sever is configured to host a deathmatch game in which two or more bots compete against each other. The goal of the deathmatch game is to kill as many other players as possible until a certain match time is reached. In addition to CERA controlled .NET bots, UT3 in-built bots and human controlled bots can also enter the game. The rich interaction between bots and game environment is used to observe behavior of bots and test for the presence of cognitive skills (*CS*$_{i,j}$). Bots behavior can be observed in several ways: the action in the video game can be watched using the original UT3 graphical user interface either in spectator mode or as another active player (see Fig. 6); additionally, specific bot internal state data can be displayed in an ad-hoc interface called CERA UT3 Viewer.

Three different remote bots have been implemented initially: *Reactive-Bot*, *Adaptive-Bot*, and *Attentional-Bot*.

ConsScale qualification of agents requires defining the particular scope of the application domain. Having a clear definition of the environment, possible agent inputs, agent physical (or simulated) capabilities, and possible actions is necessary in order to establish appropriate behavioral tests (see Fig. 1). These tests will be used to confirm or refute the presence of cognitive skills. In this case, we have considered the following limited set of agent capabilities. Actions: bots are able to run, rotate, and fire. Sensing: bots are able to see other bots, detect contact with other bots or objects, count current ammo, and detect damage caused to their body.
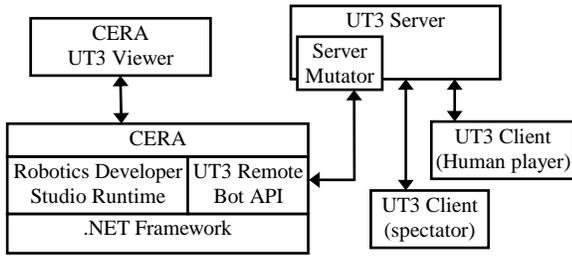
Fig. 5. Communication flow between UT3 server, UT3 Clients (video game graphical user interface) used for watching or joining the game, .NET Remote Bot library, CERA cognitive architecture, and CERA parameters viewer.

In addition to cognitive skills, architectural components are required to be tested in order to perform a *ConsScale* qualification [12]. In this case, all bots are endowed with a virtual body, sensors, actuators, and sensorimotor coordination modules; additionally, *Adaptive-Bot* includes a basic memory module, and *Attentional-Bot* also includes an attention component. CERA is designed as a test bed for MC implementations; therefore, it permits selective activation of architectural modules. Each bot is controlled by a different instance of CERA configured according to the particular bot definition.

Reactive-Bot is programmed to show several reflexes, for instance, backing up whenever it bumps into other bot. Therefore, *Reactive-Bot* can be qualified as level 2 in terms of *ConsScale* as it fulfills $CS_{2,1}$. *Adaptive-Bot* is initially endowed with just one fixed reflex (backing up on bumped); however, it is also programmed to develop new reflexes. Concretely, a reinforcement learning mechanism has been put in place, causing the bot to learn, for instance, to run away to a safer place whenever the perceived internal damage increases. This means that *Reactive-Bot* fulfills both $CS_{3,1}$ and $CS_{3,2}$. Although the name of the third considered bot is *Attentional-Bot*, it does not comply entirely with all $CS_{4,j}$. Consequently it cannot be considered as an *Attentional* agent in terms of *ConsScale*. In fact, it only fulfills 6 out of the 10 *CS* defined for level 4 ($CS_{4,1}$, $CS_{4,2}$, $CS_{4,3}$, $CS_{4,4}$, $CS_{4,5}$, and $CS_{4,7}$). This means the *CQS* measure for *Attentional-Bot* will be located between levels 3 and 4 (see Table III). Basically, this bot is able to pay attention to other bots, evaluate if they are friends or enemies, and develop escape behaviors in case of enemy attack.

Given that *Attentional-Bot* only complies with part of level 4 *CS*, its score is just slightly better than *Adaptive-Bot*'s score. Due to the exponential nature of the scale, a much better score can be obtained if more level 4 *CS* are fulfilled. In this particular case, a much higher score for *Attentional-Bot* (which would imply for instance fulfilling $CS_{4,6}$, $CS_{4,8}$, and $CS_{4,9}$) would mean that this agent is able to develop much more complex and efficient behaviors. This improvement would take place thanks to the synergy between existing skills and additional trial and error learning, self-evaluation, and planning capabilities.

Fig. 6. Screen capture of UT3 Client watching a game.

TABLE III
CALCULATION OF *CQS* FOR VIDEO GAMES BOTS

Agent	L_2	L_3	L_4	CLS	CQS
Reactive-Bot	1	0	0	1	0.18
Adaptive-Bot	1	1	0	1.250	2.22
Attentional-Bot	1	1	0.216	1.255	2.38

V. CONCLUSION AND FUTURE WORK

In this paper we have proposed a cognitive approach to artificial consciousness metrics. Consciousness has been characterized as an integrative grand function that can be developed gradually. Thanks to the definition of both qualitative and quantitative measures, artificial implementations can be evaluated in terms of the proposed scale. Furthermore, we propose to use the scale itself as a roadmap for the design of complex cognitive artificial agents.

The application of *ConsScale* has been illustrated in the domain of video game bots control design. *ConsScale* is conceived to be a general scale able to cover any implementation. Consequently, specific behavioral tests are not included as part of the scale. The required domain-specific tests have to be defined with the aim to prove the existence of the *CS* listed in *ConsScale*. Although we have tried to cover most important cognitive aspects in the current definition of *ConsScale* levels, the scale is open to additions or amendments. We believe that one of the best ways to enhance the scale is by confronting it to major MC implementations and verify if all present cognitive skills are covered. This task is currently underway as part of our main research line. Additionally, we plan to extend the video game environment described above in order to address the design and implementation of higher level agents.

REFERENCES

[1] S. Frintrop, "VOCUS: A Visual Attention System for Object Detection and Goal-directed Search," Lecture Notes in Artificial Intelligence, vol. 3899, 2006.

[2] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis and S. D. Kollias. (2005). Emotion recognition through facial expression analysis based on a neurofuzzy network. Neural Networks 18(4), pp. 423-435.

[3] M. Shanaham, "Consciousness, emotion, and imagination. A brain-inspired architecture for cognitive robotics," in *AISB Workshop Next Generation Approaches to Machine Consciousness*, pp. 26-35, 2005.

[4] J. R. Anderson, M. Matessa and C. Lebiere, "ACT-R: A theory of higher level cognition and its relation to visual attention," *Human Computer Interaction*, 12(4), pp. 439-462, 1997.

[5] A. Laird, J. E. Newell and P. S. Rosenbloom, "SOAR: An Architecture for General Intelligence," *Artif. Intell.,* 33(1), pp. 1-64, 1987.

[6] A. S. Rao and M. P. Georgeff, "BDI-agents: From theory to practice," in Proceedings of *the First Intl. Conference on Multiagent Systems*, pp. 312-319, 1995.

[7] U. Ramamurthy, B. Baars, S. K. D'Mello and S. Franklin, "LIDA: A working model of cognition," in *Cognitive Modeling*, 2006, pp. 244-249.

[8] P. O. A. Haikonen, Robot Brains. Circuits and Systems for Conscious Machines. UK: John Wiley & Sons, 2007.

[9] O. Holland. A strongly embodied approach to machine consciousness. *Journal of Consciousness Studies* 14 pp. 97-110. 2007.

[10] P. Langley and D. Choi, "A unified cognitive architecture for physical agents," in Proceedings of the *Twenty-First National Conference on Artificial Intelligence*, pp. 1469-1474, 2006.

[11] R. Arrabales, A. Ledezma and A. Sanchis, "Criteria for consciousness in artificial intelligent agents," in *ALAMAS&ALAg Workshop at AAMAS 2008*, 2008, pp. 57-64.

[12] R. Arrabales, A. Ledezma and A. Sanchis, "ConsScale: A plausible test for machine consciouness?" in Proceedings of the *Nokia Workshop on Machine Consciousness*. 2008, pp. 49-57.

[13] L. S. Vygotsky. Mind in Society: The Development of Higher Psychological Processes. 1980.

[14] M. Lewis, "The Emergence of Consciousness and Its Role in Human Development," *Ann NY Acad Sci*, 1001(1), pp. 104-133, 2003.

[15] A. R. Damasio, The Feeling of what Happens: Body and Emotion in the Making of Consciousness. London: Heinemann, 1999.

[16] S. Sasaki, T. Hongo, K. Naitoh and N. Hirai. The process of learning a tool-use movement in monkeys, with special reference to vision. *Neuroscience Research* 60(4), pp. 452-456.

[17] I. Aleksander and H. Morton, "Why Axiomatic Models of Being Conscious?" *Journal of Consciousness Studies*, vol. 14, pp. 15-27, 2007.

[18] Epic Games, Inc. Unreal tournament 3. http://www.unrealtournament3.com. Retrieved Nov. 2008.

[19] A. Sterland and J. Lissiak, "UT3 .NET Bots," http://www.codeplex.com/UT3Bots. Retrieved Nov. 2008.

[20] R. Arrabales, A. Ledezma and A. Sanchis, "Modeling consciousness for autonomous robot exploration," in Lecture Notes in Computer Science, vol. 4527, pp. 51-60, 2007.

[21] R. Arrabales, A. Ledezma and A. Sanchis, "CERA-CRANIUM: A Test Bed for Machine Consciousness Research," International Workshop on Machine Consciousness 2009, to be published.