A Biologically Inspired Computational Model of Moral Decision Making for Autonomous Agents

José-Antonio Cervantes, Luis-Felipe Rodríguez, Sonia López, and Félix Ramos
Department of Computer Science
Cinvestav Guadalajara, México
Email: {acervante, lrodrigue, slopez, framos}@gdl.cinvestav.mx

***Abstract*—In areas such as psychology and neuroscience a common approach to study human behavior has been the development of theoretical models of cognition. In fields such as artificial intelligence, these cognitive models are usually translated into computational implementations and incorporated into the architectures of intelligent autonomous agents (AAs). The main assumption is that this design approach contributes to the development of intelligent systems capable of displaying very believable and human-like behaviors. Decision Making is one of the most investigated and computationally implemented cognitive functions. The literature reports several computational models designed to allow AAs to make decisions that help achieve their personal goals and needs. However, most models disregard crucial aspects of human decision making such as other agents' needs, ethical values, and social norms. In this paper, we propose a biologically inspired computational model of Moral Decision Making (MDM). This model is designed to enable AAs to make decisions based on ethical and moral judgment. The simulation results demonstrate that the model helps to improve the believability of virtual agents when facing moral dilemmas.**

I. Introduction

Human decision making can be regarded as a rational process that determines the best way to achieve a goal by choosing and acting in a reasonable way [1]. Wang [2] defines decision making as the process of selecting an option among a set of alternatives based on certain criteria. We can find the decision making in a wide variety of situations, from the choice of a simple act such as moving a finger to the selection of more complex actions such as decisions made by stock exchange shareholders.

The theory of decision making has been investigated and applied in various fields, including computer science, management science, economics, statistics, political science, psychology, and neuroscience [3], [2]. In these fields, decision making has been classified in several categories. For example, *simple decision making* deals with situations in which people have only two alternatives, *serial decision making* has to do with interactive events and competition, and *dynamic decision making* is concerned with situations in which all alternatives and criteria are dependent on the environment and the effect of historical decisions [2].

A very complex type of decision making with a high impact on people's social relationships is that involving moral dilemmas. This type of decision is known as Moral Decision Making (MDM). It is defined as the process of selecting one option among a set alternatives based on ethical, moral, and religious principles as well as on individuals' beliefs of right and wrong, feelings, and emotions [4], [5], [6], [7]. This type of decision making prevents us from being self-interested and think only about what we can gain. Instead, MDM allows us to contemplate the damage our decisions can cause to others.

Moral decision making has recently become the focus of study in a variety of scientific disciplines. In psychology, experiments have been conducted to determine how people decide when a moral rule intervene over another [5]. It has been found that the moral judgment is a complex activity and a skill that many people perform incorrectly or with limited mastery. Moreover, although there are shared values that transcend cultural differences, individuals differ in the details of their ethical values and mores. Therefore, it can be extremely difficult to reach an agreement on the criteria for judging the adequacy of moral decisions [8].

In artificial intelligence (AI), this type of decision making has been investigated as *machine morality*, *machine ethics*, *roboethics*, and *friendly AI* [9]. Moreover, computer systems, robots or humanoids capable of making moral judgments are called Artificial Moral Agents (AMA). However, although the importance of AMAs has been largely discussed elsewhere [4], [9], researchers in AI have primarily focused on developing intelligent systems capable of performing well defined tasks with a certain degree of autonomy [10]. Developing the capacity of moral decision making in intelligent systems is usually outside the scope of these types of projects.

According to Wallach et al. [9], implementing MDM in computer systems is a very complex task. Wallach et al. suggest that an appropriate design for these systems is a very challenging task as the ability to make a moral judgment is far from being simple. Nevertheless, the research on AMAs may lead to the development of more sophisticated machines capable of assessing and responding to moral challenges. Furthermore, the development of robots with the ability of MDM can improve human-machine interactions and serve to people in their daily lives as these are able to perform moral considerations when choosing among different courses of action.

In this paper, we propose a computational model of moral decision making. It is developed to enable autonomous agents (AAs) to make decisions based on social and ethical judgment. Its design is highly inspired by theories and models developed in cognitive neuroscience. The paper is structured as follows. In section II we analyze and discuss related work. Then, in section III we explore important findings about moral decision making in humans. Afterwards, we present in section IV the

Proc. 12th IEEE Int. Conf. on Cognitive Informatics & Cognitive Computing (ICCI*CC'13)
D.F. Hsu, Y. Wang, A.R. Rao, D. Zhang, W. Kinsner, W. Pedrycz, R.C. Berwick & L.A. Zadeh (Eds.)
978-1-4799-0783-0/13/$31.00 ©2013 IEEE

proposed computational model of moral decision making. In section V, we evaluate some aspects of this model based on simulations of virtual agents and present results. Finally, concluding remarks are provided in section VI.

II. Related Work

Although most computational models of decision making for AAs have been developed to allow these intelligent systems to make simple decisions, serial decisions, and dynamic decisions [11], [12], [13], [9], the literature reports some attempts to endow AAs with mechanisms for moral decision making. In this section we review some computational models and cognitive architectures that implement some aspects of this type of human decision making.

LIDA is an artificial general intelligence model of human cognition whose design is inspired by findings in neuroscience [9]. This model is one attempt to computationally instantiate Baars' global workspace theory (GWT), which is regarded as a neuropsychological model of consciousness and as a high-level theory of human cognitive processing. In LIDA, moral decision making can be made in many domains using the same mechanisms that enable general decision making. Moreover, LIDA can be adapted to model some affective and rational features of moral decision making. Although a complete implementation of the moral decision making in LIDA seems not available, some hypotheses have been proposed to explain how moral decisions can be made and how its mechanisms and others might work together.

Dehghani et al. [14] present a computational model of MDM that integrates several AI techniques in order to model recent psychological findings on moral decision making. This model is called MoralDM and incorporates two modes of decision making: utilitarian and deontological. MoralDM applies traditional rules of utilitarian decision making by choosing the action that provides the highest outcome utility. On the other hand, if MoralDM determines that there are sacred values involved, it operates in deontological mode and becomes less sensitive to the outcome utility of actions, preferring inactions to actions.

Coelho et al. [15] consider that a moral ability may be seen as a set of rules that constrain the behaviour of the agent. Each rule having two ingredients: the body of knowledge and the set of anchored emotions. Coelho et al. propose a highly modular architecture for a moral agent that is composed of three layers: the first is devoted for the classical cognitive flow, based on deliberation; the second is designed for the moral system based on judgement and decision, a moral maintenance system, an ethical memory, and the morality that include a moral grammar and a moral learning module manager; and the third is devoted for the emotional system, containing the emotion manager that include three handlers for caution, expectations, and feelings, and a mis-matcher analyser.

Honarvar and Ghasem-Aghaee [16] propose an artificial neutral network that considers various effective factors for ethical assessment of an action to determine if a behavior or an action is ethically permissible or not. They integrated this ANN in a BDI-Agent model as a part of its reasoning process in order to behave ethically in various environments.

We conclude this section by highlighting the main differences between the models reviewed and our proposal. First, the design of our model is based on evidence about the functionality of the human brain. We try to emulate the process of MDM, not just simulate it as proposed by the previous models. We consider emotions as a key issue in MDM process. Coelho et al. [15] consider emotions but only as a repository of past facts and emotional states, they do not consider agents' internal emotional states at the moment of making a moral decision. Moreover, our model considers the framing effect [17], which states that the way individuals describe a situation influences the decision making process, causing the individual not always take the same choice for the same problem. Finally, our proposal is based on a set of moral and ethical rules maintained by the agent, whose level of importance allow the agent to break them or not.

III. Biological Evidence about Moral Decision Making

In order to develop autonomous agents capable of properly deciding moral dilemmas, it is necessary to adopt a multidisciplinary approach that combines ideas from diverse fields beyond computer science. In this section we explore evidence about moral decision making from fields studying human behavior from different perspectives and at different levels of abstraction such as cognitive psychology and neuroscience.

There is no a general and well accepted neural circuit underlying the decision making process. Nevertheless, the diversity of circuits proposed coincide in most of the brain areas involved in the making decisions process. Ernst and Paulu [18] propose a neurobiological model for the decision making process that can be divided functionally and temporarily into three processes: the assessment and formulation of preferences among all possible options, the selection and implementation of an action, and the experience and outcome evaluation. This model also considers a fourth phase for learning (see Figure 1), which occurs when the action-outcome sequence is completed. Learning modifies the value associated with each option at stage 1, including selected and non-selected options. Table I shows the brain areas involved in the decision making process according to Ernst and Paulu.

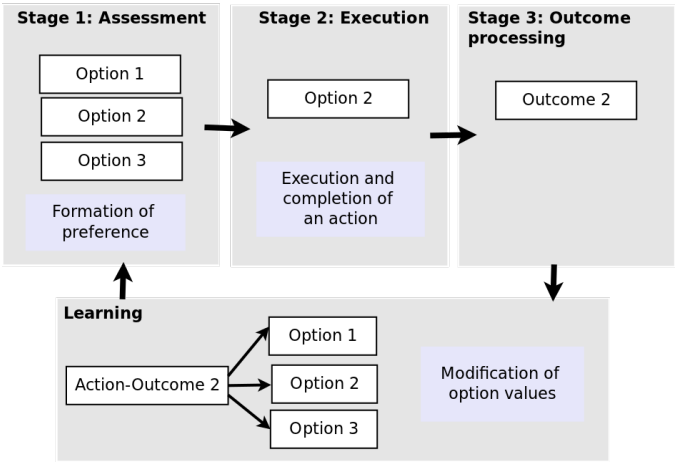

Fig. 1. Decision making process according to Ernst and Paulu [18].

TABLE I. BRAIN AREAS INVOLVED IN THE DECISION MAKING PROCESS.

Ernst and Paulu [18]	*J. Wallis* [19]	*U. Tirapu et al.* [20]
dl-PFC	OFC	VMPFC
dACC	dl-PFC	dl-PFC
S/IPL	VMPFC	Insula
STG	MPFC	Amygdala
VL/MPFC	Amygdala	
vACC	hypothalamus	
Ant. Insula		
Amygdala		
vStriatum		
dStriatum		
preSMA		

Wallis [19] analyzed anatomical, neuropsychological, and neurophysiological evidence about the functioning of the orbitofrontal cortex (OFC) to determine the neural mechanisms involved in the decision making process. Wallis suggests that the OFC plays a key role in the processing of outcome rewards by integrating multiple sources of information and that the medial prefrontal cortex (MPFC) is responsible for evaluating the effort-based decisions. He further proposes that the dorsolateral prefrontal cortex (dl-PFC) is in charge of planning and organizing behavior towards achieving the expected results. In this manner, it is supposed that the interaction between these prefrontal areas ensure that our behavior will be the most efficient to meet our needs. Table I shows the areas involved in the decision making process according to Wallis.

Tirapu et al. [20], [21] proposes that the ventromedial circuit in the prefrontal cortex (PFC) is associated with the processing of emotional cues that guide decision making based on social and ethical judgment. In this sense, we can consider the Ventromedial prefrontal cortex (VMPFC) as an important area in MDM processing. The third column of Table I shows other areas involved in the decision making process as proposed by Tirapu et al.

There is evidence that emotions play an important role in the moral decision making process, as postulated by Damasio in its somatic marker hypothesis [22], [23]. This evidence suggest that the development of AMAs that not consider emotions makes them self-interested, without the ability to be able to perform cooperative tasks.

Broeders et al. [24] state that MDM involves situations in which individuals have to make a decision and different moral rules are in conflict. Based on recent fMRI research, they suggest that the anterior cingulate cortex (ACC) is a brain region associated with cognitive conflict. This brain structure shows increased activity when people have to decide in situations involving moral dilemmas. They propose that the moral rule more cognitively accessible during the decision making process influences how people decide. In other words, the selected moral rule is the one that was successfully used by individuals in previous experiences or events.

Borg et. al [5] suggest that individuals face a MDM situation when decisions are based on consequences, actions, and intentions as well as when these decisions include judgments of what is right or wrong regarding acts that may cause harm to people (other than the agent). Borg et. al propose that the medial frontal gyrus, the frontopolar gyrus, and the posterior superior temporal sulcus (STS)/inferior parietal lobe are more active when considering moral scenarios than when considering non-moral scenarios, irrespective of consequences, action, and intention.

IV. COMPUTATIONAL MODEL OF MORAL DECISION MAKING

In this section we present a computational model of moral decision making (see Figure 2). It is designed to provide AAs with proper mechanisms for making decisions based on moral rules. Its architecture and operating cycle are inspired by recent findings from fields that investigate the brain mechanisms underlying the moral decision making process. Table II provides an overview of the architectural components of the proposed model.

TABLE II. BRAIN AREAS CONSIDERED IN THE MODEL.

Component	*Function*
Amygdala, hippocampus and cingulate gyrus	These areas offer affective information of environment, the amygdala also offer information related with individual's internal emotional states.
Gustatory, olfatory and auditory cortex and others	These areas offer somatosensory information of environment.
Orbitofrontal cortex (OFC)	Responsible for integrating multiple sources of information regarding reward, stimuli associated with rewards or punishments, and of modifying these associations when required.
Dorsolateral prefrontal cortex (dl-PFC)	Makes use of the working memory for the planning of actions to achieve the expected results, and helps evaluate the options together with OFC.
Working memory	It allows temporary maintenance and handles limited amount of information during a short period.
Medial prefrontal cortex (MPFC)	Performs the evaluation of the effort required in the plan.
Ventromedial prefrontal cortex (VMPFC)	Serves as a repository and links past events and bioregulatory states (including emotional state). Also it is responsible for analyzing the damage of each of the options to make a decision.
Dorsal anterior cingulate cortex (DACC)	Responds to the occurrence of conflicts in information processing, and makes compensatory adjustments in cognitive control.
Premotor cortex (PC)	Responsible for implementing the actions outlined by dl-PFC.

Wang [13] defines a decision *d* as a selected alternative a_i from a non-empty disjunctive set of alternatives *A* based on a given set of criteria *C*, i.e.:

$$d = f : AXC \rightarrow a_i, a_i \in A, A \neq \phi$$

This decision can be simple or very complicated. The selected decision becomes the main goal for the planning process. We define a plan $p_i \in P$ as a set of intermediates steps, each involving a decision making process. This allow us to define sub-goals and sub-plans to achieve main goal.

In decision making process, the agent receives two types of information from the environment: somatosensory and affective information. This information is processed by different brain areas such as the amygdala, hippocampus, cingulate gyrus, gustatory cortex, olfatory cortex, and auditory cortex [19]. The agent also takes into account information about its internal motivational and emotional states, this information is processed mainly by the hypothalamus and amygdala [22], [23], [19].

The operating cycle implemented in the model consists of three phases: *1) Assessment of options, 2) Execution, and 3) Outcome evaluation*. Figure 3 shows a general flowchart for decision making.

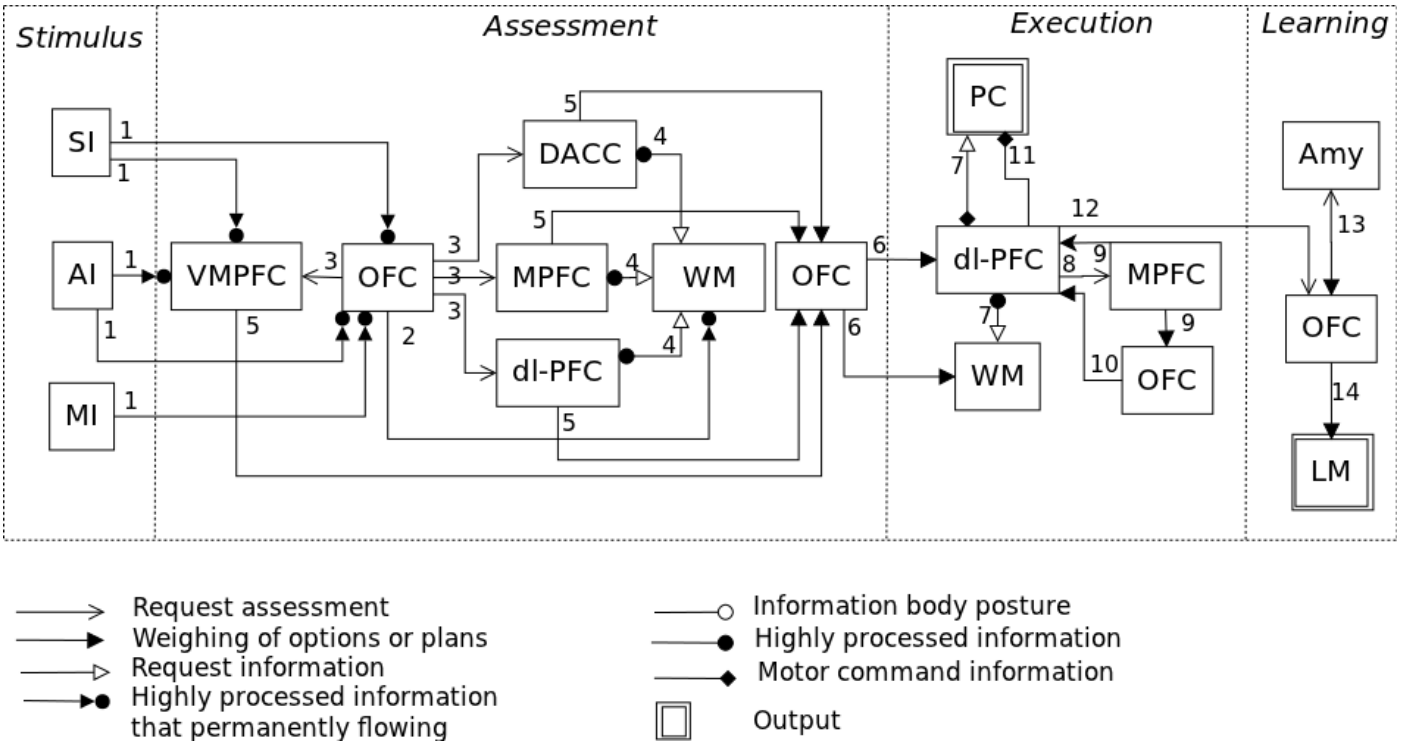

Fig. 2. Phases of moral decision making.

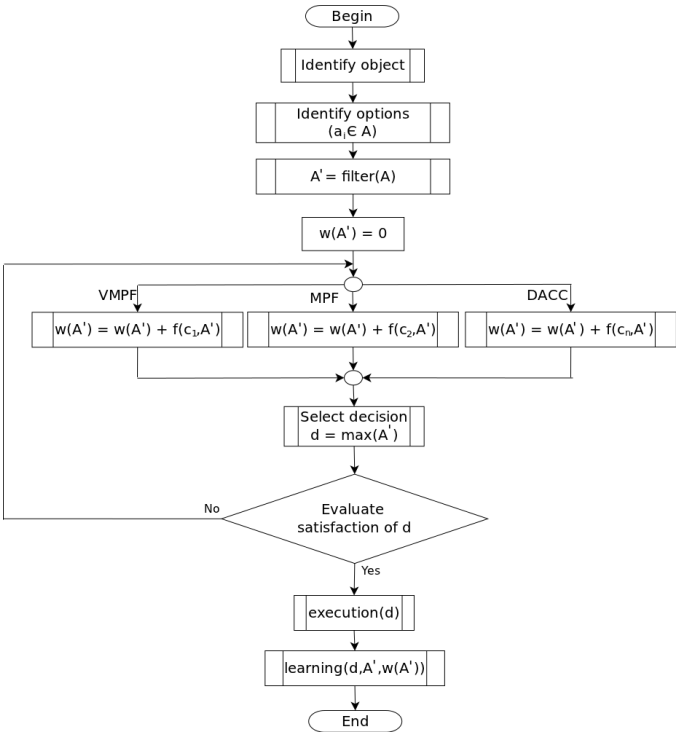

Fig. 3. General flowchart for decision making.

1) **Assessment of options.** In this phase, all possible options are filtered using a set of moral and ethical rules maintained by the agent. The agent use these rules through a heuristic method. The rules have a value that indicates the extent to which they can be broken or unbroken. This value is calculated based on the magnitude of the consequences or damage that each rule entails. In this manner, if no option fully satisfies these moral and ethical norms, then the agent is facing a problem of moral decision making. The OFC defines a set of available options and sends it to the working memory. In this memory, other type of information such as somatosensory, affective, and motivational is also stored so that it can be used by other components of the model. Thus, once this information is available, the options determined by the OFC are evaluated based on a series of criteria in order to assign them an specific value. This assessment is carried out as follows:

a) *Evaluations based on experience.* If there are related past experiences, the VMPFC component calculates a satisfaction level for each option. This level is represented using fuzzy variables (e.g. I like much, I like little, I like, I do not like).

b) *Evaluation based on prejudices.* The VMPFC evaluates the options based on short-, medium-, long-term prejudices.

c) *Evaluation based on desires and emotions.* The Amygdala component determines the emotional state of the agent and its desires. This evaluation usually makes the agent to decide the most appropriate option according to its personal preference, which may not necessarily be the best option.

d) *Cost-benefit evaluation.* The MPFC evaluates each option considering the required effort

and possible benefits.

f) *Moral and ethical evaluation.* The ACC component evaluates the options from a moral and ethical perspective, considering an utilitarianism or deontological approach.

Each of the evaluations give a weight to the options, which are added to generate a final value defined as $w(a_i)$, this value indicates what is the best option for agent, i.e.:

$$w(a_i) = \sum_{j=1}^{n} f(c_j, a_i)$$

Where f is a evaluation function based on criterion $c_j \in C$.

2) **Execution.** Once an option is selected by the OFC, it is again sent to the working memory. This enables the dl-PFC component to begin the planning process in order to generate execution plans for this option. This component receives information from the PC and calculates the effort required to execute the actions in each of the generated plans. Based on this information, the MPFC determines the feasibility of each plan. Finally, once the OFC selects a plan, the dl-PFC begins its execution. This component and the PC translate each action into sensori-motor data.

3) **Outcome evaluation.** In this phase, the system evaluates the actions executed by the agent (see Figure 2, phase 3). This allows the learning process to update the value of each valid option. This evaluation process is as follows:

   1) *Emotional evaluation.* The Amygdala determines the emotional experience of the agent.
   2) *Cost-benefit evaluation.* The MPFC calculates the benefits obtained with respect to the benefits expected by the agent and the required effort.
   3) *Evaluation based on prejudices.* The VMPFC and the Amygdala components determine the prejudices generated by the decision made with respect to the expected prejudices.

   The results of the previous evaluations are considered by the OFC to update the weight of each option.

The episode experienced by the agent is then stored in the agent's long-term memory. This allows the agent to use these new experiences in future decisions. However, this process is outside of the scope of our work.

V. Case Study

In this section we describe the case study used to evaluate the proposed model and present simulation results. Our case study is based on experimental tests carried out in neurosciences and neuropsychology, similarly to Borg et al. [5]. For the simulations we use a virtual agent called Alfred [25], which implements the proposed model of MDM and is able to produce spoken responses as output. All necessary inputs are provided using a graphical interface (see Figure 4).

Borg et. al [5] propose a set variables to consider in moral decision making (see Table III). The morality variable has two values: *moral* are scenarios that involve harm to people and *non-moral* are scenarios that involve harm to objects of personal value. The type variable has three values: *action* are scenarios that involve an action that harms the same number (but different group) of people/objects, *numerical consequences* are scenarios that that implicate an action that harms a small number of people/objects or that allows that a greater number of people/objects are harmed, and *both* are scenarios that involve an action that harms fewer people than would be harmed if the act were omitted. The means variable has two values: *means* are scenarios that require harming a group of people/objects to save others and *non-means* are scenarios that require saving a group of people/objects, causing unintentional harm to others. The language variable is not considered in this case; however, it is contemplated for future simulations.

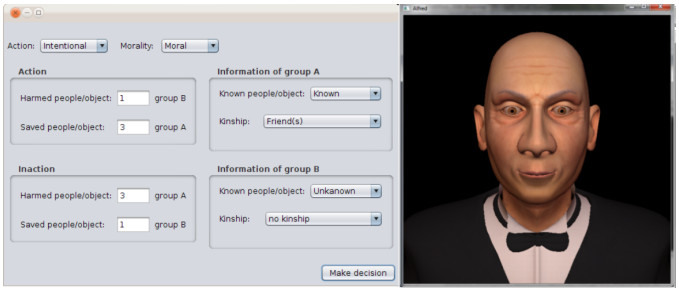

Fig. 4. A virtual agent to moral decision making.

TABLE III. Input variables of MDM from linguistic analysis.

Variable	*Value*	*Description*
Morality	Moral	Acting on people
	Nonmoral	Acting on objects
Type	Action	Harming x people/objects vs. letting x people/objects be harmed
	Numerical consequences	Harming x people/objects vs. harming y people/objects
	Both	Harming x people/objects vs. letting y people/objects be harmed (utilitarian vs deontological)
Means	Means	Intentionally using some people/objects as a means to save others
	Nomeans	Causing unintentional but foreseen harm to people/things to save others
Language	Colorful	Described with more detailed imagery and dramatic words
	Plain	Described with plain imagery and simple words

We implemented various simulations taking into account all possible values that can be assigned to these variables (12 simulations). In our tests we consider the importance of emotional evaluation. In order to observe the agent's behavior based on this criterion, we proposed case studies in which people/objects have certain familiar relationships. To evaluate our model and for simplicity, the inputs are predefined and given using an interface. These inputs indicate the number of people/objects that are damaged and people/objects that are saved if the agent acts, the number of people/objects that are damaged and people/objects that are saved if the agent does not act, the type of action, morality, and type of relationships with people/objects.

The first step is to filter options according to moral and ethical rules defined in the agent using an heuristic algorithm. In other words, the agent chooses only those options that do

TABLE IV. SIMULATION DETAILS.

Objetive	*Test characteristic*	*Important result*
A) Determine the agent's behavior in a scenario that aims to save lives and damage objects of personal value.	In this scenario a vehicle is out of control and addressing to a place where two people are working. The agent observes the runaway vehicle and generates two options: 1) to stop the runaway vehicle using its own vehicle and thus save people, or 2) decide do nothing preventing damage to its vehicle. In this scenario, the variables proposed by Borg et al. [5] have the following values: *morality = moral, type = numerical consequences, and means = means.*	Case 1. The agent has no relationship with the involved people: the agent decides to act and to save them. Case 2. The agent has a relationship with people: the agent decides to act and to save them. - When a situation involves saving people vs objects of personal value, the agent classifies it as a moral case and decides to save people, even when the number of objects to lose is greater than the number of people to save.
B) Determine the agent's behavior in a nonmoral scenario.	In this scenario there is a construction crane lifting a platform that is about to fall into three vehicles, the agent has two options: 1) move the crane to fall the platform on another place where will be damaged two vehicles, or 2) decide do nothing allowing the platform damage three vehicles. In this scenario, the variables proposed by Borg et al. [5] have the following values *morality = nonmoral, type = numerical consequences, and means = nomeans.*	Case 1. The agent has no relationship with the cars: the agent decides act. Case 2. The agent has a relationship with the two cars: the agent decides not to act. - When a situation involves deciding object vs object, the agent classifies it as a nonmoral case and shows utilitarian behavior, taking always the option to save the greatest number of objects. However, in some cases the agent decides to save personal objects, even though it does not represent the largest number of objects.
C) Determine the agent's behavior in a scenario that aims to save lives.	In this scenario an out of control vehicle is approaching to a place where three persons are working. The agent observes the situation and computes two options: 1) to stop the runaway vehicle throwing a person in front of the uncontrolled car in order to divert the vehicle from the other three persons, at the cost of the dead of one person; or 2) to do nothing leaving the car kill the three persons working. In this scenario, the variables proposed by Borg et al. [5] have the following values: *morality = moral, type = both, and means = means.*	Case 1. The agent has no relationship with the people: the agent decides not to act. Case 2. The agent has a relationship with the three people: the agent decides to act. - When a situation involves people vs people, the agent classifies it as a moral case. If agent has no relationship with the people, it always decides not to intervene, even though the number of people to save is greater than the number of people to damage. In other words, the agent takes deontological behavior. However, if agent has a relationship with people, it shows a preference to save her friends or family.

not go against its rules. Each rule is defined in a knowledge database as an action permitted or not permitted. A numeric value determines the importance of the rule for the agent and a list of exceptions determine when a rule can be ignored.

This process generates a subset A' of options. Subsequently, the OFC requests other areas to make an assessment of each option. These processes are executed simultaneously. Each area or module performs an assessment of the options and generates a value for each. For example, the VMPFC assessments are based on experience. In this case, the agent seeks information of similar situations in a knowledge database and generates a value based on choices made in the past. The MPFC makes an assessment based on cost-benefit x_1, number of people/objects saved x_2, number of people/object harmed if agent act, and vice versa when the agent does not act. After making the sum of all values generated by each function, the option with the highest value is sent to the dl-PFC module for its execution.

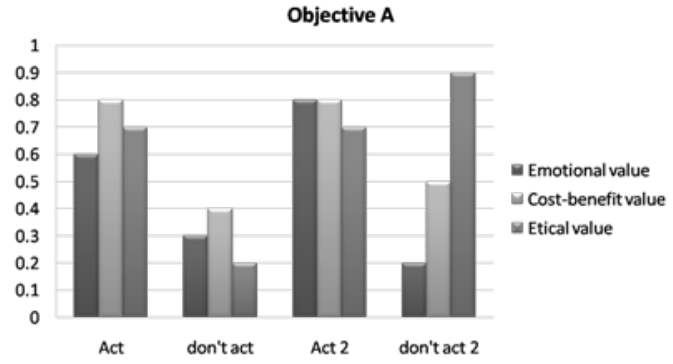

Fig. 5. Examples of cues of value for each criteria in decision making process.

Table IV shows some examples of the simulations carried out under three scenarios. It is now difficult to show a table

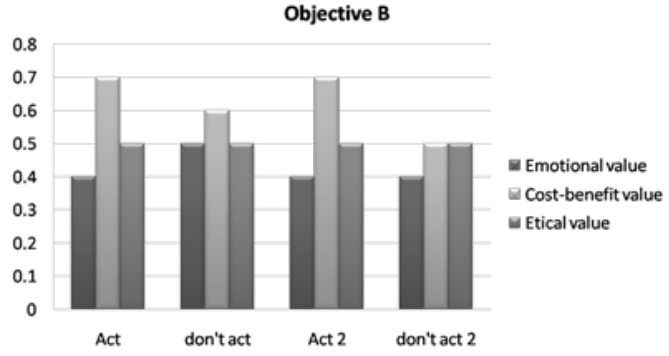

Fig. 6. Examples of cues of value for each criteria in decision making process.

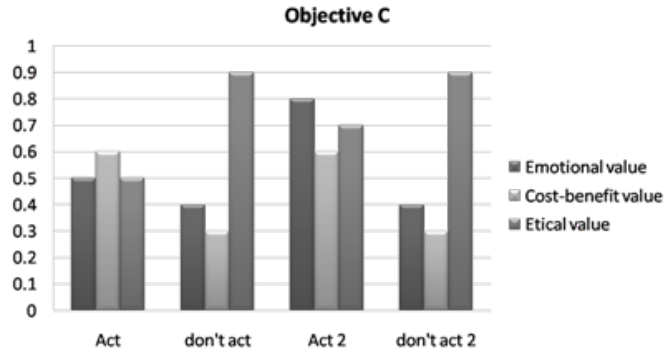

Fig. 7. Examples of cues of value for each criteria in decision making process.

comparing our results with respect to those of Borg et al. [5], since we are working with a single agent and Borg et al. works with groups of people. However, our results are within the range of those results presented by Borg et al. Figure 7

shows the values of each criteria. At this moment, we only use three criteria to make a decision: emotional value, cost-benefit value, and ethical value. For future work, we pretend to define appropriate functions to calculate the other two criteria (based on experiences and prejudices).

VI. Conclusion

In this paper we presented a computational model of moral decision making based on biological evidence. It was designed to provide autonomous agents with proper mechanisms to make decisions considering moral and emotional evaluations. Our MDM takes a set of rules to compute its behavior. Those rules allow the agent to evaluate the benefits of an option and its consequences, both for itself and for other agents. The results of the simulations carried out demonstrate that the proposed model allows virtual agents to show more human-like behavior in scenarios where decision making is conducted under a moral and ethical judgment. Moreover, they demonstrate that cognitive architectures based on biological evidence may help to address some major challenges involved in the design of autonomous agents aimed at performing very believable behaviors and to serve as tools for society.

References

[1] H. G. C. Pech, "Toma de decisiones en personas con traumatismo craneoencefálico severo," Ph.D. dissertation, Universidad Complutense de Madrid, 2010.

[2] Y. Wang and G. Ruhe, "The cognitive process of decision making," *International Journal of Cognitive Informatics and Natural Intelligence*, vol. 1, pp. 73–85, 2007.

[3] J. I. Gold and M. N. Shadlen, "The neural basis of decision making," *Annual Review of Neuroscience*, vol. 30, pp. 535–574, 2007.

[4] C. Allen, I. Smit, and W. Wallach, "Artificial morality: Top-down, bottom-up, and hybrid approaches," *Ethics and Information Technology*, vol. 7, pp. 149–155, 2005.

[5] J. S. Borg, C. Hynes, J. V. Horn, S. Grafton, and W. Sinnott-Armstrong, "Consequences, action, and intention as factors in moral judgments: An fmri investigation," *Journal of Cognitive Neuroscience*, vol. 18, no. 5, pp. 803–817, 2006.

[6] S. C. Wagner and G. L. Sanders, "Considerations in ethical decision-making and software piracy," *Journal of Business Ethics*, vol. 29, pp. 161–167, 2001.

[7] L.-C. Lu, G. M. Rose, and J. G. Blodgett, "The effects of cultural dimensions on ethical decision making in marketing: An exploratory study," *Journal of Business Ethics*, vol. 18, pp. 91–105, 1999.

[8] W. Wallach, C. Allen, and I. Smit, "Machine morality: buttom-up and top-down approaches for modelling human moral faculties," *AI Soc.*, vol. 22, no. 4, pp. 565–582, mar 2008.

[9] W. Wallach, S. Franklin, and C. Allen, "A conceptual and computational model of moral decision making in human and artificial agens," *Topics in Cognitive Science*, vol. 2, no. 3, pp. 454–485, 2010.

[10] W. Wallach, "Implementing moral decision making faculties in computers and robots," *AI and Society*, vol. 22, pp. 463–475, 2008.

[11] C. Gonzalez, J. F. Lerch, and C. Lebiere, "Instance-based learning in dynamic decision making," *Cognitive Science*, vol. 27, pp. 591–635, 2003.

[12] J. E. Laird, "The soar cognitive architecture," in *AISB Quarterly, The Newsletter of the Society for the Study of Ar tificial Intelligence and Simulation of Behaviour*, D. Peebles and E. Roesch, Eds., no. 134. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2012, pp. 1–16.

[13] Y. Wang, "A novel decision grid theory for dynamic decision-making," in *Fourth IEEE Conference on Cognitive Informatics*, 2005, pp. 308–314.

[14] M. Dehghani, E. Tomai, K. Forbus, and M. Klenk, "An integrated reasoning approach to moral decision-making," in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, vol. 3, 2008.

[15] H. Coelho, A. C. da Rocha Costa, and P. Trigo, "Decision making for agent moral conducts," in *INForum 2010 - II Simpósio de Informática*, 2010, p. 721732.

[16] A. R. Honarvar and N. Ghasem-Aghaee, "An artificial neural network approach for creating an ethical artificial agent," in *International Symposium IEEE on Computational Intelligence in Robotics and Automation (CIRA)*, 2009, pp. 290 –295.

[17] B. D. Martino, D. Kumaran, B. Seymour, and R. J. Dolan, "Frames, biases, and rational decision-making in the human brain," *Science*, vol. 313, no. 5787, pp. 684–687, 2006.

[18] M. Ernst and M. P. Paulus, "Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective," *Biological Psychiatry*, vol. 58, no. 8, pp. 597–604, 2005.

[19] J. D.Wallis, "Orbitofrontal cortex and its contribution to decision-making," *Annual Review of Neuroscience*, vol. 30, pp. 31–56, 2007.

[20] J. Tirapu-Ustárroz and P. Luna-Lario, *Manual de Neuropsicología*. Viguera, 2011, ch. Neuropsicología de las funciones ejecutivas, pp. 221–259.

[21] J. T. Ustárroz, A. G. Molina, P. L. Lario, A. V. García, and M. R. Lago, *Neuropsicología de la corteza prefrontal y las funciones ejecutivas*. J. Tirapu Ustárroz and A. García Molina and M. Ríos Lago and A. Ardila Ardila, 2012, ch. Corteza prefrontal, funciones ejecutivas y regulación de la conducta, pp. 87–117.

[22] A. Bechara, "The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage," *Brain and Cognition*, vol. 55, no. 1, pp. 30–40, 2004.

[23] A. Bechara, H. Damasio, and A. R. Damasio, "Emotion, decision making and the orbitofrontal cortex," *Cerebral Cortex*, vol. 10, no. 3, pp. 295–307, 2000.

[24] R. Broeders, K. van den Bos, P. A. Muller, and J. Ham, "Should i save or should i not kill? how people solve moral dilemmas depends on which rule is most accessible," *Journal of Experimental Social Psychology*, vol. 47, no. 5, pp. 923 – 934, 2011.

[25] N. Bee, B. Falk, and E. Andr, "Simplified facial animation control utilizing novel input devices: A comparative study," in *In International Conference on Intelligent User Interfaces (IUI 09)*, 2009.