“Let us call [the bounded rationality] model of human choice the behavioral model, to contrast it with the Olympian model of SEU\(^1\) theory.

Within the behavioral model of human rationality, one \textit{doesn’t have to make choices that are infinitely deep in time}, that encompass the whole range of human values, and in which each problem is interconnected with all the other problems in the world.

....

Rationality of the sort described by the behavioral model \textit{doesn’t optimize}, of course. Nor does it even guarantee that our decisions will be consistent.”

Simon (1983), pp. 19-23; italics added

\* This is part of Selda Kao’s doctoral dissertation to be submitted to the School of Social Science and the Department of Economics, University of Trento, during this academic year. The thesis is being prepared under the supervision of Vela Velupillai and Stefano Zambelli. We are both indebted to V. Ragupathy for invaluable intellectual and logistical advice and support, although he is not responsible for any of the remaining infelicities.

\(^1\) Subjective Expected Utility.
Abstract

In this paper, the origins and development of behavioural economics, beginning with the pioneering works of Herbert Simon (1953) and Ward Edwards (1954), is traced, described and (critically) discussed, in some detail. Two kinds of behavioural economics – classical and modern – are attributed, respectively, to the two pioneers. The mathematical foundations of classical behavioural economics is identified, largely, to be in the theory of computation and computational complexity; the corresponding mathematical basis for modern behavioural economics is, on the other hand, claimed to be a notion of subjective probability (at least at its origins in the works of Ward Edwards). The economic theories of behavior, challenging various aspects of ‘orthodox’ theory, were decisively influenced by these two mathematical underpinnings of the two theories.

JEL Codes: C61, C63, D03, D81, D83

Keywords: Classical Behavioural Economics, Modern Behavioural Economics, Subjective Probability, Model of Computation, Computational Complexity. Subjective Expected Utility.
0. A Preamble

Behavioural economics may have, finally, come of age. It is part of the curricula of graduate schools in economics, finance and management, often even one of the compulsory courses.

More than a decade ago, in a letter to Velupillai (Simon, 2000; italics added), Herbert Simon was optimistic enough to state, after a half-a-century of tireless efforts to make behavioural economics a viable alternative to orthodox neoclassical economics, that:

> The economists here [at Carnegie Mellon University] remain, for the most part, .. backward … , but I am encouraged by the great upswell, in the US and especially in Europe, of experimental economics and various forms of bounded rationality. I think the battle has been won, at least the first part, although it will take a couple of academic generations to clear the field and get some sensible textbooks written and the next generations trained.

Yet, not much more than one year earlier, at the 84th Dalhem Workshop on Bounded Rationality: The Adaptive Toolbox (Gigerenzer-Selten, p.ix), two distinguished economists claimed:

> “Bounded rationality needs to be, but it is not yet, understood.”

How, one may legitimately ask, can a ‘battle [have] been won’, with a crucial concept lying at the foundation of its ‘armory’, ‘yet to be understood’? We believe there is a case for Gigerenzer and Selten to feel that the notion of bounded rationality remains to be clarified. This is because they have been meticulous in having dissected the way the notion has been (ill-) defined by varieties of orthodox theorists, including those we shall shortly identify as some of the pioneers of modern behavioural economics. Moreover, they have also understood, with impeccable perspicacity, that boundedly rational behavior has nothing to do with either optimization, or irrationality (ibid, p.4).

---

2 In the volume on Behavioural Economics, as part of the Routledge Major Works Series in New Trends and Frontiers in Economic Analysis, edited by Shu-Heng Chen and Vela Velupillai (2012), there are – at the moment – 360 references in the five subject areas of Agent-Based Behavioural Economics, Behavioural Finance, Behavioural Game Theory, Behavioural Neuroeconomics and Modern Behavioural Economics. This list does not include the literature on Classical Behavioural Economics, which could easily add at least another 50 items to the list. By the way, none of them make the distinction between classical and modern behavioural economics.

3 This footnote added by Velupillai. As it is in our University’s graduate School of Social Science. It was once given by Velupillai, till his responsibility for the course on Behavioural Economics was abruptly terminated by the director and academic director of the previous incarnation of the graduate school, then called CIFREM. His course emphasized the distinction between classical and modern behavioural economics, emphasizing the underpinning of the former in a model of computation and the latter, at least in its origins, in subjective probability theory à la De Finetti-Savage. The current rendering of behavioural economics, as a graduate course, makes no such distinction between two kinds of behavioural economics.


5 The two paradigmatic examples of this genre, representing ‘old neoclassical’ and newclassical economics, are, respectively Frank Hahn & Roy Radner, on the one hand, and Thomas Sargent, on the other (cf., Gigerenzer-Selten, p.5, Radner (1980), Hahn (1985), pp. 15-16 and Sargent (1993), pp. 21-24.
Where we differ with Gigerenzer and Selten is their anchoring of bounded rationality and satisficing in ‘fast and frugal stopping rules for search’ without, however, providing this anchor a solid foundation in itself. Bounded rationality and satisficing, in our framework, is a natural outcome of replacing optimization with decision problems (in its metamathematical senses), whereby problem solving, in general, and human problem solving in particular, lead to structured search in computationally complex spaces that are classified in terms of solvability, decidability and computability. Optimization becomes a very special case of the solvability of a decision problem, intrinsically coupled to algorithms, which are given measures of complexity that are capable of encapsulating the notions of ‘fast and frugal’ in precise ways.

The rest of the paper is structures as follows. A broad brush discussion of the two kinds of behavioural economics is provided in the next section. Next, the analytical foundations of modern and classical behavioural economics is discussed and dissected in section 2. Section 3 is devoted to a discussion of the special role played by Herbert Simon in forging, \textit{ab initio}, classical behavioural economics and its rich vein of characterizing subfields. The concluding section suggests ways of going forward with a research program in classical behavioural economics – eventually with the hope of exposing the lacunae in the foundations of modern behavioural economics, and its ad hockeries.

1. Emergence of Behavioural Economics

Behavioral economics, which originated, almost fully developed, during the 1950s, can be classified into at least two streams - \textit{Classical} and \textit{Modern}. The former was pioneered by Herbert Simon and the latter by Ward Edwards, respectively. The two streams are clearly distinguishable on the basis of their methodological, epistemological and philosophical aspects. Despite having sharp contrasts in their approaches to understand human behavior, a clear distinction between them was not made until recently (Velupillai, 2010b). Behavioral economics, \textit{in general}, challenges orthodox economic theory and its foundational assumptions regarding human behavior, its institutional underpinnings (especially in its Classical versions pioneered by Simon), its poor prediction power and its intrinsic non-falsifiability.

The main distinctions of \textit{Modern} Behavioral Economics (henceforth MBE) and \textit{Classical} Behavioral Economics (henceforth CBE) can be classified into three aspects. First, MBE assumes economic agents are maximizing utility with respect to an underlying preference
order – to which ‘an increasingly realistic psychological underpinning’ is attributed (Camerer, et. al., 2004, p. 3); CBE assumes no underlying preference order and an economic agent’s decision making behavior, at any level and against the backdrop of every kind of institutional setting, is subject to bounded rationality and exhibits satisficing behavior. Put another way, MBE remains within the orthodox framework of optimization under constraints; CBE is best understood in terms of decision problems (in the metamathematical sense, cf. Velupillai, 2010b). Second, MBE concerns the behavior of agents and institutions in or near equilibrium⁶; CBE investigates disequilibrium or non-equilibrium phenomena. Third, MBE accepts mathematical analysis of (uncountable) infinite events or iterations, infinite horizon optimization problems and probabilities defined over σ-algebras and arbitrary measure spaces⁷; CBE only exemplifies cases which contain finitely large search spaces and constrained by finite-time horizons. The aim of this chapter is to introduce and elaborate the respective foundations of two streams in behavioral economics and their applications. Subsequently, the research ideas will concentrate on, and expand upon, some of Simon’s fundamental contributions to CBE, based on his explicit (computationally constrained cognitive) and implicit (computational complexity) foundations, always with a model of computation (usually the Turing Model of Computation) constraining and disciplining his research program.

1.1 Modern Behavioural Economics

1.1.1 Origins

“The combination of subjective value or utility and objective probability characterizes the expected utility maximization model; Von Neumann & Morgenstern defended this model and, thus, made it important, but in 1954 it was already clear that it too does not fit the facts. Work since then has focussed on the model which asserts that people maximize the product of utility and subjective probability. I have named this the subjective expected utility maximization model (SEU model).”


The origins of Modern Behavioral Economics are often claimed to have emanated from the early works by Richard Thaler, along with Kahnmann and Tversky, for example in the following quote:

---

⁶ The ‘near’ is defined, in all cases we are aware of, by uncomputable approximation processes of uncomputable equilibria.

⁷ Despite ostensible reliance on the kind of subjective probability defined by De Finetti and Ramsey, both of whom eschewed even countable infinities.
“Kahneman and Tversky provided the raw materials for much of behavioral economics – a new line of psychology, called behavioral decision research, that draws explicit contrasts between descriptively realistic accounts of judgement and choice and the assumptions and predictions of economics. Richard Thaler was the first economist to recognize the potential applications of this research to economics. His 1980 article “Toward a theory of consumer choice,” published in the first issue of the remarkably open-minded (for its time) Journal of Economic Behavioral and Organization, is considered by many to be the first genuine article in modern behavioral economics.” Camerer et al. (2004), pp. xxi-xxii.

Contrary to these claims, the real origins of modern behavioral economics can be traced back to Ward Edwards, particularly to Edwards (1954) and Edwards (1961), which provide the methodological framework within which modern behavioral economics can be identified. Edwards, in turn, draws inspiration from the famous subjective probability theorist and statistician Leonard Savage. The two papers summarize the emergence of core notions that characterize what may, with hindsight, be called a Neoclassical Theory of Behavioural Economics and offer detailed philosophical and methodological discussions related to them. More importantly, Edwards posed challenges to orthodox neoclassical notions, basing them on psychological and experimental evidences. It also introduces and provides a remarkable and detailed survey of the classic works in the field of behavioral economics till then. Other contributions of Edwards include the confirmation of intransitive behavior and introduction of experimental results and stochastic (transitivity) models of individual behavior.

The most remarkable aspect of Edwards’ papers is the formalization of weighted values and the introduction of Subjective Expected Utility (Ramsey (1931); Savage (1954)). He also sheds light on the early studies on subjective probability that happened before and after Savage’s book in 1954. The standard formulation of the objective function faced by a decision maker in an economic model under risk/uncertainty is presented as a linear combination of the values of outcomes and probabilities attached to each of these outcomes. The values of outcomes and probabilities, both, can be objective or subjective. The formulation of expected utility can be stated as:

\[
E(U) = \sum_{i=1}^{n} p_i \cdot U_i
\]

---

8 As if these were not prime motivations for Simon when he launched his program of research on behavioural economics long before some of these authors were even born. It is just that Simon’s psychological and cognitive bases for modeling realistic economic behavior was always underpinned by a model of computation.

9 So the claim that behavioral economics was not even a field till 1980 is highly questionable, even from the works by precursors to Kahneman, Tversky and Thaler.
where \( p_i \) is the probability of \( i_{th} \) outcome out of \( n \) possible ones and \( U_i \) is the value of \( i_{th} \) outcome. Based on this we can have the following classification: when subjective values are weighted with objective probabilities, it results in Expected Utility. Instead, when subjective values are weighted with subjective probabilities, it becomes Subjective Expected Utility. the other two alternatives were considered to be unimportant or proved to be unrealistic in the literature.

The classic Expected Utility formulation was first devised by von Neumann and Morgenstern, who explicitly invoked formal, ‘objective’ probability theory and were even prepared to use the frequency theory of probability\(^{10}\) – explicitly and forcefully rejected by Savage, whose work was deeply influenced by De Finetti’s foundational work on subjective probability theory. Thus, the probability with which they axiomatized expected utility maximization is actually objective. Since then, it became clear that Expected Utility fails to explain and predict individual behavior under risk – let alone uncertainty (a distinction not carefully maintained by practitioners of MBE). For vN.-M., economic problems were not clearly formulated and, besides, the mathematical tools, they felt, were improperly used. They attempted to make the qualitative notion of utility and preference measurable just like, say, force in physics. The main argument was that, for economics to be a rigorous science, formalised mathematically, preferences should be measurable. Furthermore, for preferences to be measurable, they should be numerically definable and mutually comparable. Individuals are supposed to seek and be able to choose the outcome which will give them the highest satisfaction among all the possibilities. But neither the process that underpinned ‘seeking’, nor the process of ‘choosing’ were given any procedural content, unlike the way Simon, who from the outset sought to emphasise the search processes at the foundations of choice over a complex space of alternatives.

There was a great deal of effort that was dedicated to measuring utilities and probabilities under the framework of subjective (personal) probability around the time of the early work of Ward Edwards. This empirical work went hand-in-hand with the simultaneous formalization by Savage, who built his foundations of statistics on the basis of De Finetti’s theory of subjective probability. In this scheme, the assumptions of complete preference ordering and the sure-thing principle play a crucial role, and the individuals learn and adjust their prior beliefs with the occurrence of events according to the Bayes’s theorem. These properties for

---

\(^{10}\) But not in its modern refounding and reformulation as algorithmic probability.
subjective probabilities proposed by Savage, in turn, implies that individuals with different set of subjective probabilities, over the course of their experience, will end up having close subjective probabilities which coincide with each other.

The critical point of rapid development of MBE can be attributed to the proposal of *Prospect Theory* by Kahneman and Tversky in (Kahneman and Tversky, 1979), which was considered a satisfactory replacement of expected utility theory. The theory encapsulates the idea of subjective probability\(^{11}\) (not directly) and loss aversion. Even today, loss aversion is still one of the most notable behavioral reasoning used to interpret and model decision making in different contexts.

A series of “anomalies” - resulting from the violation of transitivity and other axioms, inconsistency of some principles of neoclassical economics - have been systematically collected and investigated by contemporary behavioral economists, notably, Richard Thaler, Colin Camerer, George Loewenstein, Matthew Rabin among many others, since the late 1980s in the *Journal of Economics Perspectives*. The inconsistency in behavior is mainly observed in experimental environments, and thus the neglect of psychological and social factors are proposed as possible causes for this, according to MBE. The Neoclassical agents are now like physically weakened patients unable to predict even reasonably well, who are being examined with the benchmark idealized case of orthodox theory and its strict rational, constrained optimization, behaviour and the modern behavioral economists are assuming the role of seeking and proposing the remedies for them. The themes and fields challenged from which *anomalies*\(^{12}\) are found cover Microeconomics, Macroeconomics, Finance Theory, Industrial Organization to Game Theory to Development Economics. This has led to a collectively divided field of behavioral economics, broadly, into (at least) Behavioural micro, Behavioural macro, Behavioural finance and Behavioural game theory.

---

\(^{11}\) But neither consistently, nor meaningfully. In the whole literature on MBE, all the way from the early works of Kahneman and Tversky, there is a remarkable confusion and conflation of a variety of theories of probability, even within one and the same framework of modelling rational, psychologically underpinned, individual behaviour in economic contexts. See, in particular, Amos Tversky: *Elimination by Aspects – A Theory of Choice*, *Psychological Review*, Vol. 79, # 4, July (1972).

\(^{12}\) Velupillai refers to this trait in MBE as *anomaly mongering* in his lectures on Behavioural Economics. His point is that both the Neoclassicals, whose analogous notion is ‘puzzles’ – ‘equity premium puzzle’ being paradigmatic – and the Modern Behavioural Economists are consciously invoking Kuhn’s terminology and, therefore, suggesting that their program of research is leading to that much maligned concept of a ‘paradigm shift’. However, see below, under the subsection on Behavioural Finance.
1.1.2 Fields of Modern Behavioural Economics

Behavioural Microeconomics Some anomalies concerning preference and utility in decision making are studied. Preference reversal is believed to be a robust anomaly. This field of research attempts to challenge the commonly agreed notion in neoclassical theory that the values of goods or outcomes do exist, and people know these values directly, by highlighting the presence of framing effects, reference based effects etc. (Tversky and Thaler, 1990). It is also suggested that the assumption of a stable preference ordering should be discarded. The preference changes can be due to a variety of factors such as status quo, loss aversion, ambiguity aversion and endowment effects. Their thesis is that a consideration of these factors can make the analysis of preference more manageable and tractable\(^{12}\). The general worry is that importing psychological inspirations into existing economic models may create new complexities and reduce their predictive power (Kahneman et al., 1991). Furthermore, the difficulty and infeasibility of utility maximization was pointed out, and economists sought for possible psychological and social causes as explanation for the “mistakes”\(^{14}\) in decision making (Kahneman and Thaler, 2006).

In MBE, the majority of research is focus on suggesting more ‘realistic’ utility functions and decision contexts of modeling. There is a minority of research questioning the fundamental framework of preference, and maximization. Such as Slovic (1995) and modeling of satisficing. For formalizing satisficing, heuristic searches\(^ {15}\) are applied. Heuristics serves a guide helping decision makers to find the short cuts for relevant information. Together with satisficing, decision makers are supposed to stop searching – i.e., an exogenously determined stopping rule for the search process is activated - whenever some (exogenously determined) criteria are achieved (e.g. aspiration level). However, they are not necessarily aware of computability or undecidability which is inherent in many such procedures. If the heuristic search is programmed as a finite automaton, it will naturally terminate at some point. However, if it is programmed as a Turing Machine, then the decision maker is confronted with the famous result of the halting problem for Turing Machines. This means that the agent who is searching will either not be able determine whether the heuristic reached the exogenously determined aspiration level, or – even worse – whether it will ever do so within any reasonable, or even

\(^{12}\) Not formalized in terms of tractability in the formal hierarchy of degrees of computational complexity simply because these models are not underpinned by any formal model of computation.

\(^{14}\) In other words, ‘Anomalies’!

\(^{15}\) Without, however, any recognition that ‘heuristics’ are, formally, ‘algorithms’.
unreasonable, exogenously given time span.

**Behavioural Macroeconomics** Similar psychological and social reasons are also applied to interpret some Macroeconomic phenomena\(^\text{16}\), such as, money illusion, rigidity of (nominal) wages (loss aversion and fairness) and involuntary unemployment (gift-changing equilibrium of reciprocal preference). The most far-reaching challenge might be to address the questionable idea of the traditional notion of Discounted Utility. The presence of non-exponential discounting of utility was observed (Loewenstein and Thaler, 1989), and subsequently\(^\text{17}\) more complex ways (hyperbolic, quasi-hyperbolic discounting etc.) of discounting were invented, which are believed to be more realistic and better able to provide predictive models in the context of inter-temporal choices. Other than time discounting, there is also research on behavioral lifecycle theories (e.g. mental accounting (Thaler, 1990)) on savings and marginal propensity to consume and on regret theory, such as using counterfactual, introspective thinking and self control of future misbehavior on consumption and saving. But in no such case have non-traditional logics been utilized to derive counterfactual predictions based on introspective thinking.

**Behavioural Finance** Behavioral finance stands on the ground against the *Efficient Market Hypothesis* and it is probably one of the most developed subfields in behavioral economics. In other words, it is commonly believed that the efficient market hypothesis has virtually died out. The well known anomalies in finance include the equity premium puzzle (high risk aversion), calendar effects, status quo effect, limits to arbitrage, social preference and other stylized facts (De Bondt and Thaler, 1989; Froot and Thaler, 1990; Lamont and Thaler, 2003; Lee et al., 1990; Siegel and Thaler, 1997; Thaler, 1987a,b). Due to the nature and functioning of financial markets, huge amount of data points, at high frequencies, are available. Therefore it is also a rich ground for behavioral and (so called) computational economists to investigate and validate their models.

**Behavioural Game Theory** Similar to the other fields, behavioral game theory investigates how the results regarding strategic interaction deviate from the *orthodox* game theoretic

\(^{16}\) Akerlof and Shiller (2009) categorizes five types of animal spirits (likely to be misnamed): they are confidence, fairness, corruption and antisocial behavior, money illusion and stories.

\(^{17}\) Not quite ‘subsequently’, because the notion of hyperbolic discounting has been ‘around’ in intertemporal macroeconomic policy models at least since the early 1960s. But it is the credit of the MBE’s practice and insistence that the traditional and almost routine recourse to exponential discounting in intertemporal optimization models is being challenged.
predictions in the light of some behavioural assumptions regarding decision making in strategic situations. The psychological and social explanations such as guilt aversion and fairness criteria are incorporated into the traditional models. Behavioral game theory benefits from the fact that most of these models can be tested in laboratory environments by collecting a sufficient number of subjects. Therefore, it coexists with experimental economics and neuroeconomics which will be introduced later. A reasonably up to date survey of behavioral game theory can be found in Camerer (2003).

1.1.3 Concluding Remarks

Although Neoclassical Economic theories have been critically questioned by economists and psychologists for many decades, it is still explicitly specified that optimization, equilibrium and efficiency, on which Neoclassical economic – and its variants, such as Newclassical and New Keynesian - theories are based, is not completely rejected by behavioral economists (see, for example, the opening, programmatic, pages of Camerer and Loewenstein (2004)). The ultimate goal of behavioral economists seems to be to extend or replace neoclassical theories in a normative sense.

Modern behavioral economists have, over the years, discovered and categorized different forms of deviations from consistent behavior. A valid question here would be: why do these anomalies arise and what are they anomalies with respect to? These discoveries such as reference dependence and loss aversion, preference over risky and uncertain outcomes and time discounting, came mostly from observations in experimental environments. The anomalies and puzzles that were discovered and discussed are departures with respect to the neoclassical normative benchmark for judging rational behavior, which is expected utility maximization. These evidences or anomalies are in turn used to formulate more realistic utility functions and further, these modified utility functions are incorporated into the existing models. In some sense, Modern behavioral economists modified fractured pieces in the foundations of Neoclassical theories, but still they worked within its basic premises (preferences, utility, equilibrium and maximization).

Firstly, MBE preserves the doctrine of utility and does not go beyond it or discard it. Secondly, though the behavioral models do consider more realistic psychological or social effects, economic agents are still assumed to be optimizing agents whatever the objective functions
may be. In other words, MBE is, still, within the ambit of the neoclassical theories or it is in some sense only an extension of traditional theory by replacing and repairing the aspects which prove to be contradictory. These adjustments in turn are expected to enhance the predictive power of original theories. On the contrary, CBE does not try to endow the economic agent a preference order which can be represented by utility functions; nor, of course, do equilibria or optimization play any role in the activation of behavioural decision making by CBE agents.

1.2 Classical Behavioural Economics

It is interesting to note that even before the advent of behavioral economics, economics was still very much based on behavioral principles and psychology. For example, Adam Smith and Keynes explicitly considered psychological factors in their theories. With the rise of neoclassical economics, nearly all psychological factors were removed from normative economic theories. The phrases distinguishing “Classical” from “Modern” behavioral economics come about, partly, also for chronological reasons.

One of the most essential and concrete line separating MBE and CBE is that rational behavior is adaptive or procedural in the postulation of CBE; this makes rational behavior naturally algorithmic and the need to underpin it with a model of computation enters right on the ground floor of theory and its empirical counterparts. Given the nature of adaptive behavior and the complex environment in which it takes place, optimization principles and equilibrium analysis become meaningless and nearly infeasible. The resolving of these difficulties should not be to find approximations of sophisticated mathematical models using numerical techniques, like what we see in some parts of MBE.

As far as dynamical rational behavior is concerned, where procedure is central, Simon, Richard Day, Richard Nelson and Sidney Winter are considered as the pillars of CBE (Velupillai, 2010b). The research line was motivated by the questions ‘How does the mind work?’; ‘What kind of Mechanisms should we postulate for the Mind, based on current knowledge and research on Cognitive Science, to make sense of observed behavior?’; ‘What postulates are useful to understand and predict behavior?’; ‘What metaphors are useful to formalize intelligent procedural behavior?’; ‘How do operational Institutions Emerge and Survive’?;
Research surrounding these questions is intrinsically underpinned by cognitive psychology and the theory of computation. They lead also to what became the natural Simon framework of *Human Problem Solving*, of agents faced with complex and intractable search spaces, constrained by computationally underpinned cognitive processes facing time and resource constraints. A notable precursor for Simon, on these aspects, was Polya.

Simon is best known by the felicitous phrase he coined, “bounded rationality”, which appeared in Simon (1957) for the very first time (although it had appeared in other forms already from his classic book on Administrative Behaviour (1947). Bounded rationality generally refers to the internal cognitive limitations, and the constraints of the external environment which confront human minds in decision making contexts. This latter is more specifically contextualized by the Institutional backdrop for individual behavior. Therefore, in order to incorporate the notion of bounded rationality into the behavioral model more rigorously, one ought to investigate how human thinking is limited internally and how human beings adapt and interact with the environment, especially as members of an institution.

Simon’s insight about modeling adaptive individuals in complex economic environments can be better understood in the fragment:

“Suppose we were pouring some viscous liquid molasses into a bowl of very irregular shape. ... How much would we have to know about the properties of molasses to predict its behavior under the circumstances? If the bowl were held motionless, and if we wanted only to predict behavior in equilibrium, we would have to know little, indeed, about molasses. The single essential assumption would be that the molasses, under the force of gravity, would minimize the height of its center of gravity. With this assumption, which would apply as well to any other liquid, and a complete knowledge of the environment, in this case the shape of the bowl, the equilibrium is completely determined. Just so, the equilibrium behavior of a perfectly adapting organism depends only on its goal and its environment; it is otherwise completely independent of the internal properties of the organism.

If the bowl into which we were pouring the molasses were jiggled rapidly, or if we wanted to know about the behavior before equilibrium was reached, prediction would require much more information. It would require, in particular, more information about the properties of molasses: its viscosity, the rapidity with which it “adapted” itself to the containing vessel and moved towards its “goal” of lowering its center of gravity. Likewise, to predict the short run behavior of an adaptive organism, or its behavior in a complex and rapidly changing environment, it is not enough to know its goals. *We must know also a great deal about its internal structure and particularly its mechanisms of adaptation.*”
Simon criticized orthodox normative economics for ignoring how human beings actually behave and questioned the result that only rational agents survive the forces of competition — with orthodoxy’s Olympian assumptions (Simon, 1983) on how to formalize rational behavior, which was - at least as far as Simon was concerned – remote from any cognitive realism. Besides, the study of equilibrium requires little understanding of the characteristics of individuals in out-of-equilibrium situations, simply because normative economics has nothing to say about process and procedure. In the real world that Simon saw around him, there exists a lot of turbulence, not only generated by external shocks, that keeps the system out of equilibrium and agents needing to relocate their bearings almost ceaselessly.

Furthermore, Simon stated “decision making under uncertainty” instead of “decision making under risk” in Simon (1959). That is, an economic agent might respond to the changing environment in a personal way rather than knowing the objective probability of what outcomes might happen in the future. This property brings more difficulties on the prediction of rational individual behavior by using so-called objective characteristics of the environment.

Simon’s behavioral economics is almost comprehensively demonstrated by his encapsulation of Human Problem Solving and agents and institutions as Information Processing Systems. Although the problems which Simon dealt with are well structured problems, such as Chess playing, the combinatorial complexity of the problem is massive enough to prevent human players using minimax strategies which are suggested in traditional game theory. In this paper, only some of Simon’s massive and wide ranging contributions are covered. The underpinning of CBE and Simon’s special role will be found in later sections and chapters.

2 Underpinnings of Behavioral Economics

In this section, different underpinnings and analytic tools of MBE and CBE will be briefly mentioned. The purpose of this section is not to provide detailed theoretical and technical instructions for them, but aspire to make the clear distinctions on how the two lines are different fundamentally. It is a slightly puzzling that this distinction has never been made earlier. As one may realize from the following underpinnings and the sub-branches of MBE introduced in the previous section, MBE can be characterized as a massive magnet which attracts different resources, new tools and ways of explanations. We can almost claim that MBE has already
become a new mainstream economics, as a consequence of MBE playing the role of a revised approach of orthodox economics rather than an alternative approach. On the other hand, CBE is developed on completely different grounds from MBE. From our point of view, MBE is fostered by Orthodox Economic Theory, Game Theory, Mathematical Finance Theory and Recursive\textsuperscript{18} Methods, Experimental Economics and Neuroeconomics, Computational Economics\textsuperscript{19} and Subjective Probability Theory.

CBE, in our reconstruction of it, on the other hand, is based fundamentally on a model of computation – hence, Computable Economics – computational complexity theory, nonlinear dynamics and algorithmic probability theory.

## 2.1 Underpinnings of Modern Behavioral Economics

### 2.1.1 Orthodox Economic Theory

It is in human nature to aspire to predict, at least so the sages say and the traditional wisdom of many cultures concur\textsuperscript{20}. Microeconomics, in general, is the study of individual choices and actions. Gradually, economics has developed normative axioms\textsuperscript{21} and theories on how the individual entities (including organizations) should make choices and how they seem to make choices. There are, classically (but not necessarily exhaustively) the normative and positive approaches to behavior, respectively. In Neoclassical theory, economic agents are assumed to be fully rational and completely\textsuperscript{22} informed. It is not that they know everything, but that they can know everything and there are means to learn – epistemology - and they know how to make the best choices for themselves (even if only probabilistically). Second, in order for their choices to be tractable, axioms (completeness, reflexiveness, transitivity, and continuity) of rational preference were devised, within classical mathematical formalisms – which simply means the mathematics of (Zermelo-Fraenkel) set theory plus some variant of the axiom of choice. Individuals are assumed to have underlying preference orderings for all the alternatives

\textsuperscript{18} Not Recursion Theory.
\textsuperscript{19} Not Computable Economics.
\textsuperscript{20} In Tamil, Velupillai’s Mother Tongue, an ancient precept is: The Art of Good Government is the ability of the Minister to Predict Accurately. This is not a literal translation and does not even remotely convey the condensed wisdom in the original.
\textsuperscript{21} We are, of course, aware that axioms have no normative status, except in so far as they reflect it on the basis of the implications derived, by some deductive process.
\textsuperscript{22} Often this ‘completeness’ is probabilistic of a naïve variety.
which are knowable, although the means of getting to know them is never specified. These rational preferences are, often, represented by a utility function, which is assumed to be well-behaved. Third, the non-satiation assumption promises that the satiation point will never be reached, at least in the economic domain. Thus, the individuals are always in the state of the world where “more is better”.

In passing, it could be mentioned that there have been serious and contentious discussions in the history of the development of economic theory as to whether utility should be cardinal or ordinal, since these might consequently result in differences in the way in which economists try to measure utility. Eventually, ordinal utility seems to have reached dominance, although not very ‘consistently’; subsequently the theory of individual decision making based on preference and choice based approaches were developed.

### 2.1.2 Game Theory

Game theory is a mathematical field in economics dealing with the situations when players gain payoff by interacting with each other strategically. Its origin can be attributed to von Neumann and Morgenstern (1947). Orthodox game theory is also driven by “self-interest” and “utility maximizing” concepts, however, in order for the analysis to be tractable, infinite powers of reasoning and, often, a kind of common knowledge among players are further assumed. Together with these criteria, different notions of Nash and other kinds of strategic Equilibria, depending on the nature of games which are under consideration are defined and claimed to be able to be determined. The principles of finding Nash Equilibria include minmax criteria along with backward induction for repeated games or extensive games. Moreover, the Nash Equilibria in a game can be both pure strategies (the certain strategies) and mixed strategies (a set of probabilities over strategies).

It may be pertinent to add that no game theoretically defined Nash equilibrium is computable and no algorithm which has been claimed to determine it can be implemented without appealing to undecidable disjunctions.

---

23 This is an ultra-naïve observation where we simply report a consensus view. We disagree that game theory, even in its strategic form, originates with either von Neumann-Morgenstern or with von Neumann’s 1928 paper. Our alternative history is outlined in several of Velupillai’s recent papers on computable economics.
2.1.3 Mathematical Finance Theory and Recursive Methods

A huge amount of mathematical theories and tools have been borrowed to develop finance theories and time series analysis. In these exercises, different stochastic or random processes are imported to represent the data generating process of finance or economic time series, e.g. Brownian motion or Markov chains. The random processes applied here are based on measure theoretic concepts.

Recursive methods in macroeconomics are built on dynamic programming, Markov decision processes and Kalman filtering and again, measure theory, underpinning orthodox theories of stochastic processes and probability, plays a central role – all within one or another form of nonconstructive and non-recursion theoretic real analysis (for example for dynamic programming the notion of one or another form of contraction mapping in a suitable metric space).

Although the mathematical tools used are much more sophisticated than in non-dynamic methods – but only up to a point, economic entities are still modeled as optimizers (e.g. maximizing present values in intertemporal contexts, Value functions and Euler equations in the context of dynamic programming and optimal control formulations) where it is little realised that the analysis is around uncomputable equilibria (c.f. Ljungqvist and Sargent (2004); Stokey and Lucas (1989) (with the collaboration of Edward C. Prescott)).

2.1.4 Experimental Economics and Neuroeconomics

Experimental Economics appears as tool for examining economic theories in computational, numerical and other obviously implementable ways in which idealized subjects are placed in artificial settings that purport to mimic the theoretical environment. Narrowly speaking, it is not categorized as a branch in economics, instead, it is a methodology for researchers to support or refute specific economic theories. While, broadly, it can be considered to be cohabiting with behavioral economics. This is because – or claimed to be because - what people actually do can be observed in experimental environments, and almost of all the anomalies are found and induced from laboratory environments or field studies. The methodology of experimental economics is heavily based on so-called induced value theory (Smith, 1976). Induced value theory suggests that in the controlled laboratory environment, if subjects are suitably motivated,
experimenters can expect to obtain desirable induced values from choices of subjects on certain economic problems they are given to ‘solve’\textsuperscript{24}. This theory is obtained from non-satiation assumption, and monetary payment is the most commonly used reward for inducing real values from subjects. However, if economic agents are actually applying satisficing principle to the experiments they attend, i.e. they are satisfied by performing decently rather than trying their best or thinking hard in order to get the most reward, then results of experimental economics could be very misleading.

Neuroeconomics is the new extension of experimental economics incorporating neuroscience to obtain the data of brain activity, simultaneously, when the subject is in laboratory environment. It is also viewed as a young subfield of behavioral economics which is believed will be the main focus in the future. A popular claim is the dual system in our brain supervising our judgmental and intuitive thinking, corresponding to rational and emotional behaviors. It provides the technique to collect data in the brain for examining how and when the behavior of decision makers could deviate from rational and consistent behavior. A recent survey can be found in Camerer (2007); Glimcher et al. (2005); Rustichini (2005); a critical view of the claims of Neuroeconomics can be found in Rubinstein (2008).

The linkage of neuroscience and human behavior is seriously debatable, but we will reserve our discussion and contribution to this critique for a later exercise.

2.1.5 Computational Economics

Computational Economics is also an extension of experimental economics from another perspective, i.e. the subjects are not human subjects but software subjects. So far, there are at least two well-developed lines, which are heterogeneous agent models and agent-based modeling, and the survey for these respective lines can be found in Hommes (2006) and LeBaron (2006). A thorough critique of the excessive claims of both these lines – and other strands of supposed computational economics – is given in Velupillai & Zambelli (2011).

Heterogeneous agent models seem to have been inspired by related results on cellular automata modeling in the physical sciences, resulting in unpredictable and complex phenomena generated

\textsuperscript{24} But this is not the search for ‘solutions’ in any kind of ‘problem solving’ context, as in CBE.
by simple interaction rules. The claims in this line of research are as vacuous as those made by agent-based modellers in finance and economics. They both suffer from a serious lack of scholarship and a complete unhinging of their foundations in either serious computability theory or even a familiarity with the fruitful and frontier research in the interface between dynamical system theory, numerical analysis and computability. These interactions were the fulcrum around which von Neumann and Ulam, Conway and Wolfram and Turing (1952) pioneered their studies of emergent complex dynamics in interacting systems with simple rules of interaction.

2.1.6 Subjective Probability Theory

Subjective expected utility theory was proposed by Savage in 1954, between the period in which Edwards wrote his first and second survey papers on behavioral economics (Edwards, 1954, 1961). Savage followed the axiomatizations along the lines proposed by Ramsey (Ramsey, 1931) and De Finetti (de Finetti, 1937), and applied Bayes’ rule for updating the prior probabilities over time.

The idea of subjective expected utility ‘first’ appeared in Modern Behavioral Economics through the work of Kahneman and Tversky (1979), when building descriptive theory of decision making by individuals under risk. Their theory in turn borrowed heavily from Edwards (1962) and who in turn built on Savage (1954). Both Edwards and later Kahneman and Tversky, however do not refer to Bruno de Finetti whose contributions are not mentioned in these two papers. There seems to be some ambiguity while they talk about probabilities in their model and this gets particularly unclear when they refer to decision weights.

“In prospect theory, the value of each outcome is multiplied by a decision weight. Decision weights are inferred from choices between prospects much as subjective probabilities are inferred from preferences in the Ramsey-Savage approach. However, decision weights are not probabilities: they do not obey the probability axioms and they should not be interpreted as measures of degree or belief.”
Kahneman and Tversky (1979), p. 280

In this framework decision weight measures over stated probabilities do not obey the property of

---

25 Ignoring, for the moment, the much earlier work of Edwards, who was more than a mentor to Kahneman and Tversky.

26 Decision weights, in Kahneman and Tversky (1979), are a measure associated with each probability, reflecting the impact of probability on the overall value of the prospect.
additivity. In the Savage-de Finetti framework, the sum of the probabilities over exclusive and exhaustive events adds up to unity. In prospect theory, the sum of the decision weights is considered less than one in most of the cases. However, while they invoke the Ramsey’s approach of inferring these decision weights from choices, it naturally raises the question as to what these decision weights are? Although the propositions of decision weights are derived in the paper, it is unclear how they are different from degrees of belief – although different they must be!

In Edwards (1962), two categories of subjective probability models are introduced: additive and nonadditive ones. In Edward’s elaboration, first of all, subjective probability is a number ranging from zero to one and describing a person’s assessment of the likeliness of an event. Further, it is assumed objective probabilities exist, and they are related to subjective probabilities. Edwards argued that it is meaningless to debate whether objective probabilities can be defined, in contrast to de Finetti’s and Savage’s firm belief that there are no objective probabilities. He goes on to make a distinction between risk and uncertainty. He argues that there are some cases, such as die tosses, which have “conventional” probabilities over their outcomes. Consequently, these events which can be given objective probabilities, are defined as risky; otherwise, they are uncertain. However, both Edwards (1962) and Kahneman and Tversky (1979) considered only risky cases. Tversky and Kahneman (1992) is a revision of prospect theory including uncertain outcomes.

The concept of subjective probability is used ambiguously – to put it mildly - in modern behavioral economics. On the one hand, MBE introduced the idea of personal probability, defining it and mapping it over the objective probabilities in risky choices. This is quite different from the kind of subjective probabilities proposed by de Finetti, and does not necessarily follow the axioms of subjective probabilities.

De Finetti’s works became relatively more familiar only after the series of lectures he gave in the

27 Additivity of probability is defined as follows: If \( n \) numbers of events form a complete set of incompatible events (meaning exactly one of the events has to be true), then the probability of the logical sum (the logical sum of a group of events is true, if and only if, one of the events is true) is equal to the sum of their respective probabilities. Since \( n \) is a finite natural number, so the definition above is more precisely finite additivity. On the other hand, when \( n \) approaches to infinity, it becomes countable additivity.

28 The distinction of additive and non-additive probability, made in Edwards (1962) is that additive probabilities sum up to specific numbers, non-additive ones are not supposed to do so – then, what are they, if they do not do so?
Institut Henri Poincaré and these lectures were published in 1937. For De Finetti, the theory of subjective probability originated from his belief in a subjectivist philosophy. Subjective probabilities of outcomes, for him, are the different degrees of belief regarding the occurrences of events that people possess. These degrees of belief, however, need not be the same for all the people. In an attempt to find admissible ways of assigning numbers to different degrees of belief, de Finetti constructed axioms over events and their probabilities, especially through the logical relations (pioneered by John M. Keynes) of events. By standardizing a random quantity into 1 and 0 representing the truth and falsity of an event and by introducing the coherence criteria, de Finetti derives some basic consequences. The most important one amongst them is the concept of finite additivity, where the sum of assignments over finite events (logical sums) sums to unity. More specifically, for de Finetti, the qualitative criteria regarding coherence appears first, and then, the individuals are allowed to freely attach numbers to their degrees of belief over a complete set of incompatible events (i.e., exhaustive and exclusive events), however, within the coherence constraint. This way, a qualitative idea of coherence is linked to the mathematical expressions of (subjective) probability. The coherence principle demands consistency in assignments, based on the idea that no arbitrary gains should be available for either player by accepting certain books of bet (the Dutch Book argument). In order to satisfy the coherence principle, the sum of probabilities of the event has to be unity (the necessary and sufficient condition of coherence). Besides, it should be noted that Bayes’ conditional probability formula is derived in turn from coherence, and it is not taken as a definition in de Finetti’s theory of probability.

Before de Finetti, Frank Ramsey gave a talk in 1926 and the lecture was published in Ramsey (1931), of which de Finetti was not aware until 1937. Both of them, almost simultaneously but independently, formulated subjective probability as a degree of belief held by an individual and devoted their efforts to axiomatize it. In particular, de Finetti assumed and insisted only the use of finite additivity, because the requirement of coherence implies finite additivity. On the other hand, Ramsey simply and intuitively addressed this issue saying that it is meaningless to discuss infinite events, because he doubted a human being’s capability of handling infinite events.

“[N]othing has been said about degrees of belief when the number of alternatives is infinite. About this I have nothing useful to say, except that I doubt if the mind is capable of contemplating more than a finite number of alternatives. It can consider questions to which an infinite number of answers are possible, but in order to consider
the answers it must lump them into a finite number of groups.” Ramsey (1931), p. 183.

In contrast to finite additivity, frequentists and measure theorists advocate and justify the use of countable additivity (or denumerable additivity, infinite additivity and \( \sigma \)-additivity) by invoking the strong law of large numbers (Borel) and relative frequency in limits. Howson (2009) discusses these issues in detail and supports de Finetti’s idea of finite additivity, but not, in our opinion, in a convincing way.

In particular, we fundamentally disagree with Howson that ‘de Finetti himself would have recommended’ doing ‘probabilistic reasoning … in an informal metatheory consisting of the usual mathematics of analysis and set theory’ so that:

“Deductive consistency and probabilistic consistency are .. subspecies of the same fundamental notion of the solvability of equations subject to constraints: those of a classical truth-valuation in the deductive case, and the rules of finitely additive formal probability in the probabilistic case.” Howson, op.cit, pp. 55-6; italics added.

This is a fundamental violation of every tenet of epistemology and methodology advocated by de Finetti. Moreover, Howson does not seem to realise that it is provably hard to devise procedures to validate ‘classical truth-valuation’.

Being aware of the distinction between finite and infinite additivity, Edwards (1962, p.117) considers the infinite case to be more interesting as compared to the finite case. More recently, Bayesian approaches, together with Savage’s notion of subjective probability, are challenged by empirical evidences that suggest agents are incapable of applying the Bayesian rule to revise their prior probabilities. Case based theory, which is considered as one of the new foundations for behavioral decision theory, bases the probabilities assigned to different events on previous histories regarding similar cases and consequently, adopts a (non-algorithmic)

---

29 It may be apposite to point out that Ramsey’s equally distinguished fellow-Kingsman, a few years later, in his monumental classic on computability theory appealed to the same kind of ‘finiteness’ for the same kind of reason (cf. Turing, 1936, p. 249).

30 Refer Edwards(1962), p.117, italics added:

“As will become clear later, finite event sets immensely complicate the mathematics, and at the same time reduce the value of the model by making it inapplicable to situations in which the set of possible events is infinite. Although this paper discusses finite models below, I consider such models far less interesting than the infinite models,”
frequentist approach for the probabilities. (c.f. Barberis and Thaler (2005); Camerer and Loewenstein (2004)).

2.2 Underpinnings of Classical Behavioral Economics

"If we hurry, we can catch up to Turing on the path he pointed out to us so many years ago."

Classical Behavioural Economics was underpinned, always and at any and every level of theoretical and applied analysis, by a model of computation. Invariably, although not always explicitly, it was Turing's model of computation.

The fundamental focus in classical behavioural economics is on decision problems faced by human problem solvers, the latter viewed as information processing systems, as we emphasise right through the analysis in this paper. All of these terms are given computational content, ab initio. But given the scope of this paper we shall not have the possibility of a full characterisation. The ensuing 'bird's eye' view must suffice for now.

A decision problem asks whether there exists an algorithm to decide whether a mathematical assertion does or does not have a proof; or a formal problem does or does not have an algorithmic solution. Thus the characterization makes clear the crucial role of an underpinning model of computation; secondly, the answer is in the form of a yes/no response. Of course, there is the third alternative of 'undecidable', too. It is in this sense of decision problems that we interpret the word 'decisions' here.

As for 'problem solving', we shall assume, as pointed out many times in this paper, that this is to be interpreted in the sense in which it is defined and used in the monumental classic by Newell and Simon (1972), which is, in our opinion, an application of the theory underlying Turing (1954).

Finally, the model of computation is the Turing model, subject to the Church-Turing Thesis.

To give a rigorous mathematical foundation for bounded rationality and satisficing, as decision problems, it is necessary to underpin them in a dynamic model of choice in a

---

31 The three most important classes of decision problems that almost characterise the subject of computational complexity theory, underpinned by a model of computation -- in general, the model of computation in this context
computable framework. However, these are not two separate problems. Any formalization underpinned by a model of computation in the sense of computability theory is, dually, intrinsically dynamic. Moreover, Decidable-Undecidable, Solvable-Unsolvable, Computable-Uncomputable, etc., are concepts that are given content algorithmically, within a model of computation.

Now consider the Boolean formula:

\[ (x_1 \lor x_2 \lor x_3) \land (x_1 \lor \neg x_2) \land (x_2 \lor \neg x_3) \land (x_3 \lor \neg x_1) \land (\neg x_1 \lor \neg x_2 \lor \neg x_3) \]  
…… (2)

Remark: Each subformula within parenthesis is called a clause; The variables and their negations that constitute clauses are called literals; It is `easy' to `see' that for the truth value of the above Boolean formula to be \( t(x_i) = 1 \), all the subformulas within each of the parenthesis will have to be true. It is equally `easy' to see that no truth assignments whatsoever can satisfy the formula such that its global value is true. This Boolean formula is unsatisfiable. This is the kind of `satisfiability' we ascribe to Simon’s notion of `satisficing'.

Problem: SAT -- The Satisfiability Problem

Given \( m \) clauses, \( C_i (i=1, \ldots, m) \), containing the literals (of) \( x_j (j=1, \ldots, n) \), determine if the formula \( C_1 \land C_2 \land \ldots \land C_m \) is satisfiable.

Determine means `find an (efficient) algorithm'. To date it is not known whether there is an efficient algorithm to solve the satisfiability problem -- i.e., to determine the truth value of a Boolean formula. In other words, it is not known whether SAT \( \in \mathbb{P} \). But:

**Theorem:** SAT \( \in \mathbb{NP} \)

**Definition:** A Boolean formula consisting of many clauses connected by conjunction (i.e., \( \land \)) is said to be in Conjunctive Normal Form (CNF).
Finally, we have *Cook's famous theorem*:

**Theorem:** *Cook's Theorem*

SAT is NP - Complete

It is in the above kind of context and framework within which we are interpreting Simon's vision of behavioural economics. In this framework optimization is a very special case of the more general decision problem approach. The real mathematical content of *satisficing*\(^{32}\) s best interpreted in terms of the satisfiability problem of computational complexity theory, the framework used by Simon consistently and persistently - and a framework to which he himself made pioneering contributions.

We have only scratched a tiny part of the surface of the vast canvass on which Simon sketched his vision of a computably underpinned behavioural economics. Nothing in Simon's behavioural economics - i.e., in Classical Behavioural Economics - was devoid of computable content. There was - is - never any epistemological deficit in any computational sense in classical behavioural economics (unlike in Modern Behavioural Economics, which is copiously endowed with epistemological deficits, from the ground up).

### 3 Classical Behavioral Economics – Further Notes on the Special Role of Herbert Simon

A basic tenet of Simon’s approach to behavioural economics is that the limitations of cognitive processing should be linked, in some formal way, with the definable limitations of computation, subject to the Church-Turing Thesis (without any space or time constraints). The limits of computational complexity, on the other hand, are naturally bounded by the time and space. Behavioral models, in which agents are supposed to exercise rational behavior, whether psychologically more realistically constrained or not, hypothesizing capabilities transcending these theoretical and practical limitations are, for Simon, empirically meaningless. Simon has taken the limits of human cognition into account, transformed into computational complexity

\(^{32}\) In Simon (1997), p. 295, Simon clarified the semantic sense of the word *satisfice*:

"The term 'satisfice', which appears in the Oxford English Dictionary as a Northumbrian synonym for 'satisfy', was borrowed for this new use by H. A. Simon (1956) in *Rational Choice and the Structure of the Environment* ."
measures, for describing agents who make decisions. This is why we are convinced that computable foundations and nonlinear dynamics can both be found in *Information Processing Systems*[^33], the paradigmatic formalization of agents and institutions in the kind of behavioural economics Simon advocated.

### 3.1 Bounded Rationality

The idea of bounded rationality was first proposed by Herbert Simon in the paper titled “A Behavioral Model of Rational Choice”, which was published in 1953. It was further polished and republished with a same title as the much more famous Simon (1955) and was phrased also as “limited rationality”. In the same paper, an example was demonstrated where agents tend to be satisfied by using certain information they have and avoid information they do not really have any means of obtaining in algorithmically meaningful ways. They anticipate something acceptable in the near future without calculating any probabilities or assigning probabilities to prospective future events. Simon further described human behavior as “intendedly rational” in Simon (1957, p.196). The book “Models of Man” collected the papers which he published in early to mid 50s. It is where the phrase Bounded Rationality appeared for the first time, in the introduction of Part IV (p.196). The phrase was, then, much maligned in its uses and misuses, compared to the original definition and formalizations by Simon. Subsequently, bounded rationality became one of the frequently used terminologies of MBE. On the contrary, in Simon’s advocacy, human beings can solve their problems relying on heuristics and intuition without a given model in mind[^34]. Therefore, there seems to be a mismatch between the contemporary interpretation of bounded rationality and its original definitions. In Simon’s point of view, human beings have no capability and willingness to always find procedures to reach the best alternative, even if such a thing is meaningfully definable, or make the ‘Olympian choice.’ Reasoning capabilities, formally defined as algorithmic procedures, are constrained by the limits of computability theory and, at an empirical level, by measures of computational complexity.

[^33]: Agents and institutions and all other kinds of decision making entities, in CBE, are information processing systems which, in their ideal form are Turing Machines.

[^34]: This may well be one way for agents in CBE to transcend the limits of Turing Computability subject to the Church-Turing Thesis. However, we do not subscribe to the view that Simon assumed that the limits of Turing Computability are violable; we believe Simon could have resorted to oracle computations, when necessary, and also formalize via nondeterministic and alternating Turing Machines to encapsulate procedures – heuristics and other similar algorithms – that give an impression to the uninitiated that there are formal means to transcend Turing Computability.
Simon’s definition of bounded (limited, procedural) rationality encapsulates different notions, such as limited attention, limited cognitive capacity of computation, satisficing, and sequential decision making (naturally dynamic) (Simon, 1955, 1956). That is to say, it is not evident and admissible to assume that human begins are able to exhaust all the information and make the best choice out of it. Indeed, the notion of ‘best’ is given content via the formulation of problem solving by information processing systems in what is known in metamathematics as a decision problem. In such a framework one seeks algorithms to solve problems and classifies them as ‘easy’ or ‘hard’ using measures of computational complexity. There is no such thing as a ‘best’ algorithm or a ‘best’ heuristic.

Therefore, the dynamics of non-maximizing agents can be described adequately in the following way. The knowledge we have, and interpretation of the world where we are living in, are associated with our experience and memories. Gradually, our tastes and understanding are constructed. The process of construction is the central pre-analytic, Schumpeterian visionary (Schumpeter, 1954, p. 51, ff), stage in the decision problem. Therefore, the pursuit for stable gain in taste and knowledge also relies on what has been constructed. This is one part of requiring a program to modify itself. The unhappiness and satisfaction which are associated with our aspirations depends on whether the desires are satisfied in terms of our anticipation. The aspiration level expands with satisfaction and shrinks with disappointment. Nonetheless, the memory that is stored in our mind prevents our aspiration level from becoming null. Thus, we are in the loop of unhappiness and satisfaction, a loop given formal content via the structure of a program for a Turing Machine or a heuristic implemented on one of them (Simon, 1991).

### 3.2 Human Problem Solving

The notions of bounded rationality have been encoded implicitly and explicitly into the information processing system which was proposed in Simon et al. (1958) and analyzed thoroughly with detailed recording and interview with human subjects in Newell and Simon (1972). IPSs have shown their capability of solving problems, such as cryptarithmetic, logic, and chess games, algorithmically. In their conclusion, it is suggested that task environments of greater complexity and openness ought to be studied. Thus, we can see that they are on the track of pursuing Turing’s suggested program of research on Solvable and Unsolvable Problems (Turing, 1954).
Simon’s notion of bounded rationality, encapsulated within the formalization of an IPS is, in turn, used in simulating (representing) human problem solving. *Simulation*, even if not precisely theorized in Simon’s monumental work on Human Problem Solving, nevertheless is defined in analogy with the dynamics intrinsic to partial differential functions or their machine embodiment in the definition of the processing of information by a Turing Machine, or its specialized variants. Problem solving is the implementation, via *heuristics*, themselves algorithms, of *search processes for paths* from initial states to the target states.

The complexity of a problem solving process – the complexity, therefore, of the algorithm that is implemented in the search processes from initial conditions to ‘halting’ states – defines its hardness on a well-defined computational complexity measure. This also means that there could be problems that will be subject to the famous theorem of the *halting problem for Turing Machines*.

The methods that a problem solver uses are strongly associated with his or her memory and experience. The accumulated knowledge in the memory will form the heuristics – the current state of the program and its structure - to guide the problem solver him(her)self. Intuition is copiously invoked, and defined computationally and cognitively, in seamlessly leading the problem solver to one or another path at a node, when he or she faces a huge number of possible choices, in the *Nondeterministic Turing Machine* formulation of a problem.

### 3.2.1 Theory of Human Problem Solving

Literally, we need a problem and the problem solver to achieve problem solving, and the problem should be presented, recognized and understood. A problem is faced when one wants to do something about a particular task but does not know what *series of actions* can be done to implement it immediately. The three main factors that characterise problems are the huge size of possible solutions, the dispersion of actual solutions and the high cost of search. The problem space contains a set of elements which represent knowledge, a set of operators which generate new knowledge from existing knowledge, an initial state of knowledge, a problem which is specified by a set of desired states, and the total knowledge available to problem solvers. The problem can be further formulated (represented) by *set-predicate* formulations and *search* formulation.

**Representations** In the former representation, the set of elements includes symbolic objects
which are all possible solutions, not necessarily formally definable. Precisely, the set can be generated by a certain enumerative procedure. Thus, the problem solver will not be given the entire set, rather, is given a process to generate elements out of the set. This is exactly analogous to Brouwerian constructive spreads, arising out of free choice sequences. In a search representation, solutions as elements of a set, have the format of sequences. For instance, a proof of a theory contains a sequence of steps and chess representations contain continuations for some players.

**Task Environment** A Task Environment describes the attributes that are associated with the problem that problem solvers encounter. It consists of external and internal representations, where the former is the format in which the problem is exactly presented and the latter stands for the subjective representation the player applies. Accordingly, not only the presentation of the current problem, but also the ability and intelligence of the problem solver should be considered. This is because players with diverse abilities may perceive the problem differently. It should be made very clear that in Simon’s framework of human problem solving, as well as in Turing’s considerations of Solvable and Unsolvable Problems, concepts like ability and intelligence are precisely defined, even if pro tempore, in terms of computability theory.

**Information Processing System** The information processing system which is capable of problem solving can be characterized as follows. An IPS is a serial, adaptive (dynamic), and deterministic system which receives input and generates output. It is composed of internal building blocks such as long term memory (LTM), short term memory (STM) and external memory (EM). LTM and STM share identical patterns but are distinguished by their size. LTM can contain all the symbolic objects without limitation, while STM contains only five to seven symbols. The fact of sequential decision making is inherent in IPS; moreover, how a problem solver retrieves objects from LTM to STM relies on heuristic search. This is exactly equivalent to the partial recursive function formalization of computability or a Turing Machine definition of a computable process (cf. Davis, 1958, Chapter 1, in particular, and Part 1, in general; indeed, reading and mastering the foundational mathematics of computability theory simultaneously with an approach to problem solving in the Simon or Turing sense is the best way to understand all the equivalences inherent in all these formalizations.).

### 3.2.2 Heuristics

*Heuristic* is a method of “Rule of Thumb” that serves as a guide in searching. Intuitively, it is
an ability and process to refer to one’s own memory and experience and lead oneself to focus on appropriate subsets of knowledge. Without external help, one can learn and discover new knowledge by him/herself. Essentially, it is the ability of a Machine to reconstruct its internal structure by itself.

When an IPS receives information form the task environment, it generates the goals and the methods for the achievement by heuristic search. If heuristics cannot achieve a satisfactory solution, then either the heuristic method will be reprogrammed or the representation, namely, the internal representation in the task environment, will be reformulated. It will be clear that ‘satisfactory’ here is precisely defined by means of time and space computational complexity measures. In short, IPS and task environment are interdependent, and the process of change is learning. This is one way the human problem solver as a learner encounters him/herself as a learning machine.

In addition to bounded rationality and satisficing, Simon uncovered an interesting property, which became a recurring theme in his works, observed in many entities. In 1951, when Simon read Goodwin (1947), it inspired him to think about dynamical systems in both economic and mathematical senses. Later, the concept of Near Decomposability, culled from Goodwin’s notion of unilateral (weak) coupling, appeared in his papers and he applied it to diverse problems, such as identifying causality, counterfactuals, aggregation, organizational behavior, evolution of organisms, human and machine thinking. Near decomposability has its rigorous mathematical definition and characterisations, while conceptually, the idea can also be connected to heuristics. Especially, Simon explains in the paper Simon (2002) why near decomposability, appeared as hierarchical structure of an organism, can result in greater speed of evolution. When the hierarchical structure is applied to the problem and problem solver in human problem solving circumstances, then evolution is analogous to learning and discovery.

Near decomposability in human problem solving can be interpreted as decomposing a problem into subproblems when the subproblems are not completely independent. In Polya’s little book Polya (1945), “heuristic method” was demonstrated by an educator decomposing and reformulating a problem step by step for a student who is asked to solve the problem. Turing, at the same time, also proposed his idea of a child’s machine and education process in Turing (1950, p.456). The influence of Polya, Turing and Goodwin are unambiguously evident in Newell and Simon (1972); and in Simon et al. (1958) for their postulation of the internal
structure of minds and the representation of task environments in human problem solving.

Appendix to 3.235: Heuristics and Problem Solving

Archimedes felicitously separated the question of discovery from that of proof; the former could be done experimentally – both in the classical sense and in the sense of thought experiments. For the latter, he used both the technique of proof by reductio ad absurdum and ad contradictionem. It was only with the formalism of problem solving that this separation was healed. But that had to wait till Brouwer challenged the latter day Platonists in the early 20th century, Kolmogorov re-integrated solvability with proof and Church and Turing put it all together in one fell swoop as a computational paradigm.

There were many who lit the path between Archimedes and Simon, none more so than that father of the Bohemian Enlightenment, Bernard Bolzano. He resurrected the Aristotelian triple division of methods of inquiry into induction, deduction and retroduction. The latter as the apagogic procedure identical with reductio ad absurdum in the context of what he called HEURETIC (cf. Bolzano, 1972, Books Four and Five). This became, in the hands of Polya and Simon, heuristics – the art of guided search in a complex space.

One part of Bolzano’s methodology that is reflected in Simon’s practice comes via the influence the former had on Polya’s approach to problem solving. The other part of the influence, perhaps via Norwood Russell Hanson, was in Simon identifying the logic of discovery with the logic of retroduction in his forceful and uncompromising critique of Popper’s nihilism regarding the feasibility of a theory and logic of discovery:

“This mystical view [i.e., Popper’s view] towards discovery …. has not gone without challenge. Peirce coined the term ‘retroduction’ as a label for the systematic processes leading to discovery;

---

35 This appendix, outlining the origins of the notion of heuristics and reproducing pp. 181-2 of Velupillai (2000), is included to circumvent and prevent, if possible, questions raised about the role of a problem solving approach to economic theory and decision theory. More particularly, during the presentation of an earlier draft of this paper by the first author, questions by people who have either not read Simon on problem solving and its role in enriching economic decision processes, or who have no familiarity with constructive mathematics (or, more likely, both), made it clear that at least a mild clarification is necessary. The inclusion of the section from the defining work of Computable Economics here is also to emphasise that problem solving is the fulcrum around which constructive mathematics, computability theory, computable economics and classical behavioural economics are brought together – and that it was a fulcrum that has been around for at least about a decade and a half.
while Norwood Hanson, in his *Patterns of Discovery*, revived that term and gave us a careful account of the retroductive path that led Kepler to the elliptical orbits of the planets. … Hanson made his case for retroduction by examining historical examples of scientific discovery. He did not propose an explicit formal theory of the retroductive process .. It is the aim of this paper to clarify the nature of retroduction, and to explain in what sense one can speak of a ‘logic of discovery’ or ‘logic of retroduction’. Like Hanson, I shall proceed from examples of retroductive processes…”


Bolzano confined *his* application of the retroductive process, i.e., *the apagogic* procedure, to the *mathematical question* of ascertaining the truth of a proposition. However, ‘ascertaining the truth of a proposition’ was an instance of ‘problem solving’, which is why it entered Polya’s scheme of things seamlessly. Bolzano’s criteria for the solution of a problem, the successful demonstration of a proposition is predicated upon two requirements, even before the thought experiment is conceived: firstly, the proposition to be demonstrated must be chosen with care; secondly, it must be tested for its consequences in every conceivable way. Simon adheres to these two strictures almost religiously and his sustained criticism of ‘armchair theorizing’ reflects the importance he gives to the second of Bolzano’s criteria for the consideration of a proposition as worthy of attention. He goes beyond Bolzano, of course, because he goes that extra distance with the testing of propositions: not only as thought experiments; but also as components in real experiments. He also goes the extra distance in suggesting a theory for the careful selection of the relevant proposition, and that theory, too, is computational.

Bolzano’s testing proceeds as follows:

“If the mere clear representation of a proposition M does not lead to a judgement about it, or if this judgement does not appear reliable enough, the next stage in its *testing* is that we attempt to deduce, either from M alone or from M together with other already known premises, several consequences and from these further consequences, etc. … This procedure of showing the truth of proposition M, and thus of *solving the indicated problem*, is generally called the *reduction to absurdity*, or apagogic procedure. Examples are common in the mathematical sciences…”

ibid, p.373; first set of italics added.

Turing’s characterization of the computable numbers can be interpreted as distinguishing between numbers that are defined by *pure* existential statements and those that can be algorithmically defined. This is why standard mathematical economics is replete with existential theorems without the slightest concern over their constructive or algorithmic status. This is particularly true of modern behavioural economics, even when facile claims on heuristics are made. A subject whose fundamental entities are defined over the whole of the real number

36 Not in the preposterous Popperian sense of attempts to falsify it.
system will, naturally, not worry too much about the subset that is numerically meaningful in a constructive or algorithmic sense. It is also why economic theory has not been a pleasant playing field for those of us who would like to interpret the cardinal aim of the subject to be problem solving and who would, therefore, insist on characterizing rational agents as meaningful problem-solvers and would also indulge in hair-splitting about the importance of methods for solving problems.

The skeptical economic theorists might not find the above paragraph particularly convincing. What is the difference between problem-solving on the subset of the computable numbers or constructive numbers and problem-solving on the whole of the real number system, the skeptic may ask? There is a fundamental and foundational difference, which is best captured in an important theoretical contribution made by Kolmogorov, a long time ago:

“In addition to theoretical logic, which systemizes the proof schemata of theoretical truths, one can systematize the schemata of solution of problems. ….

…. The calculus of problems is formally identical with the Brouwerian intuitionistic logic, which has recently been formulated by Mr Heyting. … [I]t follows that one should consider the solution of problems as the independent goal of mathematics (in addition to the proofs of theoretical propositions).


Our\textsuperscript{37} belief – imparted to Velupillai by his teacher, mentor and friend, Richard Goodwin – is, therefore, that economic theory should be about problem solving\textsuperscript{38}.

3.3. Classical Behavioral Economics and Computable Economics

3.3.1 Satisficing, SAT and Diophantine Problems

The ‘senior’ author of this paper advocated that the faithful encapsulation of Simon’s bounded rationality and satisficing ought to be through models of computation in the context of decision problems. Particularly, he suggests posing problems of rational choices as SAT problems (satisfiability problem) (Velupillai, 2010a). A SAT problem looks for the truth assignments of

\textsuperscript{37} More strictly, Velupillai’s belief, in this context and given his background, is what is meant here.

\textsuperscript{38} A view to which even Lucas (2009) seems to subscribe, belatedly, not, of course, in classical behavioural economics mode, nor in any kind of constructive or computable mathematical frameworks. This analogy is very similar to the way Simon (1978, p. 17, footnote 2) referred to Becker’s notion of ‘irrationality’ (Becker, 1962), as follows:

“What Becker calls ‘irrationality’ in his article [Becker (1962)] would be called ‘bounded rationality’ here.”
the arguments which can make the global statement true. If such assignments can be found, then the SAT problem is satisfiable.

Solving SAT problems can be formulated, equivalently, as linear Diophantine equations, linear systems with nonnegative integer variables, or integer linear programming problems. Theoretically, SAT is NP-Complete (Cooke’s theorem), that is, a SAT problem is not solvable in nonderministic polynomial time in its inputs, but can be verified in polynomial time. However, the ‘senior’ author has realized very recently that Simon’s notions should be better formalized in terms of space computational complexity. In particular, SAT can be solved with a linear space algorithm. An intuitive explanation might be that in real human problem solving, subjects are never given sufficient amount of time to make decisions, rather, they are trained to restructure their short-term memory in order to process a problem in a given period of time. Subsequently, Velupillai has proved, via a duality between computability and dynamic systems, that Simon’s information processing system is capable of computation universality which is the relevant model of computation for rational choice. Furthermore, orthodox notions of rationality (through optimization) has been shown as a special (easy) case of the more general (difficult) case of SAT problem, in terms of models of computation in a decision problem context.

3.3.2 Chess and Go

Like many other strategic games, though the final target is to defeat the opponent in one’s own way, Chess players care about many other actions while the game is ongoing. For example, it is important to capture, block and otherwise threaten the opponent. These are the sub-goals that come to players’ mind alternatively, simultaneously to playing the game with the global goal, and in the pensive phases between moves. Being aware of the sub-goals, players can reduce their attention to relatively small groups of good moves and play accordingly.

Go and Chess are very fundamentally different. GO has no concrete configuration of terminal conditions, like “Check-mate” in Chess. Instead, a GO game is finished when both players pass, and the side who occupies greater territory wins. This is a most intricate ‘stopping rule’ for the program to implement the process of playing GO by a Turing Machine. The best moves in the GO games are even more meaningless than the ones in Chess. Similarly, though it is difficult to list out all the terminal positions in Chess, it is very possible to decide whether each configuration belongs to the set of Check-mate. It is only possible for some of the games of GO.
Unlike Chess, GO players rarely benefited by playing forcefully or aggressively – assuming these concepts can be given formal definitions in the relevant mathematics - because by doing that they can create unforeseeable ‘dangerous’ configurations to their own groups as well.

A Go game can be officially played on a 9 x 9, 13 x 13, and 19 x 19 board. Practically, Go games can be set from 2 x 2, 3 x 3,…, 9 x 9,…,13 x 13,…,19 x 19,…boards. The combinatorial complexity increases exponentially when the board sized is enlarged. Thus, the complexity of GO games can be expanded theoretically to countable infinities, of a kind. This is the flexibility that the Chess game may lack.

The main task in playing GO is to enclose some areas on the board, so that the stones of the opponent which are in this area have no space to escape and are captured. On the other hand, when a group of stones are in danger of being captured, the task is to create holes (eyes) to save a region. No matter how big the board size is, the warfare will be localized into separate regions on the board. When the game is being played, the attribution of some regions can be determined and it is known for both players that there is no need to fight on those regions any more. That is to say, the players will decompose the board into several blocks and try to invade or defense those regions. We conjecture, therefore, that near decomposability will turn out to be a useful way of representing some configurations in a game of GO.

In formalizing the Go games, it is reasonable to start with smaller sizes and apply them to bigger board in the idea of blocking. In spite of the fact that the complexity of a game of GO increases exponentially with the board size, human players can reduce the practical complexity drastically by decomposing the board configurations and attack them separately. However, the GO board can never be partitioned unambiguously, this is where a plausible application of near decomposability can be envisaged. Despite all the differences of the two games, there are important similarities, too; GO players need to come up with sub-goals, such as Joseki, creating Atari, making eyes and escaping from being captured etc., in order to resolve some situations.

4 A Brief Conclusion

Science, in most of the cases, is built on asking and answering – often unanswerable - questions. In order to proceed properly, it is critical in most of the cases, that appropriate questions be asked. Decision theory deals with the problems of human choices, and plenty of models have
been constructed and examined through the formalizations of orthodox mathematical economics, econometrics or experiments. Nevertheless, behavioral economics emerged based on the failure of orthodox economic frameworks. Anomalies have been collected and discovered with respect to the normative human behaviors which are predicted by orthodox economics. The central doctrines of orthodox economics are optimization subject to constrains and equilibrium analysis. Modern behavioral economics emerged as a field of finding and explaining anomalies in human decision behavior. The difficulties of solving these problems (optimization and equilibrium) have been noticed; however, their solvability has not yet been questioned and challenged in Modern Behavioral Economics.

Solvability of problems, by problem solvers, requires formal characterisations of both concepts, neither of which has ever been attempted by modern behavioural economists. They are almost defined and characterized in classical behavioural economics and computable economics, as we have argued above.

Herbert Simon introduced the notion of “bounded rationality” and “satisficing” into economic fields along with their psychological and computational underpinnings. Intuitively, computability theory tackles the solvability of a problem and computational complexity theory measures the difficulty of solving a problem. Thus, if a program is designed to mimic human thinking, naturally, the computability of a program has the counterpart in reasoning. Simon’s ideal models of economic agents can be demonstrated by an Information Processing System and its nature of adaptation can be captured in the theory of “human problem solving”. Within this framework, “anomalies” are, possibly, those that result in uncomputabilities, undecidabilities and unsolvabilities of problems, forced into solvable modes by inappropriate models, precisely definable as, for example, the use of finite automata where a Turing Machine is required, and so on.

If Simon’s postulations are taken into account, then “Olympian” rationality (coined in Simon (1983, p.19)) is merely the special case of bounded rationality, and an optimization problem is, again, the special case of a satisfiability problem (satisficing), within the formal framework of metamathematical decision problems.

Apart from making, hopefully, clear distinctions Modern and Classical Behavioral Economics, a more faithful encapsulation of Simon’s notions – with clear computable underpinnings – was
presented in this paper. In continuing work, we are expanding the scope Simon’s notions of bounded rationality and satisfying, within a formal computable formulation, an exercise already begun in Velupillai (2010) to the more general and complex cases of combinatorial game theory. Studying, for example, boundedly rational agents, choosing satisfying strategies in a game of Go will, we think, form a meaningful milestone in research along this line.

It is even possible to interpret some strands in Simon’s thinking that human beings do try to solve the formally unsolvable problems, even while they somehow find ‘only’ the methods (heuristics) to satisfactorily solve them. That is to say, they try to make good decisions for only the near future, but with long-term targets in mind. No actual agent in his or her right mind (sic!) would even dream of formulating infinite horizon optimization problems in the economic sphere, except of course those endowed with Olympian notions of rationality, solvability, computability and decidability.

REFERENCES:


Bolzano, Bernard (1837 [1972]): Theory of Science – Attempt at a Detailed and in the main Novel Exposition of Logic With Constant Attention to Earlier Authors, Edited and Translated by Rolf George, University of California Press, Berkeley and Los Angeles.


Camerer, C. F., G. Loewenstein, M. Rabin (eds.) (2004), Advances in Behavioral Economics,


Turing, A. M (1950), Computing Machinery and Intelligence, Mind, 59(236), 433-460.


