The Geometry of Desire

Luis Antunes
GUESS, LabMAg
University of Lisbon, Portugal
xarax@fc.ul.pt

Davide Nunes
GUESS, LabMAg
University of Lisbon, Portugal
dnunes@fc.ul.pt

Helder Coelho
GUESS, LabMAg
University of Lisbon, Portugal
hcoelho@di.fc.ul.pt

ABSTRACT

In the BDI paradigm, much attention was devoted to beliefs, intentions, choice and commitment, whereas desire has traditionally been seen as given. However, desire is the key connection to the agents' creator, and the ultimate source of behaviour. Desires are allowed to be incoherent, irrational, or at least a-rational. Agent environments establish a motivational context for agents to act upon. Agent societies are never truly autonomous. We argue that pre-designed utility-based behaviour search strategies not only hinder the adaptability of an agent but also prevent the emergence of novel social behaviour. In this paper, we propose a new model of desire acquisition and evolution. Agents continuously adapt their desires by means of both their intrinsic motivations, as well as a mimetic mechanism inspired in René Girard's theory. Agents acquire new goals not through fitness or novelty but out of mechanisms such as envy, imitation and competition. To achieve their goals, agents have to sometimes discard them and just overcome their neighbours.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent Agents*

Keywords

Agent theories and models::Cognitive aspects; Agent societies and societal issues::Artificial social systems; Agent-based simulation::Emergent behaviour

1. INTRODUCTION

In real and artificial societies, agents make their choices using their own criteria, often informed by complex concepts and mechanisms such as utility [19], other times following more esoteric and yet realistic rules of behaviour such as imitation [3], evolution [8] or value sharing [1].

The Belief-Desire-Intention (BDI) agent architecture [16] emanates from and represents a philosophical stance allowing the characterisation of agents in terms of mental qualities easily recognisable by other agents. Several rigorous models and techniques supplied reasoning machinery designed to deal with such qualities, and build credible scripts for themselves and interpret the actions of others. However, the behaviours that are derived from those scripts are very often unsatisfactory, especially when we hope that our model scales-up to resemble what happens in real societies, hoping to derive some understanding and perhaps useful policies from those models.

BDI used to be an architecture designed for decision and action. For some time it included an associated modal logic, which seemed to constrain the design of agent minds into a logic-based paradigm [16]. But time corrected that tendency and BDI is now seen as a terminological and conceptual basis that provides common ground (and grounding) for a dialogue between theorists and practitioners in the multiagent system (MAS) communities, but also outside, allowing to engage other scientific areas in the discussion, such as economists, other social scientists, as well as physicists, psychologists, philosophers, neuroscientists, etc.

Watching BDI as a general framework for agent mentality, we notice the key role played by intentions as a link between the agents' beliefs (what the agent knows) and desires (what the agent aims for). Several authors (notably Castelfranchi [4]) noted that desires and intentions really belong to the same mental category (pro-attitudes). Intentions are an especially constrained subset of desires, and represent what the agents will really work towards achieving. Intentions come out of desires as a result of the agents deliberation, and are managed through special mechanisms abiding to rationality principles. What cannot be derived from rational principles, whichever rationality definition we pick, is the set of desires the agents aspire to. Given that most of what remains is left to the agents' (and its designer) discretion, it is in the desire set that we can (should) locate the agents' ultimate goals, which can justify (and generate) their behaviour.

The source of desires is key to determine behaviour, both individually and collectively. In most BDI approaches, desires are given data for the problems to be addressed. But what happens when we want to confer true autonomy to agents? What happens when, as often in exploratory simulation, we don't know what exactly we are after in an experiment, both as agent designers and as interrogative scientists? What happens when we are after novelty, and seek to discover, instead of facing well-defined problems which involve search of a solution? As Ken Stanley puts it [13], sometimes we have to abandon our goals in order to achieve them.

Teleological behaviour [17] has long been the basis of multi-agent system development. Simon [18] stated that "People

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

have reasons for what they do", but what is left unsaid is that the others might not know those reasons, and moreover, people themselves might not be aware of what those reasons are, much less of their implications. Traditionally, it is possible to ground desires in some part of the agents design, such as their physiology. For instance, a person may be looking for food, much as a robot may be trying to locate a power source [8].

But a recent view on an old theory suggests other paths of inspiration and motivation. As soon as people have their basic needs satisfied, other motivations come to the front row. Human-nature interactions, observed in literary works (Cervantes, Stendhal, Flaubert, Proust, Dostoyevsky), were critically analysed by Girard in order to discover the mechanisms played by desires. Girard's theory of mimetic desire claims that motivations have much more to do with less obvious drives, such as envy, jealousy, but also admiration, identification, or membership feelings, leading to behaviours such as imitation, or the adoption of other people's desires. Girard further argues that desire is not linear, rather triangular, demanding for another participating actor, and this yields even further complexity, as it arises phenomena of transference, substitution, mediation, rivalry. Surely, other social sources of motivation arise from power, imposition, influence, marketing, education, etc.

Social science theories have time and again provided good sources of metaphors to be used in MAS. In this paper, we put forward the hypothesis that on the basis of some divergences between theory and reality, there could simply be a trivial misconception about what people ultimately want. For this reason alone, it is worth to consider whether the basic grounding of people's motivations might be far different from what has been modelled so far. Desires are such a small, basic, pervasive component of the agents' mental design, that the implications of such a move might lead to a true revolution in MAS design and applicability.

2. MIMETIC DESIRE

Scientific accounts of human motivations, and their social consequences, have always represented a challenge for researchers, because of the non-trivial nature of the micro-macro links. Individual behaviour aggregates into collective behaviour in a complex manner, and reciprocally, the society influences each individual in specific, varied and unpredictable ways (in particular, the individual perception of collective phenomena is masked by mediation, neighbourhood-specific filters, etc.). Local phenomena that cause perplexity can grow to huge proportions when they collectively build up erroneous world-scale accounts of reality. For instance, the gap between real wealth and circulating wealth does not seem to stop increasing, causing more and more problems for institutions, individual companies, and people. Micro-changes in one's individual decisions, springing from one's own rationality, can determine huge differences in global behaviour. However, there are alternatives to the top-down analytical approach that resolves every person as a strategic, perfect thinker and then proceeds to integrate their decisions in some additive way.

We can go back as far as the XVIII century to get a new vision on human motivation. Adam Smith [20] noted that basic needs could be satisfied by a frugal life. What causes people to seek and accumulate wealth is desire. While need stops with satisfaction, desire is reborn stronger every time it is satisfied. Easterlin's paradox [7] strengthened this view when he noted that beyond a certain level of welfare, the increase of economic development and global wealth does not increase the amount of happiness reported by populations. Apparently, positive achievements count far less for individuals than negative comparisons with his/her neighbours.

This aspiration for recognition has never been considered as a proper motivator for agents in multi-agent societies. The idea is to ground teleologic behaviour not on needs, but on desires, and consider their acquisition as a collectively acquired process. This process does not start from scratch, and this fact may have prevented computer scientists from adopting this principle. Artificial societies do start, whereas human and animal societies never do.

Girard [9] decoded the complex process of imitation, cooperation, competition and self-deception that leads to the formation of desires. In a nutshell, we borrow desires from others. Not only we are not autonomous, but we need a mediator to acquire the desire from, and sometimes to compete against, in the pursuit of the desired object. So, imitation is the basic mechanism by which behaviour emerges, and this makes it impossible to start from a null situation. The power of imitation has been claimed and emphasised time and again in MAS literature [3]. However, never was imitation taken as far back in the source of behaviour. From an evolutionary perspective, imitation seems to have been 'designed' by natural selection to extract useful information from the social world [10] and supply the basis for emergent cooperative behaviour. Mimetic behaviour might not lead to an exact replication of the observed behaviour but serves as the building block to acquire capabilities that might grant what is an apparent autonomy from the observed.

In traditional models, desire is a linear concept: there is the subject and the object of desire. With Girard, we realise that a third party is involved, and not a passive one. Desire becomes "triangular," as we acquire them from the models, we, either consciously or else unwittingly, admire and imitate [6]. The mediator can have multiple roles in the behaviour he/she induces.

> The triangularity of desire is a simple notion, but it has broad and complex implications. To begin with, it explains the obvious but otherwise perplexing fact that desire may not only cause rivalry – my mediator automatically becomes my rival since we desire the same object –, but also depend on it – to the point that without rivalry, desire itself threatens to languish. [6].

3. MENTAL MECHANICS OF DESIRE ACQUISITION

Triangular desire endows actors with several concomitant roles and associated drives. There is empathy between the subject and the mediator, since they share a passion for the same object. This empathy does not go without some rivalry, because often the object can be obtained by only one of the coveters.

However, the rivalry between subject and mediator is not acknowledged by either actor. A very common case is double mediation: A and B imitate each other reciprocally. A is anxious about B's desire, which alone designates a target for his own desire. Some ephemeral and random sign makes A believe that B has designs on object O. Rushing to

get there first, he thereby signals to his alter ego the stakes of rivalry. When B in turn imitates A's desire the starting illusion became reality. The first to imagine the other's desire thus seems not to have been imagining at all. This is a particularly interesting case of a self-fulfilling prophecy [6].

On a deeper level, mimetic desire does not aim at possession, but on being. Girard call this "metaphysical desire," one feels that possession yields a greater degree of ontological sufficiency. This sufficiency, autonomy, or independence, confers enhanced value to the subject, because he seems to be above and superior, indifferent to the needs of the rest, which renders a tremendous power of attraction. A consequence of this mechanism is today's publicity: models have perfect self-sufficiency. Their message is both clear and self-contradictory: "imitate me, and you will become autonomous, i.e. you will escape the infernal realm of metaphysical desire." [6].

4. DESIRE AND VALUE

In [5], Jean-Pierre Dupuy considers markets as tentative ultimate optimisers of collective behaviour. Neoclassical economics [11] has always defended such mechanisms as producing desirable equilibria, in terms of the stability of the economic landscape, but also, to a certain extent, of fairness for the individual actors there included. In MAS, the tension between social and individual welfare has also been present, but the proposed solutions always seemed quite artificial, as they imposed a dual weight on the agent's motivational drives, with social motivations being prevalent over the individual, something that severely undermined the agent's autonomy, and especially, free will (cf. among others [12]).

Dupuy notes that negative retroactions are essential to force agents into returning to equilibrium behaviours, and this is known to be a key pre-requisite for utility theory. However, utility theory has been discredited since Herbert Simon's first works [19], and even so, hard to replace in concrete agent designs. Dupuy emphasises – as did Brian Arthur [2] – the importance of *positive* retroactions, and claims that these are notably produced by imitation. Imitation causes the emergence of an external objectivity from a closed system of actors: first by a game of mirrors, with generalised behaviour adoption, later by stabilising the emerged object and forgetting of its genesis. The concept of equilibrium is hence not adapted to characterise the attractors of this mimetic dynamics. These attractors are initially arbitrary and undetermined, and then acquire value of collective action.

What mimetic adoption of desires can provide for agent societies is thus not a driving force towards equilibrium or optimality. This adoption mechanism does provide a derivation from individual self-interest, that in its diversity can represent an external source of agent heterogeneity, of variability. Scientific implications of these designs include novelty search or wide exploration of design spaces. Some reconciliation with game theory is possible, since the drives (envy, admiration, independence, subservience) that lead mimetic desire contain contradictions (cycles of comparison, rivalry, imitation) with which real people seem happy to live. For instance, comparing theoretical and real experiments with the prisoner's dilemma or centipede games, we can explain the results a posteriori, but very seldom generate them, since the drives used are unpredictable. The trivial action of buying a stranger a cup of coffee is much more easily explained by the social construction of mimetic action than by some intricate notion of utility-based intermittent self satisfaction that very hardly anyone would use to make the decision.

5. EVOLUTIONARY EXPLORATIONS OF DESIGN SPACE

The research programme to uncover the implications of this new approach on motivation grounding is guided by the purpose of providing a more realistic account of individual choice, not focused on optimising given measures. As application of this enhanced – but not wildly free – autonomy, we aim at designing and running more trustworthy simulations, possibly to rehearse and guide the deployment of policies, and in general to improve the focus and reliability of artificial social systems involving self-motivated agents, inhabiting environments that are possibly dynamical, uncertain, with unknown, or impossible to uncover, laws.

In such a setting, agents loose or dramatically reduce their potential for discovery when all their goals are previously given. We can under-specify those goals and provide agents with learning abilities, but in that case we should strive to avoid overfitting the training data. Ultimately, we will always want to ensure that our agents, even if symbolically defined, have the ability to evolve and adapt (to aprioristically impossible to know conditions, including goals).

So, we opted to design our agents with in an evolutionary framework, basing their ability to change in the acquisition of mimetic desires from their neighbours. Each agent is provided with a dynamic behaviour defined by a policy, from a collection of internally kept policies, and directed to achieving a goal. The evaluation of the agent's behaviour is performed with independent measures (of which proximity to the goal *may* be one criterion). A meta-policy rules the acquisition of new goals and policies from their neighbours.

We use Cartesian Genetic Programming [14] to allow the chosen policies to recursively and reflectively build on themselves, given the adequate operators. So, the potential for incremental complexity is unlimited, admitting not only the adaptive exploration of the agents response to the several fitness functions that the environment will successively offer (including some concomitant, conflicting, contradictory, etc.), but also the exploration of design space for the agents themselves, including incrementally complex agent designs, but also collective organisations or institutions the agents can themselves develop.

Evolutionary approaches, such as genetic programming, are typically used for off-line supervised learning in simulated environments. In this case, synthesising a successful adaptive behaviour is as dependent on the scenario modeller, just as plan-based approaches are on the agent designer. Notwithstanding this fact, these techniques (although still under-explored) have been successfully used in on-line learning scenarios [15]. Instead of focusing on a pure utility-based search in the agent behaviour design space, we use the principles of mimetic desire as the driving force behind the behaviour search.

Simulating the social behaviour search, driven by mimetic desire, is then a matter of providing the following components: *ad hoc* restrictions that are dependent on the problem, they shape the behaviour, that is not rooted in desire, but can be perceived as such; an imitation mechanism that steers an agent behaviour towards a mediator, a target for

imitation; and a competition mechanism in which one agent tries to overcome its mediator. We also need criteria for imitation, competition or cooperation, to dictate which driving force is acting on the behaviour search.

Using fitness landscapes, one can model *ad hoc restrictions* (unknown to the agents) that act like behaviour reward functions, much like what happens in reinforcement learning problems. In a preliminary exploration, we deployed our agents in partially observable fitness landscapes where they are evaluated while they operate, in effect, changing the evaluation conditions each time they test a behaviour policy. We observed that contextual information is decisive for the success of adaptation. If an agent adopts a policy that proves highly rewarding in some situation, that success may be achieved for contextual reasons that might not replicate in a different situation. In dynamic environments the problem is even more difficult, eventually the context changes and a successful policy must be adapted or even radically changed to face the new scenario. We are now exploring how decisive social drives such as imitation and competition can mitigate the problem of context.

Agents with behaviours performing badly due to unfortunate evaluation conditions benefit from imitating a mediator that evolved in a more favourable region of the search space. Imitating does not imply that agents are capable of reproducing the exact behaviour policy of the mediator. Instead, an agent evolves a behaviour with a similar phenotype triggered by different environmental conditions.

Competition is implemented by evolving one's behaviour towards being better than a mediator under the same target of competition (performing better in relation to some environmental restriction, for instance). This can pressure some agents to explore different aspects of their behaviour motivated (temporarily) by the desire to overcome their neighbours.

6. CONCLUSION

In this paper, we presented the seeds for a new approach that considers the grounding of adaptive emergent behaviour of MAS, not in the traditional utility-based and design-depended approaches but on the origin, adoption and evolution of desires.

These mental mechanics proposed by René Girard [9] are far reached in what concerns the fabrics of society. Especially important for our case is the *ad hoc* genesis of many desires and consequent behaviours. General laws are difficult to pin down, and they may be sometimes deceitful, promising to deliver what experience later proves to be impossible. The general aims of the societies we try do design are then difficult to grasp, model and program: real phenomena can always save some surprise to defeat pre-designed behaviours.

We argue that the adoption of imitation as a source of motivation can be instrumental to the evolution and adoption of desires that lead agent societies to emergent adaptive behaviour, without ignoring their contextual goals (exploring and exploiting their capabilities in relation to their environment).

7. REFERENCES

[1] L. Antunes, J. Faria, and H. Coelho. Improving choice mechanisms within the BVG architecture. In *Intelligent Agents VII, Proc. of ATAL2000*, volume 1986 of *LNAI*. Springer, 2000.

[2] W. B. Arthur. *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press, Ann Arbor, 1994.

[3] E. Borenstein and E. Ruppin. Enhancing autonomous agents evolution with learning by imitation. In *Second International Symposium on Imitation in Animals and Artifacts*, 2003.

[4] C. Castelfranchi, R. Conte, Y. Demazeau, and J. Sichman. A social reasoning mechanism based on dependence networks. *Procs. of ECAI*, 94, 1994.

[5] J.-P. Dupuy. *Pour un catastrophisme éclairé*. Seuil, Paris, 2004.

[6] J.-P. Dupuy. The problem of desire, violence, and religion in philosophy, literature, and cinema–a critical view on René Girard's mimetic and sacrificial theory, 2011.

[7] R. A. Easterlin. Does economic growth improve the human lot? In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*. Academic Press, 1974.

[8] S. Elfwing, E. Uchibe, K. Doya, and H. I. Christensen. Biologically inspired embodied evolution of survival. In *IEEE Congress on Evolutionary Computation*, 2005.

[9] R. Girard. *Deceit, Desire and the Novel: Self and Other in Literary Structure*. The Johns Hopkins University Press, 1965.

[10] J. Henrich. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53(1):3–35, 2004.

[11] H. M. Hochman and J. D. Rodgers. Pareto optimal redistribution. *The American Economic Review*, 59(4):542–557, 1969.

[12] N. R. Jennings and J. R. Campos. Towards a social level characterisation of socially responsible agents. *IEEE Proceedings on Software Engineering*, 144(1):11–25, 1997.

[13] J. Lehman and K. O. Stanley. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation*, 19(2):189–223, 2011.

[14] J. F. Miller and P. Thomson. Cartesian genetic programming. In *Genetic Programming*, volume 1802 of *LNCS*. Springer, 2000.

[15] P. Nordin and W. Banzhaf. An on-line method to evolve behavior and to control a miniature robot in real time with genetic programming. *Adaptive Behavior*, 5(2):107–140, 1997.

[16] A. S. Rao and M. P. Georgeff. Modeling agents within a bdi-architecture. In *Proc. KR'91*. Morgan Kaufmann, 1991.

[17] A. Rosenblueth, N. Wiener, and J. Bigelow. Behavior, purpose and teleology. *Philosophy of science*, 10(1):18–24, 1943.

[18] H. Simon. Rationality in psychology and economics. In R. M. Hogarth and M. W. Reder, editors, *Rational choice: the Contrast Between Economics and Psychology*. University of Chicago Press, 1987.

[19] H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99–118, 1955.

[20] A. Smith. *Theory of Moral Sentiments*. A. Millar, London, 1759.