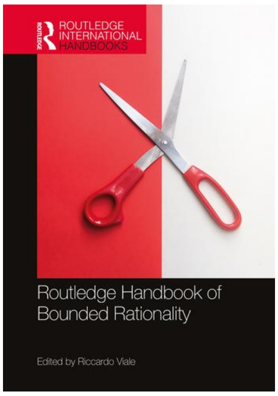


# Routledge Handbook of Bounded Rationality

Riccardo Viale

# Beyond "bounded rationality"

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315658353-38
Giovanni Dosi, Marco Faillo, Luigi Marengo
**Published online on: 29 Oct 2020**

# 33

# BEYOND "BOUNDED RATIONALITY"

## Behaviours and learning in complex evolving worlds

*Giovanni Dosi, Marco Faillo, and Luigi Marengo*

### Introduction

Let us begin by the very notion of bounded rationality. Should we assume that there is an "unbounded" rationality as a benchmark? Should one start, in order to describe and interpret human behaviour from a model which assumes that we, human beings, have complete and well-defined knowledge of our preferences, all possible states of the world, all possible actions (our "technologies"), and the mappings among them?[1]

Savage was extremely careful in limiting his choice-theoretic exercise to the normative domain and to "small worlds", i.e., stationary and isolated portions of the world wherein decision makers know the full set of possible events and can attribute probabilities to them (Savage, 1954).

Jumping from this normative small world domain to the descriptive framework where one builds theory of human behaviour in a complex world, characterized by radical uncertainty, a multiplicity of interactive agents, and persistent endogenous innovation – we suggest – is deeply misleading. Having "Olympic rationality" as a benchmark is like starting from the thermodynamic equilibrium death with full entropy in order to interpret the biological world!

Rather, the question should be: how do human agents and organizations thereof behave in complex and changing environments? Answering this question, we suggest, entails also a significant departure from what is now accepted as behavioural economics, often meant as the analysis of more or less significant deviations – called "biases" – from the "Olympic rationality". On the contrary, we suggest, human beings and human organizations behave quite distinctively from the prescriptive model derived from the axioms of rationality.

As is well known, the standard decision-theoretic model depicts agency (and, *in primis*, economic agency) as a problem of choice where rational actors select, among a set of alternative courses of action, the one which will produce (in expectation) the maximum outcome as measured against some utility yardstick. In that, agents are postulated to know the entire set of possible events of "nature", all possible actions which are open to them, and all notional outcomes of the mapping between actions and events (or at least come to know them after some learning process). Clearly, these are quite demanding assumptions on knowledge embodied into or accessible to the agents – which hardly apply to complex and changing environments. In fact, they *cannot* apply almost by definition in all environments where innovations of some kind are allowed to occur – irrespective of whether they relate to technologies, behavioural repertoires or organizational arrangements. If an innovation is truly an innovation, it could not have been in the set of events that all agents were able to contemplate before the innovation actually occurred.

Moreover, equally demanding are the implicit assumptions concerning the *procedural rationality* involved in the decision process.

As a paradigmatic illustration, take the usual decision-theoretic sequence leading from (1) representation/"understanding" of the environment (conditional on whatever available "information"), to (2) evaluation/judgement; (3) choice; (4) actions, and, ultimately, (5) consequences – determined, e.g., by the stochastic pairing of actions and "events of nature" and/or actions by other agents.

In order for this "rationalist" view to hold, at least two assumptions are crucial. First, the linearity of that sequence must strictly hold. That is, one must rule out the possibility of reversing, so to speak, the procedural sequence. For example, one cannot have preferences and representations which adapt to an action that has already been undertaken. and, likewise, one must assume that consequences do not influence preferences (i.e. preferences are not endogenous).

Second, at each step of the process agents must be endowed with, or able to build, the appropriate algorithm in order to tackle the task at hand – whether it is representing the environment, evaluating alternatives or choosing courses of action, etc.

There are, indeed, a few rather compelling reasons why these assumptions might be a misleading starting point for any *positive* theory of learning and choice.

Human agents tackle every day, with varying degrees of success, highly complex and "hard" (in the sense of computability theory) problems with their highly limited computational capabilities. Cognitive sciences have made impressive progress in the recent decades in understanding how we do that. We are bad in processing information, we cannot handle more than a very limited number of the overwhelming number of interdependencies the characterize our world, but nevertheless we go along, sometimes decently well, with simple but useful representations and simple but effective heuristics. As suggested by Gigerenzer and his group, such heuristics are not the outcome of our biases, although they may sometimes produce them (Gigerenzer et al., 1999). On the contrary it is their very simplicity which makes them "smart", and generally well adapted to the complex and fast-changing world in which we live. They require simple representations and neglect part of the available information, that is, they radically depart from that model of rationality which assumes correct representation and unlimited information processing capabilities, but, on the contrary, excel in simplicity, frugality, adaptability, i.e., features with are not even considered in the rational choice framework. "Olympic rationality", in fact, implies the availability of some inferential machinery able to extract the "correct" information from environmental signals, Bayes rule being one of them, and possibly also the most demanding in terms of what the agents must know from the start about alternative hypotheses on what the world "really is". But, again, such an inferential machinery cannot be innocently postulated. Indeed, outside the rather special domain of "small worlds" whose structure is known *ex ante* to the agents, a few impossibility theorems from computation theory tell us that a generic inferential procedure does not and cannot exist (more on this point in Binmore, 1990; Dosi and Egidi, 1991; Dosi et al., 1994). This applies even more to so-called "Rational Expectations". It has repeatedly been shown that agents cannot generically learn even in simple stationary environment, and less so in complex evolving ones. More than that: under the latter circumstances trying sophisticated forms of learning is bad for the agents – in terms of prediction and performance – and is bad for the system – in terms of its growth and stability (on both points, see the discussion and the results from Dosi, Napoletano, Roventini, Stiglitz, and Trebich, 2017).[2]

Complexity arguments also imply a radical critique to the idea that "rationality" – however defined – rather than being an approximation to the empirical behaviours of purposeful, cognitively quite sophisticated, agents, could be, so to speak, an "objective" property of behaviours in equilibrium. Add the presumption that (most) observed behaviours are indeed *equilibrium* ones. And finally postulate some dynamics of individual adaptation or intra-population selection leading there. What one gets is some version of the famous "as...if" hypothesis, suggested by Milton Friedman (1953) and rejuvenated in different fashions by more recent efforts to formalize learning/adaptation processes whose outcome is precisely the "rationality" assumed from the start (archetypical examples of this faith can be found in Sargent, 1993, and McGrattan and Marimon, 1995).

A thorough, critical, discussion of the "as...if" epistemology has been put forward by Sidney Winter, in various essays (e.g., Winter, 1971) to which we refer the interested reader (and see also Silverberg, 1988; Andersen, 1994 and Hodgson, 1988).

For our purposes here, let us just note the following:

1 Any "as...if" hypothesis on rationality, taken seriously, is bound to involve quite a few restrictions similar to those briefly overviewed earlier with reference to more "constructive" notions of rational behaviours, simply transposed into a more "ecological" dimension – whether it is the "ecology" of minds, ideas, organisations, populations, etc. That is, canonical rationality, *stricto sensu*, postulates that one decides and acts by purposefully using the appropriate procedures (or by learning them in purposeful, procedurally coherent, ways). "As...if"s of any kind apparently relax the demands on what agents must consciously know about the environment, their goals, the process of achieving them, but at the same time must assume some background mechanism that generates the available alternatives – *which must include the "correct" ones*. However, without any further knowledge of the specific mechanisms, such a possibility remains a very dubious shortcut. And it is utterly unlikely when there are infinite alternatives which ought to be scanned.
2 While "realistic" interpretations of rationality put most of the burden of explanation upon the power of inbuilt cognition, "as...if" accounts shift it to selection dynamics – no matter whether driven by behavioural reinforcements alike salivating Pavlovian dogs, or by differential reproduction of traits within populations.[3] But, then, supporters of the view ought to show, at the very least, robust convergence properties of some *empirically justifiable* selection dynamics. As it stands, in our view, nothing like that is in sight. On the contrary, except for very special set-ups, negative results are abundant in e.g., evolutionary games or other forms of decentralized interactions – no matter whether applied to biology or economics – path-dependency cannot easily be disposed of; cyclical limit behaviours might occur (cf. Posch, 1997, Marengo and Pasquali (2011), and Kaniovski et al., 1997), etc. And all this appears even before accounting for environments which are genuinely evolutionary in the sense that novelties can emerge over time.

But, even more importantly, we can add that, in complex worlds, selection is almost powerless as an optimization mechanism. If the entities under selection have some internal structure made up of many interdependent components, such structural properties pose huge constraints on the evolution, and selection alone cannot break such a constraint (Kauffman, 1993). For instance, in biological evolution, the question whether an organism is optimal is nonsensical.

No doubt that standing on two legs has given us some useful evolutionary advantage, but is also a cause of many "inefficiencies" (back and knee weakness, difficulties in giving birth, etc.). Again, Gigerenzer and colleagues suggest an idea of "ecological rationality" rather than Olympic rationality, i.e., a system of heuristics that have co-evolved (Todd and Gigerenzer, 2012) and may all be suboptimal (whatever this may mean) if taken separately, but together produce a decently working system.

Another major perspective maintains that cognitive and behavioural assumptions must keep some empirical foundations and, thus, when needed, account for constraints on memory, on the maximum levels of complexity of problem-solving algorithms, and on computational time. It is, in a broad sense, the *bounded rationality* approach, pioneered by the works of Simon (1986) and developed in quite different fashions in e.g., organizational studies (starting from March and Simon, 1958 and Cyert and March, 1963); evolutionary theories (building on Nelson and Winter, 1982; see also Dosi et al., 1988; Hodgson, 1993; Andersen, 1994); "evolutionary games" (for a rather technical overview, cf. Weibull, 1995); for insightful remarks on bounded rationality and games in general, Kreps, 1996, and also in otherwise quite orthodox macroeconomics, see e.g., Sargent, 1993). Again, this is not the place to undertake any review of this vast literature. However, few comments are required.

Of course, the very idea of "bounds" on rationality implies that, at least in finite time, agents so represented fall short of full *substantively rational* behaviours, the latter involving among other things, (1) a full knowledge of all possible contingencies; (2) an exhaustive exploration of the entire decision tree; and (3) a correct appreciation of the utility evaluations of all mappings between actions, events and outcomes (Simon, 1986).

Given that, a first issue concerns the characterization of the origins and nature of the "boundedness" itself. It is not at all irrelevant whether it relates mainly to limitations on the memory that agents carry over from the past, or to algorithmic complexity of the decision problem to be addresses, or to limited ability of defining preferences over (expected) outcomes.

Or, more radically, couldn't it be due to the fact that agents get it basically wrong (in terms of representation of the environment, etc.)?

Here the theory faces a subtle but crucial crossroads. One alternative – unfortunately found too often in economic models, and especially but not only, in game theory – is to select the bounded-rationality assumptions with extreme casualness, suspiciously well-fitted to the mathematics the authors know and to the results one wants to obtain. We have no problem in associating ourselves to those who denounce the ad hocry of the procedure. The other alternative entails the acknowledgement of an *empirical discipline* upon the restrictions one puts upon the purported rationality of the agents. No doubt, we want to advocate here the scientific soundness of this procedure, notwithstanding the inevitable "phenomenological" diversity of cognitive and behavioural representations one is likely to get. That is, whether and how "rationality is bound" is likely to depend on the nature of the decision problem at hand, the context in which the decision-maker is placed, the pre-existing learning skills of the agents, etc. Taxonomical exercises are inevitable, with their seemingly clumsy reputation. But, in a metaphor inspired by Keith Pavitt, this is a bit like the comparison of Greek to modern chemistry. The former, based on the symmetry of just four elements, was very elegant, grounded in underlying philosophical principles, utterly irrelevant, and, from what we know nowadays, essentially wrong. The latter is clumsy, taxonomic, and for a long time (until quantum mechanics) lacking underlying foundations, but is certainly descriptively and operationally more robust.

A second major issue, regards *procedural* rationality. Granted the bounds on "substantive" rational agency, as defined above, when and to what extent should one maintain any assumption of coherent purposefulness and logical algorithmic consistency of the agents?[4] In a first approximation, Simon's approach suggests such a theoretical commitment (associated indeed to major contributions to the identification of *constructive* procedures for learning and problem-solving in this vein (Newell and Simon, 1972; Simon, 1976). However, even procedural consistency might not be at all a generic property of empirical agents, and a lot of evidence from most social disciplines seems to point in this direction.

Third, and relatedly, the very notion of "bounded rationality" (of, we repeat, the vast majority of contemporary economists, though not of Herbert Simon) commits from the start to an implicit idea that "full rationality" is the underlying yardstick for comparison. In turn, this implies the possibility of identifying some metrics upon which "boundedness" and, dynamically, learning efforts could be measured and assessed. In quite a few circumstances this can be fruitfully done[5] but in others it might not be possible either in practice or even in principle. In particular, this applies to search and learning in complex functional spaces (as many problems within and outside the economic arena commonly do).[6] And, of course, this is also the case of most problems involving discovery and/or adaptation to novelty.

Since indeed these features are typical of evolutionary environments, an implication is that one might need to go well beyond a restricted notion of "bounded rationality", simply characterized as an imperfect approximation to a supposedly "full" one – which, in these circumstances, one is even unable to define what it should precisely be.

But then, again, how does one represent learning agents in these circumstances? Our somewhat radical suggestion is that evolutionary theories ought to make a much greater and systematic use of the evidence from other cognitive and social sciences as sort of "building blocks" for the hypotheses on cognition, learning and behaviours that one adopts. We fully realize that such a perspective almost inevitably entails the abandonment of any invariant axiomatics of decision and choice. But, to paraphrase Thaler (1992), this boils down again to the alternative between being "vaguely right" or "precisely wrong": we certainly advocate the former (however, compare McGrattan and Marimon (1995) for a sophisticated contrary view).

In this respect, the discussion of *routines* as foundational behavioural assumptions of evolutionary models in Nelson and Winter (1982) is an excellent example of the methodology we have in mind, unfortunately not pursued enough in subsequent evolutionary studies (see Cohen et al., 1996; Becker 2004; Becker et al., 2005).

There are, however, many other fields where a positive theory of learning in economics can draw, ranging from cognitive and social psychology all the way to anthropology and sociology of knowledge.

## Cognitive categories and problem solving

A crucial aspect of learning regards most often *cognition*, that is the process by which decision makers form and modify representations in order to make some sense of a reality which is generally too complex and uncertain to be fully understood. Hence, the necessity to acknowledge the existence (and persistence) of a systematic gap between the agents' cognitive abilities and "reality" (were there an omniscient observer able to fully grasp it). Such a gap can take at least, two often interrelated forms,[7] namely, first, the *knowledge gap*, involving incomplete, fuzzy or simply wrong representations of the environment and, second, a *problem-solving* gap between the complexity of the tasks agents face and their capabilities in accomplishing them.

Regarding both, evolutionary theories of learning might significantly benefit from that branch of cognitive studies concerned with the nature and changes of *categories and mental models* (in different perspectives, cf. Johnson-Laird, 1983; Holland et al., 1986; Lakoff, 1987; Margolis, 1987; and the presentation of a few alternative theories in Mayer, 1992). It is crucial to notice that, if one accepts any "mental model" view, learning cannot be reduced to information-acquisition (possibly *cum* Bayesian processing of it), but rather is centred around the construction of new cognitive categories and "models of the world".

In turn, a robust evidence shows that cognitive categories are not clear-cut constructions with sharp boundaries and put together in fully consistent interpretative models. Rather, they seem to display (in all our minds!) blurred contours, shaded by an intrinsic fuzziness, held around some cognitively guiding "prototypes", and organized together in ill-structured systems kept operational also via a lot of default hierarchies (cf. on all those points, see Tversky and Kahneman, 1982; Holland et al., 1986; Kahneman and Tversky, 1986; Lakoff, 1987; Hahn and Ramscar, 2001; Gärdenfors, 2004; Fehr, 2005).[8] In this domain, note, however, a subtle but fundamental difference: is "prototypization" a "bias" or an inherent property of cognitive categorization? That is, is it similar to e.g., anchoring biases, *à la* Tversky and Kahneman, in principle still linkable to Olympic rationality with variable doses of "boundedness"? Or, on the contrary, is it intimately related to the very nature of mental categories? The answer we suggest is indeed in favour of the latter interpretation.[9]

## *Framing and social embeddedness*

Cognitive categories, it has repeatedly been shown, go together with various mechanisms of *framing* by which information is interpreted and also rendered operationally meaningful to the decision makers (cf. Kahneman et al., 1982; Borcherding et al., 1990; March, 1994).

Frames appear to be indeed a ubiquitous feature of both decision making and learning. What one understands is filtered by the cognitive categories that one holds and the repertoires of elicited problem-solving skills depend on the ways the problem itself is framed. That is, framing effects occurs along all stages of the decision process – affecting representations, judgements and the selection of behaviours (cf. Kahneman et al., 1982), and, concerning the patterns of activation of experts' skills (Ericsson and Smith, 1991).

As James March put it:

> Decisions are framed by beliefs that define the problem to be addressed, the information that must be collected, and the dimensions that must be evaluated. Decision makers adopt paradigms to tell themselves what perspective to take on a problem, what questions should be asked, and what technologies should be used to ask the questions. Such frames focus attention and simplify analysis. They direct attention to different options and different preferences. A decision will be made in one way if it is framed as a problem of maintaining profits and in a different way if it is framed as a problem of maintaining market share. A situation will lead to different decisions if it is seen as being about "the value of innovation" rather than "the importance of not losing face".
>
> *1994, p. 14*

Note that in this view, "frames" include a set of (non-necessarily consistent) beliefs over "what the problem is" and the goals that should be achieved in that case; cognitive categories deemed to be appropriate to the problem; and a related menu of behavioural repertoires.

Moreover, framing mechanisms appear at different levels of cognitive and behavioural observation: they do so in rather elementary acts of judgement and choice, but are also a general organizing principle of social experience and collective interactions (Bateson, 1972; Goffman, 1974).

One can intuitively appreciate also the links between framing processes and *social embeddedness* of both cognition and action.[10]

Frames, in the broad definition given above, have long been recognized in the sociological, psychological and anthropological literature (whatever name is used to refer to them) as being grounded in the collective experience of the actors and in the history of the institutions in which agency is nested.[11]

Indeed, embeddedness seems to go a striking long way and affect even the understanding and use of cognitively basic categories, such as that of causality and the very processes by which humans undertake basic operations, such as inferences, generalisations, deductions, etc. (Luria, 1976; Lakoff, 1987; D'Andrade, 2001; Kitayama, 2002; Oyserman and Lee, 2008).

Far away from standard rationality, a long and unjustly forgotten broadly defined Austrian tradition has tried to capture cognition and decision outside the straitjacket of the "max-something" framework. Hayek's "Sensory Order" (Hayek, 1952) is probably the most sophisticated synthesis of that view which

> not only emphasizes that the only ways open to people for making sense of their environment are the ways they already possess (the environment does not dictate how they see it), and whose probabilistic analysis of how the 'ways' at a person's disposal get called upon (the probabilities being a function of recent use and cumulative use) to see if there is a match with incoming stimuli provides a 'plastic' view of the mind.[12]

Related to Hayek's Sensory Order, are the ideas of Personal Construct Psychology (Kelly, 1955), in which the organizational structure that an individual creates to make sense of the world limits her permeability to new ways of thinking. This line of thinking was first applied in economics in Loasby (1983), where it is argued that organizational change is problematic if "core" constructs are involved, even if what is going on in the firm's environment may require the development of a new construct in order to adapt and survive. We are going to develop a somehow similar argument in our model introducing the distinction between "core bits" and "non-core bits" in the environment.

A somewhat parallel and almost entirely distinct literature focuses upon the crucial role of tacit knowledge (Polanyi, 1966). Building on that notion, behavioural and evolutionary economists have made a fruitful use of habits and, collectively, routines (see below) in order to characterize behavioural patterns.

By "genuinely behavioural" we mean that interpretative tradition which tries to characterize behavioural regularities in their own right (archetypical examples are Cyert and March, 1992, and March and Simon, 1958) as distinct from the somewhat more restrained approach to the description of actual behaviours in terms of deviations from some normative notion of prefect rationality as discussed above.

First, even in the simplest setups including stationary environments, satisficing behaviour may yield a probability of surviving for ever, while maximizing ones are sure to yield death to inferior options in finite time (Dutta and Radner, 1999). That is the exact opposite to the "as if" hypothesis.

Second, heuristics tend to be "fast and frugal", meaning that they are rules of thumb for decision making that are ecologically sound, simple enough to operate when time, information and computation are limited and grounded in human psychological capabilities, such as memory and perception (Gigerenzer and Goldstein, 1996).

Third, in largely unknown environments, even if stationary in their fundamentals, higher "competence gaps" may hinder the agent's capacity to assess which behaviour is the most appropriate in which environmental conditions. Behavioural inertia is the outcome, other things being equal, of higher environmental dynamics: "uncertainty is the basic source of predictable behaviour ... [T]he flexibility of behaviour to react to information is constrained to the smaller behavioural repertoires that can be reliably administered" (Heiner, 1983, p. 585). Indeed, this insightful conjecture from a strikingly neglected path-breaking demonstration is explored in Dosi, Napoletano, Roventini, Stiglitz, and Trebich, 2017), together with its applicability to non-stationary environment.

Fourth, and closer to our concern here, memory does not involve primarily information on past events (say, the memory of an econometrician going back in her time series), but rather memory of heuristics – both in their pattern recognition side and behavioural one.

## From individuals to organizations

As already mentioned, one side of the story is, in a broad sense, cognitive. The view of organizations as fragmented and multidimensional interpretation systems is grounded on the importance of collective information processing mechanisms that yield shared understandings (Daft and Weick, 1984), or "cognitive theories" (Argyris and Schön, 1978), of the environment in which they operate, and that assist organizations to bear uncertainty, besides, as we shall see, manage environmental and problem-solving complexity. If one subscribes to the notion that organizational learning is a process of refinement of shared cognitive frames involving action-outcome relationships (Duncan and Weiss, 1979), and that this knowledge is retained – at least for some time – and can be recalled upon necessity, this is like saying that organizational learning is in fact the process of building an organizational memory. This cognitive part of the memory is made up of "mental artefacts" embodying shared beliefs, interpretative frameworks, codes and cultures by which the organization interprets the state of the environment and its own "internal states" (Levitt and March, 1988).

Together, there is an operational side to the organizational memory involving the coupling between stimuli (events and signals, both external and internal ones) with responses (actions), making up a set of rules that remain available to guide the orientation of the organization and execute its operations. In this domain, memory largely relates to the ensemble of organizational routines – patterned actions that are employed as responses to environmental or internal stimuli, possibly filtered and elaborated via the elements of cognitive memory (much more on routines in Nelson and Winter, 1982; Cohen et al., 1996; Becker, 2004; Becker et al., 2005, and the literature reviewed here). As Cohen and Bacdayan (1994) put it, this procedural side is the "memory of how things are done", bearing a close resemblance to individual skills and habits, often with relatively automatic and unarticulated features (p. 554).

Cognitive and operational memories entail an "if…then" structure. Signals from the environment, as well as from other parts of the organization, elicit particular cognitive responses, conditional upon the "collective mental models" that the organization holds, which are in turn conditional upon the structure of its cognitive memory. Cognitive memory maps signals from an otherwise unknown world into "cognitive states" (e.g., "… this year the conditions of the market are such that demand for X is high ..."). Conversely, the operational memory elicits operating routines in response to cognitive states ("… produce X …"), internal states of the organization ("…prepare the machine M to start producing piece P…") and also environmental feedbacks ("… after all X is not selling too well …"). In turn, the organizational memory embodies the specific features of what an organization "thinks" and does, and what it is "good at", that is, its distinct capabilities.[13]

## *Modelling routines, memory and learning*

For a long time all the way to the present, organizational models have run far behind the qualitative interpretations briefly discussed above. Some catching-up has occurred, however, especially in the field of modelling learning processes in high dimensional spaces with relatively limited adaptation mechanisms. A promising candidate to model routines and memory finds its roots into the formalism of Classifier Systems (CSs) (Holland, 1975; Holland et al., 1986). In a nutshell, a CS is a system of interlinked condition/action rules that partly evolves according to the revealed environmental payoffs. Aiming to balance the rather unsynchronized research efforts and respective results between empirical and theoretical research, we build a model that finds its ascendancy there, and in their application in Marengo (1992), albeit with significant modifications.[14]

Dosi, Marengo, Paraskevopoulou and Valente (2017) present a model which links Classifiers Systems and NK fitness landscape models (Kauffman, 1993). The former provides a model of a memory system that accounts for both cognitive and operational memory, while they use the latter to represent an environment in which exogenous environmental traits and organizational actions or policies interact in a complex way to determine the organization fitness or payoff. While in standard NK models (e.g., Levinthal, 1997), cognition, actions and resulting payoffs are folded together in a mapping between "traits" and their "fitness", they unfold such a map, explicitly defining the cognition/action/environmental feedbacks and modelling their (evolving) coupling. This is, we believe, a first major advance with respect to the existing literature. The organization explores a complex and possibly changing landscape in which some dimensions are outside its control (the environmental traits) and some are within (the action traits). Since the former contribute to determine the payoff of the latter, the organization must base its search over the action landscape on an internal representation (its cognition) of the environmental landscape. When the landscape is complex enough and the organization has cognitive and memory bounds, such an internal representation can only be partial, imperfect and possibly wrong. However, in practice, through the accumulation of experience, organizations can develop better representations that enable them to act successfully in such a complex environment. This is a way of saying that organizations painstakingly and imperfectly learn and develop models of their environment. However, there is an exogenous world "out there" which is indeed the object of learning, and which of course is not controlled by the organization. Rather the organization has to learn what to do – the know-how – conditional on (what it believes to be) the characteristics of the landscape mapping the combinations of state-of-the-world and actions into payoffs. This is also another major difference vis-à-vis the NK modelling style wherein the "blackboxing" renders all the landscape notionally under the control of the agent. Moreover, the CS formalism allows a straightforward study of learning via non-local search, which, if undertaken at all in NK frameworks, turns out to be quite arbitrary.

In fact, the characteristics and evolution of organizational memory mirror the characteristics and evolution of organizational routines. In the case of routines, the memory elicits a "relatively complex pattern of behaviour triggered by a relatively small number of initiating signals or choices" (Cohen et al., 1996). How small or big is the initiating set of signals in itself is an important interpretative question, which has to do with the ways the organization categorizes environmental and intra-organizational information. And, likewise, the behavioural patterns are likely to display different degrees of conditionality upon particular sets of signals. So, at one extreme, the action pattern might be totally unconditional and "robust": "perform a given sequence of actions irrespectively of the perceived state of the world". At the opposite extreme actions might be very contingent on the fine structure of the "if" part, detailing very precisely the environmental conditions triggering the action part.

## Conclusion

A multi-millennial tradition of Western thought has asked "how do people behave?" and "how do social organization behave?", from Aristotle to St Augustin, Hume, Adam Smith, Kant, to name only a few giants. However, modern economics – and more recently social sciences colonized by modern economics – have taken up the answer by one of the shallowest thinkers, Bentham: people decide their courses of action by making calculations on the expected pleasures and pains associated with them. And, indeed, this *Weltanschauung* has spread all the way to the economics of marriage, of child bearing, of church going, of torture… (some more comments in Dosi and Roventini, 2016).

Our argument is that the Benthamian view is misleading or plainly wrong concerning the motivations, decision processes and nature of the actions.

First, the drivers of human motivation are many more than one. As Adam Smith masterly argues in its *Theory of Moral Sentiments*, utility (what he called "prudence") is just one of them, and in a lot of social contexts, not the most important one.

Second, the decision processes are very rarely explicit calculations and comparisons of outcomes.

Third, the ensuing decisions very seldom look like a "rational" ("as…if") outcome of the foregoing decision processes, *even when the latter would be possible to calculate*. And in the real word, complex and evolving as it is, they rarely are.

In such circumstances, we suggest, a positive theory of individual and collective behaviours has to entirely dispose of the max U(…,…, …) apparatus, either as an actual descriptive device, and as a yardstick, whatever that means. If we are right, then also relaxations of the paradigms involving varying degrees of "bounded rationality" in the decision process and an enlargement of the arguments in the utility function (e.g., adding "intrinsic motivations", or even "altruism") are quite misleading. They are a bit like adding epicycles over epicycles in a Ptolemaic astronomy.

The radical alternative we advocate is an anthropology of a *homo heuristicus* (Gigerenzer and Brighton, 2009), socially embedded, imperfectly learning in a complex evolving environment, and with multiple drivers of his actions.

A tall task, but it is time to break away from a paradigm that trivializes the analysis of human behaviour, reducing it to sterile exercises of maximization over some arbitrary and ad hoc functions.

## Acknowledgements

We thank Riccardo Viale for his insightful comments and suggestions. GD acknowledges the earlier support of the ISIGrowth project under grant agreement No. 649186 ISIGrowth, GD and LM the current support of the GROWINPRO project grant n. 822781, both from the European Union Horizon 2020 Programme.

This work draws heavily on previous publications, in particular, Dosi, Marengo and Fagiolo (2004) and Dosi et al. (2017).

## Notes

1 Note that here and throughout we address the notion of bounded rationality used by most contemporary economists as a "full rationality" minus some frictions, memory limitations, biases, noise, etc. and not the much richer notion conceived by Herbert Simon which stood from the very cognitive and perceptual boundaries in the interactions between humans and their environments.

2 Admittedly, the point is not uncontroversial as some scholars suggest that perceptual processes and categorizations are consistent with Bayes rule (e.g. Griffith, Kemp and Tenenbaum, 2008). We are not cognitive scientists, but frankly we find it hard to believe that people learn in Bayesian manners how to swim in the Heraclitean river where you can never bathe twice. More generally, for a thorough discussion of descriptive (as opposed to normative) theories of cognition and action, see Viale (2018).

3 Incidentally, note that the outcomes of pure "Pavlovian" – i.e., reinforcement-driven, consciously blind – and "Bayesian" – apparently sophisticated rational – dynamics can be shown to be sometimes asymptotically equivalent (see the review in Suppes 1995a, 1995b, who develops much older intuitions from behaviourist psychology, e.g. Bush and Mosteller, 1955). However, in order for that equivalence to hold, reinforcements must operate in the same direction as the Bayesian inferential machinery, which is, indeed, a hard demand to make. The so-called condition of "weak monotonicity" in the dynamics of adjustment that one generally finds in evolutionary games is a necessary, albeit not sufficient, condition to this effect. Moreover, a subtle question regards the interpretative value that one should attribute to asymptotic results: what do they tell us about finite time properties of empirical observations? (We shall briefly come back to the issue below.)

4 Note that procedural rationality requires all the "linearity assumptions" mentioned above (ruling out, for example, state-dependent preferences) and also consistent search heuristics (allowing, for example, assessment rules along any decision tree which at least in probability lead in the "right" direction).

5 Promising results stem from a better understanding of the formal structure of problem-solving heuristics (cf. e.g. Pearl, 1984; Vassilakis, 1997; and, in a suggestive experimentally-based instance, Cohen and Bacdayan, 1994, and Egidi, 1996). See also below.

6 For example, in Dosi et al. (1999), we consider quantity- and price-setting as cases to the point.

7 Heiner (1983) introduces a similar concept which he calls the "C-D (competence-difficulty) gap". In his definition, such a gap reflects the agent's imperfect capabilities to correctly process the available information and act reliably. Heiner's C-D gap does not properly belong to the realm of cognitive gaps, but it rather captures their behavioural consequences.

8 "Prototypization" is easy to intuitively understand: you would give a sparrow rather than a penguin as an example of what a bird is. But with that it is also easier to understand the basic ambiguity of borderliners, fuzziness and categorical attributions by default (how should one treat a duck-billed platypus?, as a mammal? or should one create a separate category, that of ovoviviparous?). A discussion of these issues bearing on economic judgements and behaviours is found in Tordjman (1998).

9 For a thorough discussion of algorithmic processes, see Lakoff (1987) and Bonini et al. (1999).

10 On the notion of "social embeddedness" as from contemporary economic sociology, see Granovetter (1985) and several contributions in Smelser and Swedberg (2006). A discussion quite germane to the argument developed here is found in Tordjman (1998).

11 Within an enormous literature, just think of a good deal of the sociological tradition influenced by the works of Talcott Parson or of the classic Bourdieu (1977); in anthropology, among others, cf. the discussions of "embeddedness" by Karl Polanyi (1944, 1957); and Geertz (1963); see also Edgerton (1985).

12 See also Frantz and Leeson (2013) for a recent reappraisal of the Hayekian view in relation to behavioural economics.

13 Within a very large literature, cf. for instance Helfat et al. (2006) and the critical survey in Dosi et al. (2008).

14 Is this a formal account of Hayek (1952)? Yes and no. It is in the sense of the local and path-dependent nature of learning. However, it is not in so far as this genre allows for cognitive and behavioural "mutations" meant to explore the unknown. In a sense it is "Schumpeter beyond von Hayek".

## References

Andersen, E. S (1994). *Evolutionary Economics: Post-Schumpeterian Contributions*, London: Pinter.

Argyris, C., and Schön, D. (1978). *Organisational learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.

Bateson, G. (1972). *Steps to an Ecology of Mind*, New York: Ballantine Books.

Becker, M. C. (2004). "Organizational routines: A review of the literature". *Industrial and Corporate Change* 13(4): 643–678.

Becker, M. C., Lazaric, N., Nelson, R. R., and Winter, S. G. (2005). "Applying organizational routines in understanding organizational change", *Industrial and Corporate Change* 14 (5): 775–791.

Binmore, K. (1990). *Essays on the Foundations of Game Theory*, Oxford: Blackwell.

Bonini, N., Osherson, D., Viale, R., and Williamson, T. (1999). On the psychology of vague predicatives, Mind and Language 14: 377–393.

Borcherding, K., Larichev, D. L., and Messick, D. M. (Eds.) (1990). *Contemporary Issues in Decision Making*, New York: North-Holland.

Bourdieu, P. (1977). *Outline of a Theory of Practice*, Cambridge: Cambridge University Press.

Bush, R. R., and Mosteller, F. (1955). *Stochastic Models for Learning*, New York: Wiley

Cohen, M. D., and Bacdayan, P. (1994). "Organizational routines are stored as procedural memory: Evidence from a laboratory study", *Organizational Science* 5: 554–568.

Cohen, M. D., Burkhart, R., Dosi, G., Egidi, M., Marengo, L., ... Coriat, B. (1996). "Routines and other recurring action patterns of organisations: Contemporary research issues", *Industrial and Corporate Change* 5: 653–698.

Cyert, R. M., and March, J. G. (1992). *A Behavioral Theory of the Firm* (2nd ed.), Oxford: Basil Blackwell.

D'Andrade, R. (2001). "A cognitivist's view of the units debate in cultural anthropology", *Cross-Cultural Research*, 35, 242–257.

Daft, R., and Weick, K. (1984). "Toward a model of organizations as interpretation systems", *Academy of Management Review*, 9(2), 284–295.

Dosi, G., and Egidi, M. (1991). "Substantive and procedural uncertainty: An exploration of economic behaviours in changing environments", *Journal of Evolutionary Economics*, 1(2): 145–168.

Dosi, G., Fabiani, S., Aversi, R., and Meacci, M. (1994). "The dynamics of international differentiation: A multi-country evolutionary model", *Industrial and Corporate Change*, 3: 225–241.

Dosi, G., Faillo, M., and Marengo, L. (2008), "Organizational capabilities, patterns of knowledge accumulation and governance structure in business firms: An introduction", *Organization Studies*, 29(August): 1165–1185.

Dosi, G., Freeman, C., Nelson, R. R., Silverberg, G., and Soete, L. (Eds.) (1988). *Technical Change and Economic Theory*, London: Pinter.

Dosi, G., Marengo, L., Bassanini, A., and Valente, M. (1999). "Norms as emergent properties of adaptive learning: The case of economic routines", *Journal of Evolutionary Economics*, 9(1): 5–26.

Dosi G., Marengo, L., and Fagiolo, G. (2004). "Learning in evolutionary environments", in K. Dopfer (Ed.), *The Evolutionary Foundations of Economics* (pp. 255–328). Cambridge: Cambridge University Press.

Dosi, G., Marengo, L., Paraskevopoulou, E., and Valente, M. (2017). "A model of cognitive and operational memory of organizations in changing worlds", *Cambridge Journal of Economics*, 41(1): 775–806.

Dosi, G., Napoletano, M., Roventini, A., Stiglitz, J., and Trebich, T. (2017). "Rational heuristics? Expectations and behaviors in evolving economies with heterogeneous interacting agents", S. Anna School of Advanced Studies, LEM Working Paper, 2017/31.

Dosi, G., and Roventini, A. (2016). "The irresistible fetish of utility theory: From 'pleasure and pain' to rationalising torture", *Intereconomics*, 51(1): 286–287.

Duncan, R. B., and Weiss, A. (1979). "Organizational learning: Implications for organizational design", in B. Staw (Ed.), *Research in Organizational Behavior*. Oxford: Elsevier.

Dutta, P. K., and Radner, R. (1999). "Profit maximization and the market selection hypothesis", *Review of Economic Studies*, 66(4): 769–798.

Edgerton, R. B. (1985). *Rules, Exceptions and Social Order*, Berkeley, CA: University of California Press.

Egidi, M. (1996) "Routines, hierarchies of problems, procedural behaviour: Some evidence from experiments", in K. J. Arrow, E. Colombatto, M. Perlman, and C. Schmidt (Eds), *The Rational Foundations of Economic Behaviour* (pp. 303–333). London: Macmillan.

Ericsson, K. A., and Smith, J. (Eds) (1991). *Toward a General Theory of Expertise*, Cambridge: Cambridge University Press.

Fehr, B. (2005). "The role of prototypes in interpersonal cognition", in M. Baldwin (Ed.), *Interpersonal Cognition* (pp. 180–205), New York: Guilford Press.

Frantz, R., and Leeson, R. (Eds) (2013). *Hayek and Behavioral Economics*, Basingstoke: Palgrave Macmillan.

Friedman, M. (1953). *Essays in Positive Economics*, Chicago: University of Chicago Press.

Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*, Cambridge, MA: MIT Press.

Geertz, C. (1963). *Peddlers and Princes*, Chicago: University of Chicago Press.

Gigerenzer, G., and Brighton, H. (2009). *Homo heuristicus*: Why biased minds make better inference, *Topics in Cognitive Science*, 1: 107–143.

Gigerenzer, G., and Goldstein, D. G. (1996). "Reasoning the fast and frugal way: Models of bounded rationality", *Psychological Review*, 103, 650–669.

Gigerenzer, G., Todd, P. M., and the ABC Research Group (Eds) (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Gigerenzer, G., Todd, P. M., and the ABC Research Group (Eds) (2012). *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press.

Goffman, E. (1974). *Frame Analysis: An Essay on the Organisation of Experience*, Harmondsworth: Penguin.

Granovetter, M. (1985). "Economic action and social structure: The problem of embeddedness." *American Journal of Sociology*, 91(3): 481–510.

Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008). "Bayesian models of cognition", in R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 59–100). New York: Cambridge University Press,

Hahn, U., and Ramscar, M. (2001). *Similarity and Categorization*, Oxford: Oxford University Press.

Hayek, F. A. (1952). *The Sensory Order*, Chicago: University of Chicago Press.

Heiner, R. A. (1983). "The origin of predictable behaviour", *American Economic Review*, 73(4): 560–595.

Helfat, C., Peteraf, E., Singh, M., Mitchell, W., Teece, D. ... Winter, S.G. (2006). *Dynamic Capabilities: Understanding Strategic Change in Organizations*. Oxford: Blackwell.

Hodgson, G. M. (1988). *Economics and Institutions: A Manifesto for a Modern Institutional Economics*, Cambridge: Polity Press.

Hodgson, G. M. (1993). *The Economics of Institutions*. Aldershot: Edward Elgar.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., and Thagard, P. R. (Eds) (1986). *Induction: Processes of Inference, Learning and Discovery*, Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1983). *Mental Models*, Cambridge, MA: Harvard University Press.

Kaniovski, Y. M., Kryazhimskiy, A. V., and Young, H. P. (1997). Learning equilibria in games played by heterogeneous populations. IIASA Interim Report. IIASA, Laxenburg, Austria.

Kahneman, D., Slovic, P., and Tversky, A. (Eds) (1982). *Judgment under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.

Kahneman, D., and Tversky, A. (1986). "Rational choice and the framing of decision", *Journal of Business* 59: 251–278.

Kauffman, S. A. (1993). *The Origins of Order*, Oxford: Oxford University Press.

Kelly, G. A. (1955). *The Psychology of Personal Constructs*, vol. 1: Clinical Diagnosis and Psychotherapy, New York: Norton.

Kitayama, S. (2002). "Culture and basic psychological processes: Toward a system view of culture: Comment on Oyserman et al. (2002)". *Psychological Bulletin*, 128: 89–96.

Kreps, D. M. (1996). "Market, hierarchies and mathematical economic theory", *Industrial and Corporate Change* 5: 561–595.

Lakoff G. (1987). *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*, Chicago: University of Chicago Press.

Levinthal, D. (1997). "Adaptation on rugged landscapes." *Management Science*, 43(7): 934–950.

Levitt, B., and March J. G. (1988). "Organizational learning", *Annual Review of Sociology*, 14(1), 319–338.

Loasby, B. J. (1983). "Knowledge, learning and the enterprise", in P. Wiseman (Ed.), *Beyond Positive Economics?* London: Macmillan.

Luria, A. R. (1976). *Cognitive Development: Its Cultural and Social Foundations*, Cambridge, MA: Harvard University Press.

March, J. G. (1994). *A Primer on Decision Making: How Decisions Happen*, New York: Free Press.

March, J. G., and Simon, H. A. (1958). *Organizations*, New York: Basil Blackwell.

Marengo, L. (1992). "Coordination and organizational learning in the firm", *Journal of Evolutionary Economics*, 2(4): 313–326.

Marengo, L., and Pasquali, C. (2011). "The construction of choice: A computational voting model", *Journal of Economic Interaction and Coordination*, 6: 139–156.

Margolis, H. (1987). *Patterns, Thinking and Cognition*, Chicago: University of Chicago Press.

Mayer, R. E. (1992). *Thinking, Problem Solving and Cognition*, New York: W. H. Freeman.

McGrattan, E. R., and Marimon, R. (1995). "On adaptive learning in strategic games". In A. Kirman, and M. Salmon (Eds), *Learning and Rationality in Economics*, Oxford: Basil Blackwell Publishers.

Nelson, R. R., and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*, Englewood Cliffs, NJ: Prentice Hall.

Oyserman, D., and Lee, S. W. S. (2008). "Does culture influence what and how we think? Effects of priming individualism and collectivism", *Psychological Bulletin*, 134: 311–342.

Pearl, J. (1984). *Heuristics: Intelligent Search Strategies for Computer Problem Solving*, Reading, MA: Addison-Wesley.

Posch, M. (1997). "Cycling in a stochastic learning algorithm for normal form games", *Journal of Evolutionary Economics*, 7: 193–207.

Polanyi, K. (1944). *The Great Transformation*, Boston: Beacon Press.

Polanyi, K. (Ed.) (1957). *Trade and Market in the Early Empires*, Glencoe, MN: Free Press.

Polanyi, M. (1966). *The Tacit Dimension*, London: Routledge & Kegan Paul.

Sargent, T. J. (1993). *Bounded Rationality in Economics*, Oxford: Clarendon Press.

Savage, L. (1954). *The Foundations of Statistics*, New York: Wiley.

Silverberg, G. (1988) "Modelling economic dynamics and technical change: mathematical approaches to self-organisation and evolution", in G. Dosi, C. Freeman, R. Nelson, G. Silverberg, and L. Soete (Eds.), *Technical Change and Economic Theory*, London: Pinter.

Silverberg, G., and Verspagen, B. (1994). "Learning, innovation and economic growth: A long-run model of industrial dynamics", *Industrial and Corporate Change*, 3: 199–223.

Simon, H. A. (1976). "From substantive to procedural rationality", in S. J. Latsis (Ed.), *Method and Appraisal in Economics* (pp. 129–148). Cambridge: Cambridge University Press.

Simon, H. A. (1986). "Rationality in psychology and economics", *The Journal of Business*, 59(4).

Smelser, N. J., and Swedberg, R. (Eds) (2006). *The Handbook of Economic Sociology*, 2nd ed., Princeton, NJ: Princeton University Press.

Suppes, P. (1995a). "A survey of mathematical learning theory 1950–1995", mimeo, Stanford University.

Suppes, P. (1995b). "Learning by doing, or practice makes perfect", mimeo, Stanford University.

Thaler, R. H. (1992). *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, New York: Free Press.

Tordjman, H. (1998). "The formation of beliefs on financial markets: Representativeness and prototypes", in N. Lazaric and E. Lorenz (Eds), *Trust and Economic Learning*, London: Edward Elgar.

Tversky, A., and Kahneman, D. (1982). "Judgments of and by representativeness", in D. Kahneman, P. Slovic, and A. Tversky (Eds), *Judgment under Uncertainty: Heuristics and Biases* (pp. 84–98). Cambridge: Cambridge University Press.

Vassilakis, S. (1997). "Accelerating new product development by overcoming complexity constraints", *Journal of Mathematical Economics*, 28(3): 341–373.

Viale. R. (2018). "The normative and descriptive weaknesses of behavioral economics-informed nudge: Depowered paternalism and unjustified libertarianism", *Mind and Society*, 17: 53–69.

Weibull, J. (1995). *Evolutionary Game Theory*, Cambridge, MA: MIT Press.

Winter, S. G. (1971). "Satisficing, selection and the innovative remnant", *Quarterly Journal of Economics*, 85: 237–61.