COGNITIVE BIASED ACTION SELECTION STRATEGIES FOR SIMULATIONS OF FINANCIAL SYSTEMS

Marco Remondino and Nicola Miglietta
e-business L@B, University of Turin, Italy

Keywords: Reinforcement learning, Action selection, Bias, Ego biased learning.

Abstract: In agent based simulations, the many entities involved usually deal with an action selection based on the reactive paradigm: they usually feature embedded strategies to be used according to the stimuli coming from the environment or other entities. This can give good results at an aggregate level, but in certain situations (e.g. Game Theory), cognitive agents, embedded with some learning technique, could give a better representation of the real system. The actors involved in real Social Systems have a local vision and usually can only see their own actions or neighbours' ones (bounded rationality) and sometimes they could be biased towards a particular behaviour, even if not optimal for a certain situation. In the paper, a method for cognitive action selection is formally introduced, keeping into consideration an individual bias: ego biased learning. It allows the agents to adapt their behaviour according to a payoff coming from the action they performed at time t-1, by converting an action pattern into a synthetic value, updated at each time, but keeping into account their individual preferences towards specific actions.

1 INTRODUCTION

Agent Based Simulation (ABS) is one of the most interesting paradigms to represent complex social systems. They allow to capture the complexity by modeling the system from the bottom, by defining the agents' behaviour and the rules of interaction among them and the environment. ABS, in this field, is not only about understanding the individual behaviour of agents, or in optimizing the interaction among them, in order to coordinate their actions to reach a common goal, like in other Multi Agent Systems (MAS), but above all it's about re-creating a real social system (e.g.: a market, an enterprise, a biological system) in order to analyze it as if it were a virtual laboratory for experiments. Reactive agents or cognitive ones can be employed in multi agent systems (Remondino, 2005); while the former model deals with the stimulus-reaction paradigm, the latter provides a "mind" for the agents, that can decide which action to take at the next step, based on their previous actions and the state of the world. When dealing with the problem of action selection, reactive agents simply feature a wired behaviour, deriving from some conditional embedded rules that cannot be changed by the circumstances, and must be foreseen and wired into them by the model designer. Reactive agents are good for simulations, since the results obtained by employing them are usually easily readable and comparable (especially for ceteris paribus analysis). Besides, when the agent's behaviour is not the primary focus, reactive agents, if their rules are properly chosen, can give very interesting aggregate results, often letting emergent system properties to come out at a macro level. Though, in situations in which, for example, learning coordination is important, or the focus is on exploring different behaviours in order to dynamically choose the best one for a given state, or simply agent's behaviour is the principal topic of the research, cognitive agents should be employed, embedded with some learning technique. Besides, if the rules of a reactive agent are not chosen properly, they bias the results; being chosen by the designer, they thus reflect her own opinions about the modeled system. Since many ABS of social systems are formulated as stage games with simultaneous moves made by the agents, some learning techniques derived from this field can be embedded into them, in order to create more realistic response to the external stimuli, by endowing the agents with a self adapting ability. Though, multi-agent learning is more challenging than single-agent, because of two complementary reasons. Treating the multiple agents

Remondino M. and Miglietta N. (2009).
COGNITIVE BIASED ACTION SELECTION STRATEGIES FOR SIMULATIONS OF FINANCIAL SYSTEMS.
In *Proceedings of the International Joint Conference on Computational Intelligence*, pages 534-539
DOI: 10.5220/0002311405340539

as a single agent increases the state and action spaces exponentially and is thus unusable in multi agent simulation, where so many entities act at the same time. On the other hand, the actors involved in real Social Systems have a local vision and usually can only see their own actions or neighbours' ones (bounded rationality) and, above all, the resulting state is function of the aggregate behaviours, and not of the individual ones. While, as discussed in Powers and Shoham (2005), in iterated games learning is derived from facing the same opponent (or others, sharing the same goals), in social systems the subjects can be different and the payoff is not a deterministic or stochastic value coming from a payoff matrix, but rather a variable coming from the dynamics of interaction among many entities and the environment, not necessarily contained within a pre-defined scale. Besides, social models are not all and only about coordination, like iterated games, and agents could have a bias towards a particular behaviour, preferring it even if not the best of the possible ones. In the following paragraph evidence is given, coming from Behavioural Finance (BF), that human beings are not completely rational and are often biased in their perceptions.

The purpose of this work is not that of supplying a optimized algorithms; instead, the presented formalisms mimic the real cognitive process by human agents involved in a social complex system, when they face an individual strategic decision.

The work is divided in two parts: in the first part the most important cognitive distortions analyzed by BF are introduced, while in the second part a novel technique is introduced, which keeps into account some of the described perception errors.

2 BEHAVIOURAL FINANCE

The classic theory about expected utility supposes the presence of optimizing behaviours and of complete decisional rationality for the individuals. This is not always true in the real world and many empirical evidences prove that the economic agent features systematic distortions, compared to the prescriptions coming from the theories of markets efficiency. This is studied and formalized by BF. The cognitive distortions taking part in human behaviour are divided into three categories: the *heuristics*, the *biases*, and the *framing effects*.

Heuristics are rules proposed to explain how individuals solve problems, give judgments, take decisions when facing complex situations or incomplete information. The justification for their existence is founded on the assertion for which the human cognitive system is based on limited resources and, not being able to solve problems through pure algorithmic processes, uses heuristics as efficient strategies for simplifying decisions and problems. Even if they succeed in most cases, they could bring to systematic errors. At a psychological level, when the number and the frequency of information increases, the brain tries to find some "shortcuts", allowing to reduce the elaboration time, in order to take a decision anyway. These shortcuts are defined heuristics (or rules of thumb). On one side, they allow to manage in a quick and selective way the information; on the other side, they could bring to wrong or excessively simplified conclusions. The most significant heuristics are: representativeness, availability and anchoring. The first shows how agents tend to make their choices on the basis of stereotypes that could lead to errors caused by wrong estimates. When referring to the availability, the individuals tend to assign a probability to an event, based on the quantity and on the ease with which they remember the event happened in the past. Once again, the heuristic error is the consequence of a simplified cognitive model. Anchoring it the third heuristic behaviour that could generate errors in the decision process; it's the attitude of the individuals to stay anchored to a reference value, without updating their estimates. It's at the bases of conservative attitudes often adopted by economic agents. Last but not least, also "affect heuristics" could impact decision making; by following their emotions and instincts, sometimes more than logically reasoning, some individuals could decide to perform a decision in a risky situation, while not to perform it in other – apparently safer – ones.

The biases are distortions caused by prejudices towards a point of view or an ideology. Bias could be considered a systematic error. The most common biases are the over-optimism, confirmation bias, control illusion, and the excessive self-confidence. Many individuals have excessive confidence in their own means, thus overestimating their capabilities, knowledge and the precision of their information. Confirmation bias is a mental process which consists in giving the most importance, among the information received, to those reflecting and confirming the personal believes and, vice versa, in ignoring or debasing those negating inner convictions. On the contrary, the hindsight bias consists in the error of the retrospective judgment, i.e.: the tendency of people to erroneously believe, after an event has taken place, that they would have been able to correctly predict it a priori. Another basic behavioural principle is the so called "aversion to ambiguity", often referred to as "uncertainty aversion". This can be synthesized in the sentence "*People prefer the familiar to the unfamiliar*" and describes an attitude of preference for known risks over unknown risks, which can bring people to running an higher, though known, risk, over a potentially lower, but unknown, one.

That it is not the same as "risk aversion", which is the reluctance of a person to accept a bargain with an uncertain payoff rather than another bargain with a more certain, but possibly lower, expected payoff.

The term "framing" is referred to a selective influence process on the perception of the meaning of words and sentences; these distortions are derived from Prospect Theory, whose aim is to explain how and why the choices are systematically different from those predicted by the standard decision theory.

Prospect Theory is alternative to that of Expected Utility, when it comes to understanding the human behaviour under uncertainty conditions, and adopts an inductive and descriptive approach. This theoretical foundation can be interpreted as a synthetic representation of the most significant anomalies found in decisional processes under uncertainty.

Some behaviours have been seen as violations to the Expected Utility: the certainty effect, the reflect effect and the isolation effect. The certainty effect is referred to the fact that, when facing a series of positive results, people tend to prefer those considered as certain or almost certain, when compared to others with an higher expected value, but not certain. Many other important framing effects are derived from the certainty effect, e.g.: aversion to certain lost, bringing people to secure choices, even if less economically worthy.

The reflect effect happens when turning the previous situation upside down, i.e.: instead of considering the probability of a positive outcome, that of a negative outcome is indeed considered. While when considering positive situations the individuals are risk-averse, they tend to become risk-seeking when all the alternatives seem to be negative (they often choose the least certain ones, even when apparently worst, possibly hoping that they will turn less negative). Isolation effect is the tendency to disregard the common elements among more possible choices, just focusing on the differential elements. This can lead to errors, since apparently equal aspects of different situations can be indeed different: there could be several ways to decompose a real problem, and many situations are indeed complex, thus stressing the interaction among the parts.

In the following paragraphs, Reinforcement Learning (RL) is formally described as a technique for learning in artificial agents, and then a new approach is introduced, with the aim of injecting some of the analyzed perception errors in the existing algorithms.

3 REINFORCEMENT LEARNING

Learning from reinforcements has received substantial attention as a mechanism for robots and other computer systems to learn tasks without external supervision. The agent typically receives a positive payoff from the environment after it achieves a particular goal, or, even simpler, when a performed action gives good results. In the same way, it receives a negative (or null) payoff when the action (or set of actions) performed brings to a failure. By performing many actions overtime (trial and error technique), the agents can compute the expected values (EV) for each action. According to Sutton and Barto (1998) this paradigm turns values into behavioural patterns; in fact, each time an action will need to be performed, its EV, will be considered and compared with the EVs of other possible actions, thus determining the agent's behaviour, which is not wired into the agent itself, but self adapting to the system in which it operates.

Most RL algorithms are about coordination in multi agents systems, defined as the ability of two or more agents to jointly reach a consensus over which actions to perform in an environment. In these cases, an algorithm derived from the classic Q-Learning technique (Watkins, 1989) can be used. The EV for an action – $EV(a)$ – is simply updated every time the action is performed, according to the following, reported by Kapetanakis and Kundenko (2004):

$$EV(a) \leftarrow EV(a) + \lambda(p - EV(a)) \quad (1)$$

Where $0 < \lambda < 1$ is the learning rate and p is the payoff received every time that action a is performed. This is particularly suitable for simulating multi stage games (Fudenberg and Levine 1998), in which agents must coordinate to get the highest possible aggregate payoff. For example, given a scenario with two agents (A and B), each of them endowed with two possible actions a_1, a_2 and b_1, b_2 respectively, the agents will get a payoff, based on a payoff matrix, according to the combination of performed actions. For instance, if a_1 and b_1 are performed at the same time, both agents will get a positive payoff, while for all the other combinations they will receive a negative reward. ABS applied to social system is not necessarily about coordination among agents and convergence to the optimal behaviour, especially when focusing on the aggregate level; it's often more important to have a realistic behaviour for the agents, in the sense that it should replicate, as much as possible, that of real individuals. The aforementioned RL algorithm analytically evaluates the best action based on historical data, i.e.: the EV of the action itself, over time. This makes the agent perfectly rational, since it will evaluate, every time he has to perform it, the best possible action found till then. If this is very useful for computational problems where convergence to an optimal behaviour is crucial, it's not realistic when applied to a simulation of a social system. In this kind of systems, learning should keep into account the human factor, in the shape of perception biases, distortions, preferences, prejudice, external influences and so on. Traditional learning models represent all the agents in the same way – i.e.: as focused and rational agents; since they ignore many other aspects of behaviour that influence how humans make decisions in real life, these models do not accurately represent real users in social contexts.

4 EGO BIASED LEARNING

Even if preferences can be modified according to the outcome of past actions (and this is well represented by the RL algorithms described before), humans keep an emotional part driving them to prefer a certain action over another one, as described in paragraph 2. That's the point behind learning: human aren't machines, able to analytically evaluate all the aspects of a problem and, above all, the payoff deriving from an action is filtered by their own perception bias. There's more than just a self-updating function for evaluating actions and in the following a formal reinforcement learning method is presented which keeps into consideration a possible bias towards a particular action, which, to some extents, make it preferable to another one that has analytically proven better through the trial and error period. As a very first step towards that direction, *Ego Biased Learning,* introduced by Marco Remondino*,* allows to keep personal factor into consideration, when applying a RL paradigm, by modelling two perception errors described in paragraph 2: Anchoring and Affect Heuristics.

4.1 Dualistic Case

In the first formulation, a dualistic action selection is considered, i.e.: $A(a_1, a_2)$. By applying the formal reinforcement learning technique described in equation (1) an agent is able to have the expected value for the action it performed. We imagine two different categories of agents (α_1, α_2): one biased towards action a_1 and the other one biased towards action a_2. For each category, a constant is introduced $(0 < K_1, K_2 < 1)$, defining the propensity for the given action, used to evaluate $\overline{EV(a_1)}$ and $\overline{EV(a_2)}$ which is the expected value of the action, corrected by the bias. For the category of agents biased towards action a_1 we have that:

$$\alpha_1\colon \begin{cases} \overline{EV(a_1)} = EV(a_1) + (|EV(a_1)| * K_1) \\ \overline{EV(a_2)} = EV(a_2) - (|EV(a_2)| * K_1) \end{cases} \tag{2}$$

In this way, K_1 represents the propensity for the first category of agents towards action a_1 and acts as a percentage increasing the analytically computed $EV(a_1)$ and decreasing $EV(a_2)$. At the same way, K_2 represents the propensity for the second category of agents towards action a_2 and acts on the expected value of the two possible actions as before:

$$\alpha_2\colon \begin{cases} \overline{EV(a_1)} = EV(a_1) - (|EV(a_1)| * K_2) \\ \overline{EV(a_2)} = EV(a_2) + (|EV(a_2)| * K_2) \end{cases} \tag{3}$$

The constant K acts like a "friction" for the EV function; after calculating the objective $EV(a_i)$ it increments it of a percentage, if a_i is the action for which the agent has a positive bias, or decrements it, if a_i is the action for which the agent has a negative bias. In this way, the agent α_1 will perform action a_1 (instead of a_2) even if $EV(a_1) < EV(a_2)$, as long as $\overline{EV(a_1)}$ is not less than $\overline{EV(a_2)}$. In particular, by analytically solving the following:

$$EV(a_1) + (|EV(a_1)| * K_1) \geq EV(a_2) - (|EV(a_2)| * K_1) \tag{4}$$

We have that agent α_1(biased towards action a_1) will perform $a_1$as long as:

$$EV(a_1) \geq EV(a_2) * \frac{1 - K_1}{1 + K_1} \tag{5}$$

Equation number 5 applies when both $EV(a_1)$ and $EV(a_2)$ are positive values. If $EV(a_1)$ is positive and $EV(a_2)$ is negative, then a_1 will obviously be performed (being this a sub-case of equation 5), while if $EV(a_2)$ is positive and $EV(a_1)$ is negative, then a_2 will be performed, since even if biased, it wouldn't make any sense for an agent to perform an action that proved even harmful (that's why it went down to a negative value). If $\overline{EV(a_1)} = \overline{EV(a_2)}$, by definition, the performed action will be the favorite one, i.e.: the one towards which the agent has a positive bias.

In order to give a numeric example, if $EV(a_1) = 50$ and $K_1 = 0.2$ then a_1 will be performed by agent α_1 till $EV(a_2) > 75$. This friction gets even stronger for higher K values; for example, with a $K_1 = 0.5$, a_1 will be performed till $EV(a_2) > 150$ and so on.

By increasing the value of K_1, the positive values of $EV(a_1)$ turns into higher and higher values of $\overline{EV(a_1)}$. At the same time, a negative value of $EV(a_1)$ gets less and less negative by increasing K_1, while never turning into a positive value (at most, when K_1, $\overline{EV(a_1)}$ gets equal to 0 for every $EV(a_1) < 0$). For example, with $K_1 = 0.1$, $\overline{EV(a_1)}$ is 10% higher than $EV(a_1)$.

Since a_2 is the action towards which the agent α_1 has a negative bias, it's possible to notice that the resulting $\overline{EV(a_2)}$ is always lower (or equal, in case they are both 0) than the original $EV(a_2)$ calculated according to equation 1. In particular, higher K_1 corresponds to more bias (larger distance among the objective expected value), exactly opposite as it was before for action a_2. Note that for a $K_1 = 1$ (i.e.: maximum bias) $\overline{EV(a_2)}$ never gets past zero, so that a_2 is performed if and only if $EV(a_1)$ - and hence $\overline{EV(a_1)}$ - is less than zero.

4.2 General Cases

The first general case (more than two possible actions and more than two categories of agents) is actually a strict super-case of the one formalized in 4.1. Each agent is endowed with an evaluation biased function derived from equations (2) and (3). Be $\alpha(\alpha_1, \alpha_2, \ldots, \alpha_n)$ the set of agents, and $A(a_1, a_2, \ldots, a_m)$ the set of possible actions to be performed, then the specific agent α_k, with a positive bias for action a_h will feature such a biased evaluation function:

$$\alpha_k: \begin{cases} \overline{EV(a_1)} = EV(a_1) - (|EV(a_1)| * K_1) \\ \ldots \\ \overline{EV(a_{h-1})} = EV(a_{h-1}) - (|EV(a_{h-1})| * K_1) \\ \overline{EV(a_h)} = EV(a_h) + (|EV(a_h)| * K_1) \\ \overline{EV(a_{h+1})} = EV(a_{h+1}) - (|EV(a_{h+1})| * K_1) \\ \ldots \\ \overline{EV(a_m)} = EV(a_m) - (|EV(a_m)| * K_1) \end{cases} \tag{6}$$

This applies to each agent, of course by changing the specific equation corresponding to her specific positive bias. Even more general, an agent could have a positive bias towards more than one action; for example, if agent α_5 has a positive bias for actions a_1 and a_2 and a negative bias for all the others, the resulting formalism is equation (7) and, in the most general case, for each $\overline{\mathrm{EV}(a_i)}$ we have the equation (8). In case that two or more $\overline{\mathrm{EV(a)}}$ have the same value, the agent will perform the action towards which it has a positive bias; in the case explored by equation (7), in which the agent has the same positive bias towards more than one action, then the choice among which action to perform, under the same $\overline{\mathrm{EV(a)}}$, is managed in various ways (e.g.: randomly).

$$\alpha_5: \begin{cases} \overline{EV(a_1)} = EV(a_1) + (|EV(a_1)| * K_1) \\ \overline{EV(a_2)} = EV(a_2) + (|EV(a_2)| * K_1) \\ \overline{EV(a_3)} = EV(a_3) - (|EV(a_3)| * K_1) \\ \ldots \\ \overline{EV(a_m)} = EV(a_m) - (|EV(a_m)| * K_1) \end{cases} \tag{7}$$

$$\overline{EV(a_i)} = EV(a_i) \mp (|EV(a_i)| * K_{\mathrm{i}}) \tag{8}$$

As a last general case, the agents could be a different positive/negative propensity towards different actions. In this case, the K variable to be used won't be the same for all the equations regarding an individual agent. For example, given a set of $K(K_1, K_2, \ldots, K_n)$ and a set of actions $A(a_1, a_2, \ldots, a_m)$, for each agent (α_k) we have:

$$\alpha_k: \begin{cases} \overline{EV(a_1)} = EV(a_1) \mp (|EV(a_1)| * K_1) \\ \ldots \\ \overline{EV(a_m)} = EV(a_m) \mp (|EV(a_m)| * K_{\mathrm{n}}) \end{cases} \tag{9}$$

Besides being a fixed parameter, K could be a stochastic value, e.g.: given a mean and a variance.

5 FUTURE DEVELOPMENTS

Many of the described cognitive biases are derived from the fact that humans are social beings. While individual preferences are very important as a bias factor for learning and action selection, when dealing with social systems, in which many entities operate at the same time and are usually connected over a network, other factors should be kept into consideration. In particular, the preferences of other individuals with which a specific agent is in touch can affect choices, by modifying the objective perception mechanism described in equation 1. Once again, if the goal is that of representing agents mimicking human behaviour, then it's not realistic to consider perfect perception of the payoffs deriving from past actions. Fragaszy and Visalberghi

(2001) agree that socially biased learning is widespread in the animal kingdom and important in behavioural biology and in evolution. It's important to distinguish between imitation and socially biased learning; the former is limited to imitating the behaviour of another individual (possible with some minor changes), the latter is referred to modifying the possessed behaviour after the observation of others' behaviours. While imitation is passive and mechanical, social learning supposes a form of intelligence in selecting how to modify the past behaviour, taking into account others' experience.

Box (1984) defines socially biased learning as: *a change in behaviour contingent upon a change in cognitive state associated with experience that is aided by exposure to the activities of social companions*. From this definition, it's evident that the first part is already taken into account by RL methods (equation 1) and by the ego biased learning proposed in the previous sections. What is still lacking is the bias coming from social companions. They should be able to perceive the outcome that other agents had from the actions they performed. Not all the agents are perceived in the same way; some of them can be considered more reliable, and thus their experience will be more valuable as a bias. Other cognitive distortions analyzed by BF will thus be formally incorporated in RL algorithms.

6 CONCLUSIONS

Many evidences coming from the real world prove that individuals are not completely rational; their perceptions are biased and distorted by emotions, preferences and so on. Behavioural Finance is the discipline that studies and formalizes these biased behaviours. In order to endow artificial agents with a realistic behaviour, in this work a formal method for action selection is introduced, called Ego Biased Learning. It's based on one step QL algorithm (equation 1), but it takes into account individual preference for one or more actions, thus being a very first step in formalizing human distortions in a RL algorithm. This method is designed to be used in simulation of social systems employing MAS, where many entities interact in the same environment and must take some actions at each time-step. In particular, traditional methods do not take into account human factor, in the form of personal inclination towards different strategies, and consider the agents as totally rational and able to modify their behaviour based on an analytical payoff function derived from the performed actions.

Ego Biased Learning is first presented in the most simple case, in which only two categories of agents are involved, and only two actions are possible. That's useful to show the basic equations defining the paradigm and to explore the results, when varying the parameters. After that, some general cases are faced, i.e.: where an arbitrary number of agents' categories is allowed, along with an equally discretionary number of actions. There can be many sub-cases for this situations, e.g.: just one action is preferred, and the others are disadvantaged, or an agent has the same bias towards more actions, or in the most general situation, each action can have a positive or negative bias, for an agent. This technique represents two of the most common perception errors studied by BF: Anchoring and Affect Heuristics. In future works, other biases will be introduced in the learning mechanism, and formally described.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the key support of Prof. Anna Maria Bruno, full professor of *Economia e Gestione delle Imprese*, University of Turin.

REFERENCES

Box, H. O., 1984. *Primate Behaviour and Social Ecology*. London: Chapman and Hall.

Chen M. K., 2008. Rationalization and Cognitive Dissonance: do Choices Affect or Reflect Preferences? *Cowles Foundation Discussion Paper No. 1669*

Mataric M. J., 2004. Reward Functions for Accelerated Learning. In *Proceedings of the Eleventh International Conference on Machine Learning.*

Mataric, M. J., 1997. Reinforcement Learning in the Multi-Robot domain. *Autonomous Robots*, 4(1).

Fudenberg, D., and Levine, D. K. 1998. *The Theory of Learning in Games*. Cambridge, MA: MIT Press

Fragaszy, D. and Visalberghi, E., 2001. Recognizing a swan: Socially-biased learning. *Psychologia, 44.*

Powers R. and Shoham Y., 2005. New criteria and a new algorithm for learning in multi-agent systems. *In Proceedings of NIPS.*

Sharot T., De Martino B., Dolan R.J., 2009. How Choice Reveals and Shapes Expected Hedonic Outcome. *The Journal of Neuroscience, 29(12):3760-3765*

Sutton, R. S. and Barto A. G., 1998. Reinforcement Learning: An Introduction. *MIT Press, Cambridge, MA. A Bradford Book*

Watkins, C. J. C. H. 1989. Learning from delayed rewards. *PhD thesis*, Psychology Dep. Univ. of Cambridge.