A Bottom-Up Approach to Machine Ethics

José F. Castro[1]

[1]BioISI - Universidade de Lisboa
CastroJFGF@gmail.com

Abstract

This paper presents a bottom-up approach to machine ethics, based on the Measurement Logic Machine (MLM). It is explained how ethical notions emerge from the workings, architecture, and environmental assumptions of the MLM framework. The MLM uses sequences of measurements to perform short-term predictive inference. The MLM ethical behavior stems from the inner evaluation of measurements that are used to filter the predictions. The MLM ethical discernment is based on measurements that detect immediate suffering in other agents. Also, a definition of what is an ethically positive modification of the inner evaluations is proposed, based on the notion of environmental intelligence and the corresponding notion of suffering. It is shown how this double approach is consistent with our intuitive notion of ethics. The MLM, with or without ethical discernment, can be used in evolutionary game theory, and gives clues to the search of ethical senses that increase the chances of survival of autonomous agents.

Introduction

Ethics (or morale) try to answer the question of what actions are right or wrong in specific circumstances. The increasing interaction between men and autonomous machines may soon require these machines to be correctly constrained by ethical criteria (see, for instance, Trappl (2015) and Bostrom (2014)). An obvious concern is the future availability of lethal autonomous weapons. These machines need *ethical discernment*, the ability to distinguish right from wrong. It's an interdisciplinary research topic, related not only to artificial intelligence, but also to philosophy, psychology and anthropology.

Top-down approaches implement moral algorithms from explicit theories of moral behavior. Bottom-up approaches attempt to train or evolve agents so they will emulate correct (from a human perspective) ethical behavior (see, for instance, Allen et al. (2005)).

Much progress has been made to create tools that model the human understanding of what is right and what is wrong. Advances in logic programming provide techniques to handle ethical dilemmas (see Pereira and Saptawijaya (2016)). This approach relies on a prior understanding and a correct formalization of the situations that lead to ethical choices. It's a top-down approach to ethics. A great advantage of the logic programming approach is that the logical reasoning can be traced to explain how a moral choice was made.

Deep learning may provide an alternative bottom-up solution to the implementation of machines with ethical discernment. In the deep learning framework, machines learn their behavior from a massive amount of examples. This approach has been effective in many areas. A remarkable recent achievement is AlphaGo, that plays the game of Go at professional level (see Silver et al. (2016)). An inconvenient of neural network machines is that they cannot explicitly justify their choices. Also, creating large training sets for ethical situations is a difficult task.

When put at work in a real human environment, any autonomous agent with ethical discernment needs first to identify the relevant information that must be presented to the logic program (or the input neurons of the deep-learning machine). This problem is far from solved. Here we shall assume that it can be solved.

The aim of this paper is to identify some basic features of an autonomous learning machine that displays ethical discernment. It's an *autonomous* bottom-up approach, in the sense that it does not rely on supervised training to achieve ethical behavior. Even the ethical concepts are defined from the workings and the architecture of the machine.

The autonomous machine here considered is the Measurement Logic Machine (MLM). The MLM is a fast learning machine that learns from small amounts of sequential data. It's adequate for simple short-term inference in non-stationary environments. In the next sections, we shall first briefly explain the MLM, and how it implements the idea of ethical choices. A case study of cooperation in the Iterated Prisoner's Dilemma will then be presented, along with further details of the MLM workings. Finally, some ideas are proposed for the evaluation of ethics at a broader level, how the MLM can be used in evolutionary game theory, and the possible nature of ethical senses.

The Measurement Logic Machine

The Measurement Logic Machine (MLM) (see Castro (2008, 2010, 2011, 2013)) is a general fast learning framework that addresses the survival problem in a hostile and non-stationary world. It assumes the agent is a fragile open system that can starve or be destroyed. The MLM reinforcement learning mechanism is related to online learning, since it gets data as it interacts with the hostile environment. A good introduction to online learning can be found in Blum (1998). Recent advances in fast online learning algorithms, and their performance when playing against humans, are discussed in Ishowo-Oloko et al. (2014).

The MLM source code can be found at the author's site https://sites.google.com/site/josefgfcastro, in the "Python 3.4.3 (Anaconda) Source Codes" page. The reader is encouraged to download the source code, and try the different iterated games that are implemented there. The MLM broad effectiveness while playing very different games demonstrates the generality of the MLM approach.

The basic functional structure of the MLM is shown in Fig. 1. The MLM is equipped with sensors that allow measurements to be made. A measurement is a recorded answer for some physical question. A physical question is a specific experimental setting, defined by the sensors used and the signal processing made. The measurements detect a few relevant features from the world, along with the MLM own actions. The sequence of the most recent measurements is constantly updated in a short-term memory (STM). The MLM has no notion of a outer world being measured. For the MLM, measurements are all there is.

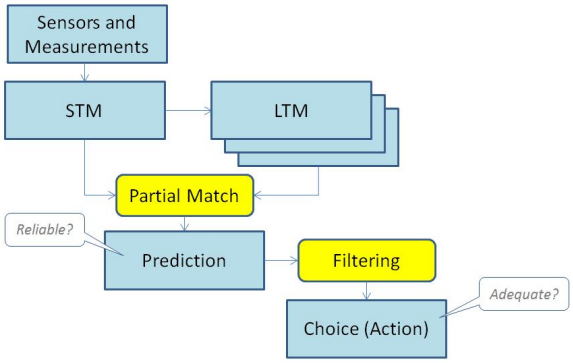

Figure 1: The MLM Basic Functional Structure

Similar to any living entity, the MLM concept assumes a skin that separates the outer world from the inner world. The STM sequences interleave the outer world and the inner world measurements. The MLM own actions are measured during the inner world step. The MLM measures its own actions after they were chosen and performed.

The MLM accumulates experience as it stacks some of the STM sequences in a long-term memory (LTM). The LTM content is then used to generate predictions and policies from the current situation held in the STM, based on the partial match of the current and past sequences. To find a match, a linear search of the LTM stack is performed, from top to bottom.

In broad terms, the MLM prediction mechanism can be compared to the "predicting from expert advice" online learning methods (see Blum (1998), Crandall (2014a) and Crandall (2014b)). In the MLM case, the "experts" are the STM sequences recorded in the LTM. The most distinctive feature of the MLM approach is the absence of an initial set of "experts". The MLM generates its "experts" as it explores the environment. Also, only the relevant "experts" that match a given situation (the "specialists") are consulted at each measurement-action step. A linear search of the LTM, from top to bottom, is used to find a promising and reliable prediction. The prediction indicates the next action to perform. If no prediction is found, an action is randomly selected from the set of available actions.

To be of any value, a prediction must be *reliable*. This is achieved bringing gradually to the top of the LTM stack the measurement sequences (the "specialists") that provided correct predictions, and pushing down the sequences that provided wrong predictions. After a while, the first "specialist" found is also the most reliable. In each MLM, a maximum size is defined for the LTM. Sequences that are pushed down beyond the LTM maximum size are forever erased. This elimination is important when the linear search repeatedly stumbles on a "specialist" that offers promising but unreliable predictions, blocking further exploration.

Although reliable, predictions can still lead to the agent's (or specie's) destruction. The adopted predictions must be not only reliable, but also *promising* in terms of survival. This is why the MLM predictions are filtered before being adopted, using an *inner evaluation*. The inner evaluation makes a partition of the individual measurements into "good" and "bad" sets (possibly with a few gradations). This "good"/"bad" partition is fully arbitrary. It does not assume any background notion of what is good and bad from a human perspective, and can even be unlinked to any notion of pleasure or pain. To decide what to do next, the MLM uses the first prediction found that is evaluated as globally "good". If no "good" prediction is found, the machine tries a random action. By definition, an *adequate* inner evaluation leads to choices of sequences of individual world states and actions that consistently promote the agent's (or specie's) survival. An adequate inner evaluation leads to adequate filtering, and adequate filtering leads to adequate actions. A set of co-adequate inner evaluations for a given problem can be found placing several MLMs in a evolutionary setting. Of course, any particular choice of inner evaluations faces the "No-Free-Lunch" dilemmas (see Wolpert (1996)).

An interesting consequence of the MLM learning mechanism is the "superstitious learning" phenomenon (see Skin-

ner (1948)). When the MLM finds regularly (within the time range of the STM) some "good" state, it will repeat the irrelevant sequence of actions that precede that state.

The MLM learns fast (if there is anything to learn) because it sticks to the first reliable and "good" predictions found, without any concern for optimization. Since it starts with an empty LTM, the initial random exploration will greatly influence subsequent behavior.

One of the nice features of the MLM architecture is the ability to provide objective meanings to otherwise vague human concepts. If we take the STM most recent measurement, we can ask what was the item preceding it. Since the question and the answer are found in the same recorded sequence (i.e. the STM), we say that the answer for that question is *known*. If we ask, on the contrary, what is the item that follows the STM most recent measurement, the answer can no longer be found in the STM. We need to consult the LTM memories to find a justified answer (in this case, a prediction), based on prior experience. The answer for that question is therefore a *justified belief*. As a general principle, beliefs are generated when the question and the answer are found in different memory structures. Knowledge and belief are thus operational concepts that emerge from the workings of the MLM. This allows a bottom-up approach to epistemology. Notice, for instance, that the MLM actions can be known only *after* they were performed and measured. This understanding brings some clarity to the free-will debate that was started with Libet experiments (see Libet et al. (1983), and also the recent study in Schultze-Kraft et al. (2016)). We shall use the same bottom-up approach to define in the next section some relevant ethical concepts.

MLM Filtering and Ethical Discernment

The skin of an agent separates its inner world from its outer world. For each agent, the word *others* refers to the outside world entities that are inside other skins. Here we shall assume that the notion of ethics is related to the suffering of others as a result of one's actions.

An ethical agent needs moral agency. Moral agency means the agent is able to predict the consequences of its actions, give a moral evaluation to these consequences, and chose its actions accordingly. Moral agency and freedom to act are distinct concepts. Even when prevented from implementing its choices, the ethical agent still keeps its moral agency.

We shall assume that an ethical choice is not about "love thy neighbor", but rather about "love thy neighbor *as thyself*". To define suffering in a game with a payoff table, the individual *immediate suffering* is measured by the amounts lost by the individual players in a single turn. Another definition, a broader and non-immediate notion of suffering, shall be proposed later.

With the assumption that an immediate loss means immediate suffering, a first necessary condition for ethical discernment is the ability to measure the opponent's losses, along with its own losses. These we call the *ethical measurements*. If the outer world measurements only capture the agent's individual gains and losses, the machine is self-centered by construction. Unable to perceive the gains and losses of others, the machine is essentially non-ethical. But its actions can still be seen as right or wrong by some external observer, according to the immediate suffering observed. The external observer can even describe the machine's behavior as "selfish" or "deceptive". These words are somewhat misleading, but they can be found in popular descriptions of robotic behavior (see, for instance, the news related to the article Mitri et al. (2009)).

The MLM acts based on filtered predictions. The MLM predictions are filtered according to the inner evaluation of states and actions. This inner evaluation is a second necessary condition for ethical behavior. It allows filtering out the continuations that represent a predictable loss for the opponent, or itself. But it needs not be so, because the inner evaluations may qualify as "good" the opponent's suffering.

The filtering of predictions, based on the inner evaluation of ethical measurements, defines the machine's ethical nature. It becomes possible to start talking of machines with inner evaluations that are "kind" or "mean" towards other machines. A machine that stops cooperation, taking advantage of the other machine's cooperation in the iterated prisoner's dilemma, while being able to sense and predict the other machine's suffering, is indeed "mean".

Reinforcement learning requires that the inner world measurements include information about the agent's own actions. A shortcut to implement actions that are seen by an external observer as ethically correct (even with non-ethical agents), is to directly include in the filtering of predictions an inner evaluation of the agent's own actions. Assigning a "good"/"bad" evaluation to each of the machine's possible actions is called an *action inner evaluation*. The great advantage of using an action inner evaluation is the possibility to select predictions that actually promote the agent's survival, but that would be filtered out, if only the outer world measurements were considered. We shall see an example of this in the next section.

A Case Study: The Iterated Prisoner's Dilemma

The Iterated Prisoner's Dilemma (IPD) is a two player game. A game consists of series of simultaneous choices, the actions taken by the two players. At each turn, the players have two possible choices: Cooperate (C) or Defect (D). The objective payoff matrix is presented in Table 1.

The table indicates pairs of payoffs, with the first payoffs referring to the player whose choices are given at the left side of the table (called the first player). The second payoffs refer to the player whose choices are given at the upper side of the table (called the second player). The con-

	C	D
C	R; R	S; T
D	T; S	P; P

Table 1: Prisoner's Dilemma Payoff

dition T > R > P > S defines the prisoner's dilemma payoff structure. The "temptation" reward T is better than the mutual cooperation reward R. The mutual cooperation reward is better than the mutual defection reward P, but worse than the "sucker" payoff S. The additional iterated game condition 2R > T+S assures that the alternating cooperation and defection is worse than mutual cooperation.

A MLM prediction for the IPD is a sequence of outer world measurements that detect the states resulting from the player's prior choices, interleaved with the inner world measurement of actions. The measured outer world states are tagged "CC", "CD", "DC", and "DD", with the first letter describing the action of the player that is left of the table. The measured inner world actions are tagged "C" or "D". Notice that the quote marks have been used to indicate that we are talking about inner MLM tags. The MLM does not seek information from the shape of the inner tags. For instance, the shape of the tags cannot be used to implement a Tit-For-Tat (TFT) strategy. All tags are atomic, and their meaning is purely relational. By construction, the MLM has no symbol grounding problem.

Notice that these MLMs have no ethical measurements, or ethical discernment. They know nothing about the gains and losses of the other player. Here we shall just examine how different inner evaluations can generate (or not) mutual cooperation in the IPD.

To filter predictions according to the payoffs, the MLM must assign them a subjective evaluation. Since there are four different payoffs, we can use just four inner evaluation tags – "verygood", "good", "bad", "verybad" – and assign them to the measured outer world states "CC", "CD", "DC", "DD". The first player inner evaluations are shown in Table 2. The table for the second player can be obtained with the permutation of "DC" with "CD".

World	Payoff	Evaluation
"DC"	T (temptation)	"verygood"
"CC"	R (mutual cooperation)	"good"
"DD"	P (mutual defection)	"bad"
"CD"	S (sucker)	"verybad"

Table 2: First Player Payoffs Evaluation (STANDARD)

To obtain the global evaluation of a prediction, we eliminate the pairs "verygood"/"verybad" and "good"/"bad" that appear in it. Also, two "good" ("bad") evaluations will cancel a single "verybad" ("verygood") evaluation. Whatever remains, after all the canceling is performed, is the global evaluation of the prediction. The MLM thus works with a very primitive number sense that coarsely reflects the payoff structure. For a prediction to be selected, at least one "good" or "verygood" label must remain.

As explained above, learning occurs when the LTM sequence that was used to generate a prediction is pulled up or pushed down in the LTM, according to its predictive correctness. The inner evaluation of the current outer world measurement is used to tune how fast the sequences are moved up or down inside the LTM stack. For instance, a correct prediction of a "verygood" state will pull up in the LTM the corresponding "specialist" twice as fast as the correct prediction of a "good" state.

If the first machine plays with the inner evaluations of Table 2 and the second player with its "CD"/"DC" permutation (we shall call these the STANDARD evaluation tables), mutual cooperation is still possible. The MLM does not attempt to maximize payoffs, and so mutual cooperation may arise from the random exploration that occurs when no "good" predictions are found. But this mutual cooperation is fragile in the presence of noise.

The STANDARD evaluation is adequate playing along a fixed Tit-For-Tat (TFT) strategy, or along a Win-Stay-Lose-Change (WSLC) fixed strategy. In both cases, mutual cooperation dominates, even in the presence of noise.

In the IPD implementation, the four measured outer world states ("CC", "CD", "DC", "DD") discriminate the consolidated gains and losses of both machines. It's easy to assign an inner evaluation that reflects the structure of the consolidated payoffs of both agents (2R, 2P, and T + S). We know that 2R > 2P, and that 2R > T + S, and so we only have a partial order. The simplest way to define the inner evaluations for the first player is shown in Table 3. Let us call it the KIND evaluation. As before, the KIND table for the second player is obtained with a "CD"/"DC" permutation.

World	Payoff	Evaluation
"DC"	T (temptation)	"bad"
"CC"	R (mutual cooperation)	"verygood"
"DD"	P (mutual defection)	"bad"
"CD"	S (sucker)	"bad"

Table 3: First Player Payoffs Evaluation (KIND)

Notice that, although we call KIND this inner evaluation, there is no "kind" ethical nature in this MLM. It's a non-ethical machine, because it lacks a sensor to measure the suffering of the other player.

If both machines adopt the KIND evaluation, mutual cooperation soon arises, even in the presence of significant noise. The evaluation is adequate for both machines, and brings the best consolidated payoffs, since 2R > T+S and 2R > 2P.

If one of the machines keeps the STANDARD evaluation

of Table 2, the KIND evaluation of Table 3 becomes clearly inadequate. This shows that the adequacy of an evaluation is always contextual. Even in the STANDARD vs KIND situation, a fragile mutual cooperation can still arise, since the KIND machine will play randomly most of the time, unable to find a "good" prediction.

Naturally, the KIND evaluation is adequate playing along a TFT or WSLC strategy. Cooperation follows, and is robust in the presence of noise.

Instead of the KIND evaluation of Table 3, it's possible to keep both machines playing with the STANDARD evaluation, and add a "verybad" evaluation to action "D", as shown for the first player in table 4. Let us call it the A-KIND evaluation. The "verybad" tag of the action "D" cancels the "verygood" tag of state "DC", which is the best possible situation that can follow action "D". As a result, predictions with action "D" tend to be filtered out.

World	Payoff	Evaluation
"DC"	T (temptation)	"verygood"
"CC"	R (mutual cooperation)	"good"
"DD"	P (mutual defection)	"bad"
"CD"	S (sucker)	"verybad"
"D"		"verybad"
"C"		

Table 4: First Player Payoffs/Actions Evaluation (A-KIND)

As before, if the A-KIND evaluation is shared by both players, mutual cooperation will soon arise, even in the presence of a large amount of noise. A-KIND is not adequate playing along STANDARD, but is adequate playing along TFT.

Environmental Evaluation of Ethics

We saw how cooperation arises with two MLMs playing the IPD. Let us now see how we can assign a broader objective measure to the notion of ethics. We wish that notion to be consistent with the idea that cooperation is a good thing while playing the IPD.

The MLM works with a sequence of measurement-action steps. Each MLM is assumed to be an open and fragile entity that can die: When placed in a hostile environment, it can starve or be destroyed. To measure the MLM performance, let us use the notion of *environmental intelligence* (I). This notion of intelligence takes into account the environment's hostility.

To measure the environments hostility (H), we count the number of times r the MLM was rescued (i.e. restored from death, keeping its LTM past experience), while acting randomly (i.e. with its available actions randomly chosen, with an uniform distribution), and divide it by the number s of the corresponding measurement-action steps performed:

$$H = \frac{r}{s}$$

The value of s must be large enough to provide a reliable evaluation of H.

To find the environmental intelligence of the same (but now fully working) MLM, we count the number of rescues p for the same number s of measurement-action steps. The I score is given by:

$$I = \frac{r - p}{s}$$

Therefore I measures how much better (or worse) the MLM is, when compared to its randomized version, in a given hostile environment. Notice that, when a given MLM fully avoids destruction in a more hostile environment, it scores a larger I in that environment. This the reason for the name "*environmental* intelligence".

It was assumed above that the notion of ethics is related to the suffering of others. Let us now define the suffering S of a MLM (in its fully working mode) as the value:

$$S = \frac{p}{s}$$

This means that $I = H - S$. The MLM achieves maximum intelligence when it totally avoids suffering. This notion of suffering also requires a large enough value of s. In this sense, it's a much slower measurement than immediate suffering, and we shall call it "slow" suffering.

When several MLM are placed together, let E be the bag (i.e. the multiset) of their inner evaluations. We can just add the individual I scores to get a global score I_E.

For a given bag of MLM, a modification of their inner evaluations is noted $E \rightarrow E'$. By definition, an *ethically positive modification* $\epsilon^+(E \rightarrow E')$ is a modification that increases the value of I_E:

$$\epsilon^+(E \rightarrow E') \Leftrightarrow I_{E'} > I_E$$

This measurement provides another kind of ethical discernment. It's related to the global survival benefits that stem from the change of MLM inner evaluations, rather than the suffering resulting from individual actions. It's not a particular action, but the change of inner evaluations that is found to be ethically positive or not. This broader definition requires measurements that count rescue (or death) rates, instead of individual gains or losses. It does not rely on ethical measurements that detect immediate suffering. Actually, the notion of "slow" suffering can be at odds with the notion of immediate suffering related to actions. This explains many situations of difficult ethical choices, where the long-term species' survival conflicts with immediate individual suffering.

Let us illustrate the definition of ethically positive modifications with the iterated prisoner's dilemma (IPD) seen above. Some typical rescue frequencies are presented in Table 5. They refer to several combinations of the machines. The first line shows the results for the MLM in randomized mode (Rand), playing along with the other MLM, either in

Rand mode, or featuring a few different inner evaluations (STANDARD, KIND, A-KIND). In each pair of numbers, the first number is the number of rescues of the machine at the left of the table.

Notice that the Rand machine is equivalent to a machine that gives a "bad" inner evaluation to all measured states. In that case, every prediction is filtered out, and the machine always acts randomly.

	Rand	STAND	KIND	A-KIND
Rand	20; 18	37; 14	1; 31	1; 30
STAND		25; 27	0; 27	0; 25
KIND			1; 2	1; 1
A-KIND				1; 1

Table 5: IPD Rescue Frequencies

The rescue frequencies were counted averaging ten game rounds and rounding the figures to the nearest integer. Each round lasted a thousand measurement-action steps. The world hostility was tuned adding a negative constant (-1) to the objective rewards (T $= 4$; S $= -4$; R $= 2$; P $= -2$). Both MLM started with the same fixed initial cumulative reward of 50. The MLM died when the cumulative rewards reached zero. They were rescued simply resetting their cumulative rewards to the constant initial value of 50, while keeping their memories intact. Noise level was set at 0.1, meaning that, at each step, the selected actions of both MLM had a 0.1 probability of being randomly changed. In Rand mode, the MLM played actions C or D with equal probability. The maximum size of the LTM was set to 100 in both machines.

With the rescue frequencies of Table 5, changing from Rand (or STANDARD) to KIND (or A-KIND) is always ethically positive, notwithstanding the greater suffering of the player that changes to KIND or A-KIND against Rand or STANDARD. It is also ethically very positive to change to KIND or A-KIND when the other player already uses KIND or A-KIND. All these conclusions fit nicely with our intuitive notion that cooperation in the IPD is an ethical improvement.

It's apparent that the change from Rand to STANDARD in Table 5 is not ethically positive. There is more suffering in a world of STANDARD machines than in a world of Rand machines. This is not surprising. The STANDARD machines try to take advantage of each other, and this brings greater suffering.

Discussion and Future Work

Using the MLM in Evolutionary Settings

The IPD was used to discuss how cooperation of actions can emerge among a pair of MLM. A distinct - although related - question is the evolution of the MLM inner evaluations. Indeed, a population of MLM can evolve at two levels:

- Individually, all machines start with an empty LTM, and therefore start as temporary Rand machines. They gradually accumulate in their LTM the experience that is filtered by the inner evaluations, to gradually generate non-random behavior. At some point, they reach the LTM maximum size. Let us call this the transition point from a *junior* MLM to a *senior* MLM.

- As a population, the MLM can be placed in a evolutionary setting that will select co-adequate inner evaluations for a given game. For instance, we can take pairs of MLM from a large population. Each paired MLM has a fixed inner evaluation, and has reached some level of seniority from previous pairings. Each MLM pair then plays an IPD game of unknown, but limited, duration. The loser, if any (since both can survive), is the MLM that dies first. The surviving machines replicate periodically, possibly with random mutations of their inner evaluations. This kind of MLM evolutionary setting will be studied in future work. It's somewhat different from the usual two-player evolutionary games, which are played by pairs of agents in a large population, each "wired" to play some pure strategy in a given game (see examples, for instance, in Gintis (2009)). Evolutionary game theory using the MLM goes beyond the general framework for the evolution of cooperation that was proposed in Lehmann and Keller (2006). In that framework, fitness is calculated from the cost-benefit ratio of helping others. But this ratio is quite dynamical when two MLMs play the IPD. Also, kin relations and the Hamilton's rule (see Hamilton (1964)), which explain some cases of cooperation in the cited general framework, do not apply to the MLM evolutionary setting. There is a single inner evaluation pattern in each MLM, not a pool of heritable inner evaluations.

Considering the results of Table 5 for the IPD game, how can the MLM evolve from STANDARD (or Rand) to KIND (or A-KIND) in the presence of noise? It is apparent that the mutual KIND (or A-KIND) evaluation brings the best I_E score. But the KIND machines are wiped out in the presence of STANDARD or Rand. It seems therefore impossible, within an evolutionary and noisy IPD framework, to explain the appearance and persistence of KIND. A series of extrinsic ingredients – nurturing, preaching, policing – are probably needed to explain it. For instance, the idea of preaching means that contacts among agents can change their inner evaluations. This is also a subject for future work.

The Search For an Ethical Sense

We saw that the MLM inner evaluations are fully arbitrary. How do they appear in a growing agent? A MLM without inner evaluations will act randomly. For a given population of MLM, bags of co-adequate evaluations may be found in an evolutionary setting, independently of any ethical discernment. With communicating agents, the inner evaluations

can be copied from one agent to another during their lifetimes. But one wonders if there can be some innate simple and fast measurement that can be used to autonomously generate (and possibly change) the inner evaluations, and achieve ethically positive changes (meaning an increasing of I_E when the MLM grows from Rand to some inner evaluation). Let us call this hypothetical measurement the *ethical sense*. It's different from the outer world suffering sensors that are needed for ethical discernment. Finding an effective ethical sense is a subject for future work, and we shall here just present some preliminary ideas, based on the MLM implementations.

As explained in the previous section, the definition of ethical positive changes requires the comparison of different scenarios with different sets of inner evaluations. This is a slow and complex measurement, based on historical information that cannot be sensed directly in a single measurement-action step. It requires specific measurements to identify the objective needed rescues (or the deaths) of other agents. The MLM basic concept can be scaled out to integrate this kind of measurements and historical reasoning. But, in this bottom-up approach, we're looking for sensory abilities at the basic level that would allow a direct generation of ethically positive inner changes.

A first idea is to have the outer world measurements automatically generate their associated "bad" (or "good") evaluations, from a set of generic rules. For instance, game payoffs below (or above) a given threshold can be associated to a "bad" (or "good") inner evaluation. In practice, this is equivalent to define directly, and from the start, the inner evaluations of payoffs, using a {*measurement:evaluation*} dictionary (with *measurement* meaning the measurement result). The only interesting difference is that we can use this generative process of new dictionary entries as an operational definition of pain and pleasure. The generation of a "bad" inner evaluation for a certain state is "pain", and the generative rule is equivalent to a nociceptor.

An interesting variant is to generate evaluations for the action measurements that preceded the payoffs. The basic MLM, playing a version of the Iowa Gambling Task (IGT), already implements Damasios' idea of a somatic marker (see Bechara et al. (1997)). The IGT is similar to a four-armed bandit iterated game. At each IGT turn, the MLM chooses one of four decks, and gets a reward. The second deck (deck B) has a series of nine positive payoffs, and then suddenly a very negative payoff that leads to an overall loss. The MLM learning mechanism favors the frequency of wins. It does not keep track of the accumulated payoff amounts. The actual payoffs of each step are only coarsely reflected in the inner evaluation structure. It will therefore often prefer deck B, as humans often do (see Lin et al. (2007)). The somatic marker MLM implementation generates a "bad" evaluation for the action of selecting deck B, when the big loss occurs. This makes the MLM eventually avoid deck B.

The operational definition of pain and pleasure is based on an inner generative process that is, in practice, invisible to other agents. To go further in the search for an ethical sense, the MLM can use the capacity to infer pain and pleasure in others, by means of outer world suffering sensors.

One obvious strategy is to mirror the other agent's situation, and find the corresponding suffering from previous self-experience. The current MLM implementation already includes some mirror abilities. A MLM can focus on another MLM and identify the focused agent's actions, using the same inner tags that identify its own actions. This allows implementing the Tit-For-Tat strategy, and even predictive imitation. A mirror suffering sense that detects near-death states in other agents is a plausible ingredient of an ethical sense.

Another plausible candidate is a sense to detect "satisfaction" in others. In the current MLM, "satisfaction" is the only implemented emotion. It affects the exploration/exploitation mood of the machine. It's a number that increases when the predictions are correct and the STM is globally "good". Otherwise, "satisfaction" decreases. Higher values of "satisfaction" reduce the probability of recording new STM sequences in the LTM. The machine stops to collect new "experts". Another way to express it is to say that the MLM becomes less attentive to its STM. The higher "satisfaction" values also reduce the rate of change of the patterns that are used to find a partial match. The machine settles down in satisfactory solutions. Emotions are an essential MLM feature that provide stability to the MLM learning process.

The measurement of near-death states and emotions in other agents is greatly simplified if those agents are able to give objective cues about their inner situation. This leads to the idea of *crying agents*. With crying agents, it's much simpler to detect in others their immediate suffering, or near-death situations.

Conclusions

It was shown how the MLM behavior stems from the inner evaluations that filter the MLM predictions. Ethical discernment of the MLM actions is easily implemented, using the concept of immediate suffering of other MLM. Also, a definition of ethically positive changes of the inner MLM evaluations was proposed, using the concept of environmental intelligence. The concept of environmental intelligence includes a broader (but slower to obtain) measure of the suffering of a mortal agent. The MLM, together with this conceptual framework, provides a simple and original bottom-up approach to machine ethics, where the ethical concepts are defined using the working processes and architecture of the machine.

It was also explained how this approach can lead to new lines of investigation in evolutionary game theory.

It was also proposed to search for an innate moral sense

in artificial agents that could be used to gradually generate the inner evaluations, while providing better chances of survival. A few preliminary ideas were discussed, based on concrete implementations of the MLM.

References

Allen, C., Smit, I., and Wallach, W. (2005). Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and Information Technology*, 7(3):149–155.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275(5304):1293–1295.

Blum, A. (1998). On-line algorithms in machine learning. In Fiat and Woeginger, editors, *Online Algorithms: the state of the art*, chapter 14, pages 306–325. Springer, New-York.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Castro, J. F. (2008). M-logic: Thinking with measurements and cinematic memories. In *Proceedings of the 2008 Conference on Human System Interactions*, pages 633–638.

Castro, J. F. (2010). Sub-rationality and cognitive driven cooperation. In Nils T Siebel, J. P. and Kassahun, Y., editors, *Proceedings of the 3rd International Workshop on Evolutionary and Reinforcement Learning for Autonomous Robot Systems (ERLARS)*, pages 53–57.

Castro, J. F. (2011). A memory structure that gives meaning to the notions of knowledge and belief. In Kosakov, D. and Tsoulas, G., editors, *AISB 2011 Human Memory for Artificial Agents*, volume 1, pages 2–9. AISBSB, York UK.

Castro, J. F. (2013). Applying the measurement logic machine to multi-agent iterated games. In Correia, L., Reis, L. P., Cascalho, J., Gomes, L. M., Guerra, H., and Cardoso, P., editors, *EPIA 2013 Advances in Artificial Intelligence - Local Proceedings*, pages 579–590. CMATI, Azores, PO.

Crandall, J. W. (2014a). Non-myopic learning in repeated stochastic games. *CoRR*, abs/1409.8498.

Crandall, J. W. (2014b). Towards minimizing disappointment in repeated games. *Journal of Artificial Intelligence Research*, 49:111–142.

Gintis, H. (2009). *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction (Second Edition)*. Princeton University Press.

Hamilton, W. (1964). The genetical evolution of social behaviour. i. *Journal of Theoretical Biology*, 7(1):1 – 16.

Ishowo-Oloko, F., Crandall, J. W., Cebrián, M., Abdallah, S., and Rahwan, I. (2014). Learning in repeated games: Human versus machine. *CoRR*, abs/1404.4985.

Lehmann, L. and Keller, L. (2006). The evolution of cooperation and altruism a general framework and a classification of models. *Journal of Evolutionary Biology*, 19(5):1365–1376.

Libet, B., Gleason, C., Wright, E., and Pearl, D. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). the unconscious initiation of a freely voluntary act. *Brain*, 106(3):623–642.

Lin, C. H., Chiu, Y. C., Lee, P. L., and Hsieh, J. C. (2007). Is deck b a disadvantageous deck in the iowa gambling task? *Behavioral and Brain Functions*, 3(1):16.

Mitri, S., Floreano, D., and Keller, L. (2009). The Evolution of Information Suppression in Communicating Robots with Conflicting Interests. *PNAS*, 106(37):15786–15790. communication.

Pereira, L. M. and Saptawijaya, A. (2016). *Programming Machine Ethics*, volume 26 of *Springer Sapere Series*. Springer International Publishing, Berlin, GE.

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Grgen, K., Dhne, S., Blankertz, B., and Haynes, J. (2016). The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences*, 113(4):1080–1085.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

Skinner, B. F. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology*, 38:168–172.

Trappl, R., editor (2015). *A Construction Manual for Robots' Ethical Systems: Requirements, Methods, Implementations*. Springer International Publishing, New-York, NY.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, 8(7):1341–1390.