Microfoundations, Behaviour, and Evolution: Evidence from Experiments[∝]

Francesco Bogliacino
Universidad Nacional de Colombia
For correspondence: Carrera 30, No 45-03, Bogotá (Colombia)
email: fbogliacino <at> unal.edu.co

Cristiano Codagnone
Università degli studi di Milano, Universitat Oberta de Catalunya

Abstract

The article discusses whether and to what extent experiments can contribute to a research paradigm based on the study of human behaviour in complex evolving environments and on the problem of asymmetric adjustment among different components of economic system along certain trajectories, focusing on the possibility that experimental evidence may represent an external consistency check on this type of heterodox modelling.
It considers the evidence on rationality of human agents, and the possibility to identify a microfoundation alternative to *homo oeconomicus,* discussing the evidence on humans as strong reciprocators, as trusting individuals and as embedded in social norms.

Keywords: Experiments; Causality; Heuristics; Learning; Bounded Rationality; Altruism; Punishment; Trust; Norms
JEL Classification: C9, C18; D01, D03,

[∝] A previous version of this paper circulated with the title "Microfoundations for Structures and Evolution: Evidence from Experiments", we change it as requested by the guest editor. This paper is forthcoming on Structural Change and Economic Dynamics. We thank the guest editor, Mario Pianta, for comments, and two anonymous referees for their suggestions and critiques that helped us a lot to improve this article from the first version. This work is grounded in endless discussions with colleagues and friends. A special thanks should go to Gianluca Grimalda, Giuseppe Veltri and to all the attendants to the *Seminario Experimental* at Los Andes. Furthermore, we acknowledge useful inputs from those attending the Workshop "Explaining Economic Change" at Rome La Sapienza.

1. Introduction

Rational choice theory is usually prone to paradoxes. Take the example of coordination, discussed by Hollis and Sugden (1993). Imagine that two persons (agent 1 and 2) should decide independently where to meet; both prefer a successful date to lack thereof and, between the two possible meeting places, both strongly prefer A to B. Common sense would dictate that they would meet at A. However, meeting at B is a Nash equilibrium since nobody has an incentive to deviate from B as a response to the other going to B. Furthermore, it is very difficult to *rationally* (in a mathematical precise sense) eliminate B as a strategy since excluding B for agent 1 requires to hold a belief that the counterpart would never go to B, which could only be supported on the belief that the agent 1 would never go to B in the first place, which brings us back and forth in an infinite regression. What happens if we leave axiomatic reasoning aside and ask people to make such a decision? Unsurprisingly, they go to B (Camerer, 2003).

This alternative approach to the theory of choice and game theory stands on a methodological tool, called experiment, or Randomized Control Trials (RCTs). This quest for stylized facts is, in our opinion, very similar to the approach of evolutionary theory, which defends their agent based modelling because of an external consistency with empirical evidence. In this special issue Dosi and Virgillito (2017) argue in favour of alternative microfoundations for human behaviour, and we claim that RCTs are a key tool to answer this question.

Some of those intuitions are far from new. Ashraf et al. (2005) recognized that most of the intuitions behind bounded rationality in individuals (from dual thinking process to loss aversion, from overconfidence to concern for fairness) could be actually traced back to Adam Smith, in particular to the Theory of Moral Sentiments (Smith, 1759). Nevertheless, there is an added value in supporting these classical intuitions with robust evidence. In fact, RCTs provide one advantage with respect to other empirical evidence. It is cumulative and aimed at *replicability*. In fact, observational data regarding choices and behaviour can always be interpreted in different and possibly contrasting ways. As Smith (2010) claims, decisions are as sensitive to small variations in context as they are to structural elements and rules of the game (in the case of strategic interaction). This goes back to the intuition by Hume (1739) that we tend to interpret actions that may be influenced by "circumstances unknown to us" as pure motivation. Ultimately, hypothesis testing implicitly includes further auxiliary assumptions (Duhem-Quine thesis; Quine, 1951), e.g. context independence or the existence of a common mind frame between the experimenter and the participant (*anthropotheorization*, as stated by Smith, 2010). Only with the kind of cumulative, trial and error, case-by-case work that experimentalists have been developing, we can understand what kind of auxiliary assumptions may be at stake and what kind of circumstances are intervening to affect human behaviour.

This holds true, provided that we understand correctly the theoretical domain in which causal identification takes places and the limits that are implicit in it (Heckman, 2010). Since this

methodological discussion is very technical, we prefer to include it in the Appendix of the article.

In this article, we will present a selective number of empirical *facts*, which could be helpful to heterodox scholars, especially those involved in the study of human behaviour in complex evolving environments (Dosi and Virgillito, 2017) and on the problem of asymmetric adjustment among different components of economic system along trajectories (both observed or "required", e.g. for full employment) of structural dynamics (Scazzieri, 2017).

Experiments focus on individuals, in isolation or in interaction, but this is not overlooking the role of *structures*. On the one hand, experiments must include institutional settings, but the *aggregation* of behaviour is fundamental because it allows studying emerging order as a property of a certain context (or institution), beyond individual motives. This is actually one of the main lesson from the *founding father* of experimental economics, Vernon Smith. On the other hand, agency (as opposed to structure) is a necessary object of study if we want to learn how structures evolve, and not only how they affect behaviour. In the specific case of human cooperation, which we will discuss at length, we must admit that free-riding is a possibility and is evolutionary stable, thus the understanding of how it became possible to see large scale cooperation in human, require to see how we can institutionally control free riding when this strategy is available.

Dosi et al (2005) in their quest for the foundations of a theory of evolutionary learning, among other important contributions, make four points that are relevant for our purpose here, namely that: (i) evolutionary theories should systematically use evidence from other cognitive and social sciences as 'building blocks' for the hypotheses on cognition, learning and behaviours that one adopts (p. 263); (ii) it is important to consider the link between the framing effects and processes demonstrated by cognitive psychologist and social embeddedness (*ibid*., pp. 265-266); (iii) the heuristics and biases object of the Kahneman-Tversky (K-T) apply both to one-off decision and learning processes (*ibid*., pp. 266-267); (iv) establishing the evolutionary microfoundations of socio-economic change is patently a very difficult challenge, also a result of lack of robust models, stories, and evidence in the various disciplines that should provide the building blocks (ibid., p. 272). Looking at experimental evidence responds to the first point in Dosi et al (2005) and, we believe, contribute to this research programme by shedding light on two subjects that are key to the evolutionary/structuralist research program: a) the presence of limited cognitive ability (*bounded rationality*) and its implication for evaluations, choices and learning (with specific emphasis on heuristics and biases); and b) the analytical basis to explain human behaviour, (i.e. the correct anchor of microfoundation) in a way that enable linking eminently cognitive aspects (i.e. such as framing and others) with the issue of social embeddedness. Once accepted that both over-socialization (structure driven), as in standard sociological approaches such as functionalism and structuralism, and under-socialization (agency driven), such as *homo oeconomicus*, are bad reference points, behavioural experimentalism can be

used to go beyond such dichotomy.[1] On the other hand, obviously the experimental evidence presented cannot alone tackle the daunting task of building the evolutionary microfoundations of structural evolutionary change; it provides insights that can help explain emergent properties ruling out 'anthropomorphic' interpretation of dynamics (Dosi & Virgillito, 2017), as long as they are integrated within further theoretical reasoning and complemented by other sources of empirical evidence. For this reason, as we further substantiate both at the beginning of sections 3 and 4, we present and discuss the experimental evidence as a sort of 'neutral' input without entering into, for instance, the theoretical (i.e. are deviations from the standard rational choice axioms or rather manifestation of 'ecological rationality' as opposed to 'constructivist rationality'[2]) and normative (i.e. whether new welfare criteria can be established once behavioural evidence show preferences to be context dependent[3]) debates surrounding experimental evidence on heuristics and biases. The insights from behavioural experimental evidence can either be used in the attempt to amend and make the standard rational choice paradigm more 'realistic'[4], or rather to depart and radically change the paradigm. Put it sharply, we are not interested in discussing such dichotomy, not only because it would be beyond our scope and space here, but also because the evidence presented is sufficiently robust that theorists aiming at external consistency validation (Dosi and Virgillito, 2017) should consider it and use it regardless of ongoing theoretical and normative debates surrounding such evidence.

In trying to define RCTs, we think that it is better to follow the economists' practice (i.e. as opposed to the experimental practices of psychological disciplines). In economics, experimentalists faced strong opposition (sometimes even radical, as in Harrison, 1982) forcing them to develop strong protocols and procedures, e.g. avoid deception, use of performance related measurement, and better transparency with regards to replicability and data access (Guala, 2005; Friedman and Sunder, 1994).

We define an experiment as the measurement of human behaviour under controlled environment. Smith (1994) defines it as the conjunction of three elements: an environment, an institution, and a behaviour.

The environment is a complete specification of endowments, technologies, and preferences available to participants. Of course, some of the elements of the environment, such as the preferences, are not observable. The original intuition by Smith (1976) was to develop an

[1] The social embeddedness cited in Dosi et al (2005), was formulated by Granovetter (1985) exactly as a way to move away from the dichotomy that Wrong (1961) had identified between the 'cultural dopes' of structural-functionalism (over-socialised conception of man) and the under-socialised and atomistic homo oeconomicus.
[2] We use here the expressions as presented in Vernon Smith Nobel Prize lecture as published in the American Economic Review (2003), although the paradigm of 'ecological rationality' was first presented by Gigerenzer et al (1999).
[3] See a discussion of the concept of 'preferences purification' in Infante et al (2016).
[4] In this respect see the critique to mainstream behavioural economics as accepting the normative benchmark of rational choice and replicating an 'as if' approach in disguise (e.g. Infante et al 2016; Berg & Gigerenzer, 2010).

induced value approach, where incentives were used to provide the value/cost profiles that fitted into the research question. In the original paper, the proposal for the theory of induced value rests only on the principle of non-satiation (i.e. higher reward is always chosen with respect to lower reward), but current versions of the theory propose three axioms (Friedman and Sunder, 1994): (1) the subject must prefer more reward to less (*monotonicity*); (2) the reward depends on the action (*salience*); (3) influence of other aspects on the subject's welfare from participations are negligible with respect to the reward itself (*dominance*); if possible, one should let the *explicit* variation in reward to be the only source of variation (Smith, 1976).[5] As a result, a set of good practices could be enumerated, going from no deception, to simplified and neutral language, and of course to the use of monetary reward (which makes monotonicity easier).

Institutions are full specification of language and communication rules, and how contracts are made. Finally, behaviour is what the participant brings to the lab, and of course, is changing in response to environment and institutions.

In RCTs, the study of a specific institution or environmental element involves its controlled variation following standard experimental design (Friedman and Sunder, 1994), while confounding factors are either fixed or randomized by design, controlled by measurement or, at worst, assumed away by thought experiments, depending mostly on feasibility constraints (Camerer, 2003).

Nevertheless, one of the main concerns of experimentalists is that, while a controlled environment allows reproducing the specific assumptions of theories to be tested, they may not reproduce data generating process in the field (Smith, 1989).

This has popularized the use of field experiments in recent times. Harrison and List (2004) propose the following taxonomy.

(1) Artefactual field experiments are lab experiments with "field" samples. This is mainly a way to address problems of double selection (into school and into experiments) that characterize student samples, although a similar selection in principle exists in the experimental site. There is now a widespread feeling that too much inference has been grounded on samples of WEIRD people (Western, Educated, and from Industrialized, Rich and Democratic countries; Henrich et al. 2010). In many situations, participant from this cultural *milieu* occupy the extreme end of the human distribution of behaviour (Henrich et al. 2010).

[5] Smith (1982) suggests that to provide rigorous testing hypotheses, one would need also *privacy*, i.e. only information on alternative own rewards, to avoid other regarding preferences as violation of induced value theory. Since other regarding preferences are today one of the main topic of interest, this is not normally followed. Besides, in games of strategic interaction common information is necessary as a minimal requirement to common knowledge postulated by standard theory (although the two do not coincide -Smith, 1989- because participants can freely map common information into their level of knowledge).

(2) A framed field experiment is an artefactual field experiment but with "field" context with respect to information, task or commodities. The point made by Harrison and List is that a context-free environment may not constrain the elements that participants may transpose from one setting to the other. Instead, they claim that context may be very relevant for individual performance, and not providing one may induce the participants to bring one on their own. This is not at odd with the general quest for simplicity or transparency by Smith (1982), since the latter clearly claims that inference about non-laboratory environments requires a principle of *parallelism*, i.e. the holding of a *ceteris paribus* condition.

In their paper, Harrison and List (2004) provide an illuminating example of how the commodity could be part of the way in which the subjects solve the task, from experiments with the Hanoi Tower (originally studied by Hayes and Simon, 1974). The Hanoi Tower consists of a set of disks of ordered radius, which form a tower on one of three pegs. The participants should rebuild the tower on the last peg, moving one disk at a time and never putting a smaller disk below a larger disk. Manual experiments show that participants learn the solution by using violations of the rule, i.e. putting the tower in the final position and reasoning backward (*backward induction*). Computer versions of the experiments do not allow such a move, and provide different results. Their telling conclusion is: "This example also illustrates that off-equilibrium states, in which one is not optimizing in terms of the original constrained optimization task, may indeed be critical to the attainment of the equilibrium status" (Harrison and List, 2004: 1024).

(3) A natural field experiment is the same as a framed field experiment but in a real setting and with subjects not knowing that they are participating. In this case, the argument is that subjects may be naturally helped by contextual cues to interpret tasks and stimuli, which may not work in the lab. Moreover, the participation is suspected to generate a sort of "demand effect", which may increase performance.

In view of the key aspect of the experimental approach anticipated above the aim of this article is to discuss whether and to what extent experiments can contribute to a research paradigm based on the study of human behaviour in complex evolving environments and on the problem of asymmetric adjustment among different components of economic system along certain trajectories, focusing on the possibility that experimental evidence may represent an external consistency check on this type of heterodox modelling. To this purpose, in the following sections, we will discuss experimental evidence with respect to rationality and its bounds (Section 2), and, the issue of the possibility of a reference point for human behaviour as it emerged from experimental evidence (Section 3). Section 4 concludes. In the Appendix we include a general issue of causality and validity related to the interpretation of experiments.

2. Bounded Rationality: Heuristics, Biases, Learning, Dual Process of Reasoning, Degree of iterations

Dosi and Virgillito (2017) cite Simon (1969, p. 267) to present a minimalist definition of the economy as a complex evolving system and elaborate on the concept of emerging properties, which would rule out any form of 'anthropomorphization' as the basis of interpreting any form of systemic dynamics. On the other hand, the work of Simon on bounded rationality is cited by different authors in ways that attempt to improve the realism of the neo-classical view of human action or to radically depart from the standard rational choice paradigm.

While Dosi and Virgillito (2017) seem to propend for the ecological rationality approach to heuristics, in this section we discuss several approaches but we focus mostly on the heuristics and bias research programme launched by Tversky and Kahneman (1974) and, especially, the influence this has had on the development of behavioural economics. In this respect the clarification and disclaimer anticipated in the introduction should be further illustrated and substantiated before proceeding further in our exposition.

First, it is clear to us that one main critique to behavioural economics evidence is that it represents a much less radical departure from standard rational choice theory in that it sticks to, and is defined with respect to, such theory intended as a sort of normative benchmark. Infante el al (2016), for instance, consider surprising that, most behavioural economists moving from the impulse of recognising and demonstrating empirically that individuals' mental processes produce decisions not adhering to the axioms of rational choice, nonetheless stick to the rational choice normative benchmark or the 'inner rational agent' that is hampered by an irrational psychological self. Their critique, however, is from a welfare behavioural economics perspective and motivated by challenging the normative basis for policy interventions (i.e. nudges). This clearly lays outside the scope of our discussion here, although their point that cases of rational choices are in need of an explanation as much as those or irrational ones is clearly relevant. Berg & Gigerenzer (2010), on the other hand, argue that the similarity between behavioural economics and neoclassical economics methodological foundations are greater than usually reported and that the former is another form of 'as if' economics in disguise. These authors argue that behavioural economics fails in its objective of increasing the empirical realism of the discipline for they stick to the expected utility approach and make it even less realistic by adding parameters to it. According to Berg & Gigerenzer, a more rigorously empirical science requires less focus on extending as-if utility theory in view of biases and deviations, and more research on the between decision processes and the environments in which they occur, which is the 'ecological rationality' paradigm. It is probably worth stressing that one might draw a distinction between the original research programme as defined by Kahneman and Tversky and the way it was later mainstreamed into behavioural economics. Originally Tversky and Kahneman, as recognised also by Berg & Gigerenzer (2010), state that the heuristics and deviations from rational choice were to be considered only descriptively and that the normative and descriptive study of choice are separate undertakings (Tversky & Kahneman, 1986). The problematic character of retaining rational choice as the normative benchmark

and the need to actually provide explanation for when individuals behave according to such benchmark was stressed by Kahneman himself (1996).

Yet, it is of marginal interest to us in this paper whether experimental evidence can be easily incorporated into the neo-classical model or rather makes it impossible such reconciliation and calls for a radical departure from it. Whether a heuristic is interpreted as a deviation from a normative standard rational behaviour or as an efficient way of dealing with complex environments, is less relevant with respect to the fact that robust and replicated empirical evidence shows that such heuristic is used by agents, for this is no doubt a potential source of external consistency that modellers can use and integrate with other theoretical and empirical evidence as they deem appropriate for their research programme. Obviously, the 'ecological rationality' in the two slightly different ways as has been used by Smith and Gigerenzer[6] –that one may juxtapose to the standard paradigm allegedly retained by behavioural economics –also bear relevance for the objective of building evolutionary microfoundations of socio-economic change and is mentioned both at the beginning and at the end of this section.

Second, the predominant focus on the Tversky and Kahneman heuristics and bias research programme and on subsequent behavioural economics evidence is also a pragmatic choice.[7] First, behavioural economics has produced a much larger body of experimental evidence and of attempts to incorporate these insights into microeconomics and macroeconomics than the 'ecological rationality' paradigm a la Gigerenzer. Second, the latter approach does not deny the empirical *fact* of heuristics but rather the lens through which we can analyse it.

2.1 Two blades of bounded rationality

[6] In a footnote to his Nobel Prize lecture, with regard to the expression 'ecological rationality' Smith writes that '*After finishing this paper I found that my use of the term below had been used by Gerd Gigerenzer et al. (1999)…*' (2003, p. 465). On the other hand, Berg and Gigerenzer (2010) claim a similarity between theirs and Smith approach to ecological rationality since, in their view, what Smith calls markets and institutions are equivalent to what they instead call heuristics. Although the idea of an ecological rationality as representing the outcome of the interaction between agents and the environments (as radically different from what Smith calls 'constructivist rationality) is indeed similar in both Smith and Gigerenzer, we are less convinced on the analogy made between the heuristics about which Gigerenzer writes and the market and institutions experimentally studied in Smith research programme. The former are cognitive processes analysed at the individual levels, whereas the market and institutions studied by Smith are the rules of the game in the artificial microeconomic system created in the lab. We suspect that Berg and Gigerenzer mean "social order" as produced by human interaction that was the focus of Smith's Nobel Lecture.

[7] In reality this was a point of tension between Tversky and Kahneman as well, at least according to the biography by Lewis (2016). Kahneman was actually more inductive and oriented towards grasping the psychological intuition, while Tversky bothered a lot over axioms. Ultimately, this clash is the same that we find between the ecological rationality and the departure-from-rationality approach (the fact that they published together means that some compromise is possible). For an interesting discussion on the induction versus deduction approach, McCloskey (2005) suggests that the approach of qualitatively theorizing (e.g. existence proofs) plus focus on statistical significance is a deductive approach that is not beneficial to economics insofar it drives efforts away from the "how much?" question that is at the core of *real* sciences such as physics. This statement is essentially the same as that made by Dosi and Virgillito (2017) and Dosi, Marengo and Fagiolo (2005).

According to Camerer and Lowenstein (2003, pp. 3-4), the conviction that increasing the realism of the psychological underpinnings of economic analysis will improve economics on its own terms is at the core of behavioural economics; they actually consider some of the neoclassical assumptions as 'procedurally rational' precisely in the sense specified by Simon. Various models that relax the standard assumptions and incorporate bounded rationality into microeconomics have been thoroughly reviewed by Crawford (2013). Among the models that relax the optimisation assumption but retain intact the 'preferences apparatus' Crawford places, for instance, Simon (Simon, 1955), Cyert and March (Cyert & March, 1963), Newell and Simon (Newell & Simon, 1972), Nelson and Winter (Nelson & Winter, 1982), and Rubinstein (Rubinstein, 1998). A second category of models, according to Crawford, are optimisation-based but modify other assumption of the neoclassical model (i.e. the preferences apparatus, learning mechanisms, etc.), including what he calls a branch of behavioural decision theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991), behavioural game theory (Camerer, 2003), beliefs based learning models (Camerer & Ho, 1999; Camerer et al., 2002), and 'level K' models (Camerer et al., 2004; Chong et al., 2016). These models start from one aspect of Simon's concept of bounded rationality, namely the limitation of human cognitive capacities.

Yet, Simon, as cited in Gigerenzer & Gaissmaier (2011, p. 457), envisaged two dimensions of bounded rationality:

> "Human rational behavior (and the rational behavior of all physical symbol systems) is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1990, p. 7).

Stressing the blade of the 'task environment' Gigerenzer has developed the 'ecological rationality of heuristics' claiming that in complex environments allegedly heuristics can make us smarter (Gigerenzer et al., 1999); Gigerenzer has been the staunchest critique of Kahneman and Tversky's research programme on heuristics and their eventual link to systematic judgement and choice errors (bias).[8] Whereas the common explanation is that heuristics are used to save time at the cost of accuracy, the ecological rationality approach posits that given the tasks the environment poses to actors, using heuristics can have a 'less is more effect' for 'a heuristic is ecologically rational to the degree that it is adapted to the structure of the environment' (Gigerenzer et al. 1999, p. 13). Accordingly, a heuristic is 'a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods' (Gigerenzer & Gaissmaier, 2011, p. 454). The contrast with the definition proposed by Kahneman & Frederick (2002), according

[8] In the notes to the introduction of his best seller Thinking Fast and Slow (2011), Kahneman calls Gigerenzer as their most 'persistent critique' and cite his main critiques and their response (Gigerenzer, 1991, 2010; Kahneman & Tversky, 1996). On the other hand, in their stock taking exercise on behavioural economics Camerer & Lowenstein (2003) considered the 'ecological rationality' critique as a reasonable one and stressed that heuristics can be both good and bad.

to which using a heuristic means judging a target attribute by substituting it with another attribute that comes more readily to mind, is evident, for the latter envisages implicitly the concept of bias and the former does not. An alternative definition, reported in Gigerenzer & Gaissmaier (2011, p. 454) is the one proposed by Shah & Oppenheimer (2008), who suggest that all heuristics rely on the reduction of efforts by various mechanisms: examining fewer cues, reducing the effort of retrieving cue values, simplifying the weighting of cues, integrating less information, and examining fewer alternatives.

2.2 Heuristics, biases, and the behavioural economics experimental evidence

Whereas this stream of empirical work was launched by two psychologists such as Tversky and Kahneman (1974), two stock taking exercises on behavioural economics start by tracing the possible origin of behavioural economics to classical economists such as Adam Smith and other thinkers (Camerer & Lowenstein, 2003; Thaler, 2016). Allegedly, the early psychological realism present in economics was paradoxically side-stepped and basically abandoned as a result of the neoclassical revolution that '*constructed an account of economic behaviour built up from assumptions about the nature—that is, the psychology—of homo-economicus*' (Camerer & Lowenstein, 2003, p. 5). According to Chetty (2015), behavioural economics can contribute to improve standard theory by making it more realistic and 'evidence based', with useful inputs for both microeconomics and macroeconomics. Before entering more deeply into behavioural economics, however, we must look at the key contributions that came from cognitive and social psychology, which contributed first to shaking the axiomatic edifice of rational choice (Von Neumann and Morgenstern, 1944; Kreps, 1988), showing that judgment relies on heuristics and choices are reference dependent (Camerer, 1995; Camerer & Lowenstein, 2003). In the 1960s, cognitive psychology started to use a metaphor of the brain as a processor of information and not as an impulse response mechanism (Neisser, 1967, 1976); starting from such premises, psychologists first, and later behavioural economists, presented empirical evidence on human limits on computational power or will power, which implicitly and later explicitly challenged that assumptions of standard economic theory.

Empirically it has been shown through experiments that human behaviour is heavily context dependent, 'a function of both the person and the situation' (Barr et al., 2013, p. 440). There is no given ordering of preferences but they are constructed (Slovic, 1995), the framing of the situation affects the final choice (Tversky & Kahneman, 1974), the ordering is affected by the endowment available at the timing of decision (Thaler, 1980); present bias and hyperbolic discounting push individual to change planned choices (Loewenstein & Prelec, 1992). The preferences of individuals are defined from a reference point (Tversky & Kahneman, 1991) being it the *status quo* option, the endowment, or some particular option that is chosen after a first phase of coding, in which the multiple alternatives are simplified. Much of the available experimental evidence unravel the 'bayesian rule' of the standard model (Camerer & Lowestein, 2003, pp. 9-10). For instance, researchers showed that people

engage in wishful thinking and other self-serving biases, in which the prior conjectures drive the selection of empirical evidence (Babcock & Loewenstein, 1997; Kahneman & Frederick, 2002). One important dimension in human judgment and decision making that came to the fore later than other limitations to the standard of rationality is the role of 'Affect Heuristic' (Slovic et al., 2002). A strong early proponent of the importance of affect in decision-making was Zajonc (Zajonc, 1980), who argued that affective reactions to stimuli are often the very first reactions, occurring automatically and subsequently guiding information processing and judgment. According to Zajonc, all perceptions contain some affect. "We do not just see 'a house': We see a handsome house, an ugly house, or a pretentious house" (1980, p. 154). He later adds, "We sometimes delude ourselves that we proceed in a rational manner and weight all the pros and cons of the various alternatives. But this is probably seldom the actual case. Quite often 'I decided in favour of X' is no more than 'I liked X'. We buy the cars we 'like,' choose the jobs and houses we find 'attractive,' and then justify these choices by various reasons" (1980, p. 155). Affect also plays a central role in what have come to be known as dual-process theories of thinking, knowing, and information processing. Dual process theories, vividly and accessibly pictured by Kahneman (2011) with the distinction between System 1 and System 2, posit the existence of an automatic and a reflective system. It is conceptually helpful and empirically grounded to recognise that most of the time we do not think too much (i.e. we go in automatic mode) and sometimes we give lots of thought to a problem (i.e. we go in a more reflective mode).

With the possible exception of the 'affect heuristic' and the related literature on the role of emotions (Bogliacino, Codagnone, Veltri et al. 2015), the literature on heuristics remained strictly cognitive, always worked in relation to a rationalistic benchmark from which the ideas of systematic errors (bias) emerged.

As Kahneman put it in his Nobel lecture their research programme aimed at mapping bounded rationality by way of exploring systematic biases (2003, p. 1449). This work demonstrated how individuals employ heuristics such as availability, representativeness, and anchoring and adjustment to make judgments and how they use simplified strategies such as elimination-by-aspects to make choices (Tversky and Kahneman, 1974; Kahneman et al., 1982; Tversky, 1972).

The representativeness heuristic, for instance, has been shown to be used in countless experiments when people are called to make probabilistic judgements: they judge conditional probabilities (conditioning on data or class) by how well the data represents the hypothesis or the example represents the class. Using this heuristic is a time saving shortcut to make a judgement with minimal cognitive effort (system 1), sometime with very negative results for the individuals.

While heuristics are the cornerstone of behavioural economics on *evaluation*, the milestone on *choice* is represented by prospect theory (Kahneman & Tversky, 1979), which constitutes the alternative to Expected Utility theory. Prospect theory is based on reference dependent

(alternatives are valued from a reference point), probability weighting (i.e. probability enter non-linearly into choices) and finally, and most important, loss aversion (losses impact more than gains; Kahneman et al., 1991; Tversky & Kahneman, 1991).

Despite shaking the edifice of rational choice especially for what concerns the preferences apparatus (Camerer & Lowenstein, 2003; Chetty, 2015; Thaler, 2016), behavioural economists still strive to contribute to mainstream microeconomics and macroeconomics, and claim the possibility to retain generality and produce predictions. The behavioural perspective is allegedly able to account for the way in which people "build" preferences, i.e. materially choose (Payne et al., 1992) and, very importantly, how the context shapes this process of preference building (Goldstein & Weber, 1995; Loewenstein, 2000). Indeed, although more realistic, behavioural models are still characterized by generality and relative parsimony of the parameters included, allowing them to be tractable and to offer clear and testable predictions (Camerer & Lowenstein, 2003). It is beyond the scope and space of this essay to review all the potential contribution that behavioural economics can make to microeconomics and macroeconomics, and we limit ourselves to a few examples reported in (Camerer & Lowenstein, 2003), particularly those that may be relevant also from the perspective of structural dynamic of macro-economy as complex adaptive system.

Loss aversion, for instance, is more realistic than the standard utility function over wealth and can be parameterized in a general way, as the ratio of the marginal disutility of a loss relative to the marginal utility of a gain at the reference point (Camerer & Lowenstein, 2003, pp. 4-5). Another example cited by Camerer & Lowenstein (2003, pp. 11-12) are models that attempt to incorporate sub-optimal heuristics based judgement of probability in quasi-bayesian models; individual get some hypotheses wrong and use information incorrectly but otherwise use Bayes' rules. Possible the most interesting from the perspective of complex adaptive system is behavioural game theory and related experiments where learning mechanisms are studied and documented (Camerer & Lowenstein, 2003, pp. 4-5).

Finally, a number of behavioural patterns studied in behavioural economics could be incorporated into macroeconomics (Camerer & Lowenstein, 2003, pp. 31-33) such as for instance: a) rigidities in prices and wages (commonly attributed to generic exogenous factors), could be explained in terms of loss aversion or concerns about fairness (as shown by the experimental literature on gift exchange, Fehr et al. 1998); b) a behavioural life-cycle of savings could hinge upon the 'mental accounts' (Thaler, 1999); c) a Keynesian important concept as 'money illusion' can be empirically substantiated by the findings of behavioural experiments (Shafir et al., 1997).

There have been several attempts to model the deviation from standard models documented by behavioural economics (Crawford, 2013, p. 512). As anticipated, Crawford (2013) distinguish two types: a) those that relax the optimisation hypothesis but retain the preferences apparatus; and b) those deemed 'optimization-based model' that instead retain optimisation but work on relaxing other part of the neoclassical model (i.e. the preference

apparatus) and looking also at learning mechanisms. Optimisation-based model dominate more recent behavioural economics contributions. We focus on the latter for they are potentially more relevant to structural and evolutionary dynamics by focussing on learning mechanisms and, thus, dealing with the dynamics of the system.

Some of the classical contributions seen earlier are cases of optimisation based models that: a) relax assumptions on preferences, incorporating reference dependence (Kahneman and Tversky 1979; Tversky and Kahneman 1991; Bogliacino and Ortoleva, 2015); or b) allow social preferences and reciprocity (Rabin, 1993; Fehr and Schmidt 1999; Bolton & Ockenfels, 2000; Andreoni & Miller, 2002; Charness & Rabin, 2002; Sobel, 2005); or c) drop time consistency, to allow present-biased preferences (Laibson, 1997; O'Donoghue & Rabin, 1999; Frederick et al., 2002); or d) weaken customary assumptions about the accuracy of individuals' models or inferences to allow for various "heuristics and biases" (Rabin 2002).

2.3 Behavioural game theory

Behavioural game theory, on the other hand, can be considered as an 'optimisation-based learning model' that, by relaxing the assumption that players play a Nash equilibrium from the start of play, look for adaptive and learning rules and models that may or may not lead to equilibrium. In this respect, a large range of learning theories have been presented and tested experimentally. In "beliefs-based" adaptive learning models players' adjustments are directly motivated by optimization, even though their beliefs are based on oversimplified models of others' decisions (Woodford, 1990; Milgrom & Roberts, 1991; Crawford, 1995; Camerer and Ho 1999; Camerer, Ho, and Chong 2002). They modify their guesses about what other players will do, based on what they have seen, and choose strategies which have high expected payoffs given those updated guesses. A particular case is experience weighted attraction (EWA), a one parameter theory of learning (Camerer and Ho 1999; Camerer, Ho, and Chong 2002) according to which players respond weakly to 'foregone payoffs' from unchosen strategies and more strongly to payoffs they actually receive. Learning direction theory is a simple alternative where players determines their *ex post* best response and adapt their strategy accordingly. A player is assumed to observe not only his own decision and its realized payoff, but also to have enough information about the game to assess the payoffs that would have resulted from alternative decisions (Selten and Stoecker 1986).

Other models assume heterogeneity in the degree of cognitive iteration: in "Level-K, or LK" model, agents of Level K best-respond to agents of Level K-1(Crawford et al., 2013), whereas in "Cognitive Hierarchy or CH" (Camerer, et al., 2004) models, agents best-respond to all lower level counterparts, weighted by their true proportion.

One recent generalization is Chong et al. (2016) that propose a generalised model nesting a variant of the LK model. This is a strand of non-equilibrium structural model, allowing players to be possibly surprised by their rivals in a situation of interaction, contrary to Nash equilibrium predictions. They posit that agents are heterogeneous in terms of their strategic

thinking capacity. The proportion of players with thinking level k (k = 0, 1, … ∞) is denoted by the density function *f(k)*. These models start with an explicit assumption on how non-strategic level-0 players behave. Higher level players' behaviours are then determined iteratively by assuming that they best-respond to lower level players. Hence, the behavioural predictions of non-equilibrium structural models depend critically on level 0's behavioural rule because higher level rules are defined iteratively based on lower level ones. Currently, there is a lack of plausible models or supporting empirical evidence to determine level 0; there are two rules currently used: a) in one model, level 0 is assumed to randomize uniformly across all possible actions (e.g., Camerer et al., 2004); b) In another model, saliency is used to derive level 0's rule from the payoff structure. Chong et al (2016) use a model based on the hypothesis that level-0 players use minimum avoidance as the saliency principle in choosing their actions. The empirical results found by the authors testing their model on experimental data suggest that the never-worst strategies are more than twice preferred over the minimum strategies. They call this effect the minimum aversion effect.

2.4 Critiques and limits

We can now go back shortly to the critique moved from the approach of 'ecological rationality' and integrate them with some other critiques that can be derived from sociological thinking.

As argued by Gigerenzer & Gaissmaier (2011, p. 452), Kahneman and Tversky research programme has led to the opposition between logic and statistical principles on the one hand, and heuristics as mostly producing bias on the other. People use heuristics but should rather avoid them. According to these authors, in real settings, the rational model does not work, heuristics are rational ecologically, and can produce effective and efficient choices (Gigerenzer & Gaissmaier 2011, p. 453). In another contribution, Gigerenzer makes the basic critique that behavioural economists are even more loyal to the ideal of the methodological individualism than the neo-classical authors they criticize (2015). They go all the way to make us predictably and hard-wired to irrationality (Gigerenzer, 2015, p. 364). Various other authors concur with this vision and stress that behavioural economics rests on an even more reductionist individualism than neoclassical economics (Frerichs, 2011; Gigerenzer, 2015; Rebonato, 2014; Streeck, 2010), as if homo oeconomicus is substituted by 'homo behaviouralis' o 'homo naturalis'. Sociological critiques hold that behavioural economics still search for the micro-foundations of a 'universal nature' and that is inspired by a cognitive universalism obliterating synchronic and diachronic social and cultural differences (Streeck, 2010; Zerubavel, 1997). Cognitive schemas are grounded in culturally, historically, and sub-culturally specific traditions. Observing that our actions can be deliberate or automatic, hot or cold, representing different strategies (or lack thereof) and having different effects is not sufficient and beckons the social, cultural, and historical conditions that either enable or constrain individual actors or groups of similar actors from switching their action strategies today or across time (Cerulo, 2010, p. 121; DiMaggio, 2002, pp. 277-278).

3. Is homo oeconomicus a valid theoretical anchor?

Similarly, with what we state in the introduction and at the beginning of the previous paragraph, we need to introduce a caveat in the interpretation. What we will be presenting in this section is evidence (in controlled experiments) that agents behave in a certain way, and how robust is this evidence with respect to changes in samples, contexts and (artificially introduced) institutions. We will not be arguing that the best way to interpret this evidence will be through multi-arguments utility function (e.g. introducing other people payoff in my own utility function) or some alternative research program. As Smith (2003) argued, one could interpret social phenomena through the lens of rational constructivism, i.e. social facts such as institutions are deducted from conscious human reason, or through ecological rationality, a sort of Hayekian order emerging through some loosely define cultural of biological evolutionism. Although we have our preferred explanation, such a theoretical discussion will be out of the scope of this paper, and has partly been addressed in other mayor contributions (e.g. Cooper and Kagel, 2009). Our main concern with stylized facts is driven by the quest by evolutionary and Schumpeterian economist for "external consistency" of models of complex environment (Dosi and Virgillito, 2017).

Still, there is some limitation to the way RCTs can be harmonized with rational choice approach. One the standard defense of neoclassical economics is that rational choice is merely a *representation* of choices. To the extent in which we want to *test* hypotheses, in RCT we deal with hypothesis as capturing some *procedural explanation* on how choices are performed. Otherwise, prediction would be simply impossible in economics and one would bring the argument to its logic implication and fully adhere to the economics-as-rhetoric interpretation by McCloskey (1983).

3.1 At the root of social order

One of the domain in which experimental evidence proved to be more fruitful is the study of other-regarding preferences, or, to put it differently, to what extent *homo oeconomicus* is a good theoretical benchmark to explain human action.

A typical objection to the question *tout cours* is that utility theory is mute on the "argument" of preferences (Binmore and Shaked, 2010), and as such, no threat would come from empirical evidence showing that people care about others.

A counter objection is that if one takes Walrasian theory as part of the overall neoclassical edifice, then self-regarding preferences do become a typical restriction of the theory (Eckel and Gintis, 2010), otherwise most of the standard results are difficult to sustain. However, the question is a fundamental one at a more general level: on the one hand, the simple fact that *free riding* is a profitable behaviour in many social interaction situations obviously raises the question whether a group of "cooperators" would survive invasion by a homo oeconomicus, to put it in blatant evolutionary terms. The contradiction is apparent in the

sense that evolutionary dynamics penalizes cooperators,[9] but human societies provide ample and systematic evidence of cooperation, both among hunters-gathered and modern societies, at a level and scale which are not common among animals or insects, except for the cases of bees and ants, where genetic proximity is way closer (Fehr and Fischbacher, 2003).

Moreover, the very existence of *order* has been 'tormenting' most of sociology and political science since the Seventeen century. Hobbes (2011) was certainly among the first to understand that political order is not granted and that it is easy to imagine *a state of nature*, where chaos and violence would rule. Traditionally, sociology has come out with a *oversocialized* conception of man (Wrong, 1961), which accounts for compliance with social order through a sort of internalization of *superego*; yet, even from a psychoanalysis point of view the superposition of normative principles on *biological* instincts is supposed to generate discomfort and, as such, creates the space for anomic behaviour. A much more diverse perspective coming from sociology is Parson (1937), where the issue of normative expectation may constitute some underlying principle of norm internalization.

How to deal with empirical evidence in this matter is a very delicate issue. Take the ultimatum game (Guth et al. 1982; Guth and Tietz, 1990). This is one of the most common task used in experimental studies in behavioural sciences: it is a single round negotiation, where a party, the proponent, offers to share a sum of money to a counterpart, the respondent. It is a take-it-or-leave-it offer, thus if the respondent rejects, the outcome is zero for both.

Assume for a moment that money can be considered the scale on which preferences are defined. Would you take a positive offer as evidence of other regarding preferences?[10] A *prima facie* answer would be positive, but it would also be mistaken.

First of all, respondents consistently reject very low offers, thus a rational proponent would correctly anticipate this behaviour and offers positive amounts. This can be easily seen by studying a variant of the same game without the strategic response by the respondent: this is called dictator game and is equal to the ultimatum except for the role of the respondent, which becomes dummy. The evidence shows that sharing offers in the dictator are still positive, but significantly lower than in the ultimatum (Forsythe et al. 1994; Guth and Tietz, 1990; Thaler, 1988).

Secondly, the fact of observing a positive offer by the proponent is not a straightforward violation of self-regarding rational choice in the sense that any split is a Nash equilibrium of the game, while only {sharing the minimum amount; accepting any positive amount} is a *subgame perfect Nash equilibrium.* Unfortunately, backward induction, on which subgame perfect refinement is defined stands on a more stringent concept of rationality (*Common*

[9] Defection is an evolutionary stable strategy in evolutionary game theory. "A strategy is evolutionary stable if a population of individuals using that strategy cannot be invaded by a rare mutant adopting a different strategy" (Axelrod and Hamilton, 1981: 1392).

[10] By "other regarding preferences" we mean any preference formulation where other people's payoff enter as an argument of the function.

Knowledge of Rationality) as formally proved by Aumann (1995). Similar problems are generated by most of the *refinements* of equilibrium concepts, for example in signalling games, which makes the case for experimental evidence even stronger (Camerer, 2003).

Thirdly, one could argue that one shot interactions may be subject to "comprehension" problems, and repetition is needed for robust measurement. Nevertheless, any "good", including other welfare may be subject to diminishing marginal utility, while most likely marginal utility of money is constant (Eckel and Gintis, 2010). As a result, repetition would generate decreasing demand for other players' welfare, under standard assumptions and would confound with comprehension-based explanations.

Fourth, one would wonder to what extent variations in the amount of money at stake induce the same demand of other regarding welfare.[11] Again, larger sums of money may induce people to take the situation more seriously, may be making them more selfish, but they also imply changes in incentives. A very nice experiment by Andreoni and Miller (2002) investigates this issue working on a variation of the dictator game. In their protocol, participants decide on the allocation of different amounts of tokens, but each token may be equivalent to different money units when assigned to their own welfare or the opposite party, e.g. in some cases the relative price of giving could be three, while in the baseline DG it is one. The evidence shows that choices tend to meet the Generalized Axiom of Revealed Preferences (*If A is indirectly revealed preferred to B, then B is not strictly directly revealed preferred to A*), in the sense that they violate the axiom far less than a random choice would do, and choices are razionalizable under simple functional forms for almost half of the participants. If choices between monetary values and other persons' welfare are consistent with GARP, increasing the stakes would decrease the demand of the latter good, but would also induce an income effect, whose direction depends on the Engel curve.

This is just a small set of concerns coming from the problem of evaluation and interpretation of empirical evidence. It is paramount that only experimental design can disentangle these different elements at play, whereas for observational data, all these problems of interpretation would be magnified. Testing against the data, the hypothesis of homo oeconomicus or a rational-other-regarding-agent is far from straightforward, and without experiments identification becomes very difficult.

[11] Discussing the role of stakes is beyond the scope of this paper, since the methodological literature on variations in incentives is large and raises a number of questions. It suffices to say that sometimes the critique stands on a sort of naïve principle that people will necessarily put more effort if this raises a higher reward. As discussed by Camerer and Hogarth (1999), this is a sort of labour theory of production (where the output is the task in the experiment), while we may legitimately assume that other inputs such as experience, cognitive capital etc. are also involved, and one may expect phenomenon such as capital labour substitution. We agree with them that the importance of incentives is an empirical issue and is task related, i.e. should be evaluated depending on the specific experimental procedure.

In the remaining of this Section, we will address three specific issues on which RCTs helped shedding some light: the nature of humans as strong reciprocators, the internalization of social norms and the extent to which people trust each other.

3.2 Strong reciprocators

The first theoretical hypothesis behind the presence of large division of labour and cooperation in humans has been kin selection (Hamilton, 1964), but this has limited predictive power given the large presence of non-kin cooperation. The same goes for the theory of direct reciprocity (Axelrod and Hamilton, 1981): the success of tit-for-tat strategy in repeated prisoners' dilemma may explain altruism in small groups or where the possibility of repeated interaction is significant.

Alternatively, one can consider group selection as a mechanism that explains the survival of altruism (Wilson and Sober, 1994), but the presence of migration and the evolutionary stability of defection in within-group dynamics are forces that play against this type of explanations.

Indirect reciprocity, i.e. the need to seek approval through one's behaviour, has recently been proposed as yet another theoretical alternative (Nowak and Sigmund, 1998). Although the evidence for reputation formation is strong, it does raise the issue of how much people overestimate the long run indirect benefit of image establishment and what could be the evolutionary advantage of truth telling in this respect, when obviously lying is less costly (Fehr and Fischbacher, 2003), although there is experimental evidence that people incur a cost to comply with truth-telling (Gneezy, 2005).

One way out of this dilemma is the hypothesis that humans are strong reciprocators,[12] in the sense of performing altruistic punishment to castigate free riding (Fehr and Gachter, 2002). The threat of being punished may spur altruistic behaviour, supporting cooperation. Of course, this still poses the problem of the evolutionary stability of altruistic punishment, since defection continues to be an evolutionary equilibrium in single populations. However, as pointed out by Boyd et al. (2003), there is a fundamental difference between altruistic contribution and altruistic punishment: the relative disadvantage of the former, with respect to free riders, is independent from the share of defectors, while, obviously, the cost disadvantage of altruistic punishers decreases if free riding becomes less present (because punishment just become a threat and costs are not incurred into). It has been shown that for a set of parameters for which altruism is not supported by groups' dynamics, altruistic punishment does survive, and simple forces such as mutation and conformism may stabilize cooperation (Boyd et al. 2003). Moreover, altruistic punishers can invade a population of defectors, provided that punishment is larger than the cost of contribution (Fowler, 2005).

[12] Models accounting for the origin of strong reciprocity tend to posit concerns for fairness or aversion for inequality. Probably the most common is Fehr and Schmidt (1999), where preferences are defined by two parameters capturing respectively the guilt of having more than the counterparts and the envy of receiving less. Concern for fairness is present in Falk and Fischbacher (2006).

Finally, in recent theoretical development on gene-culture co-evolutions (Gintis, 2007; Henrich, 2004) based on dynamics in which fitness is shaped by a culturally embedded landscape and mutation contributes to alter the environment, including the cultural aspects, it is possible to show that social learning (in the form of imitation) may support the survival of strong reciprocators in single population dynamics.

What is the evidence for altruistic punishment?

If we start from ultimatum game, rejection can be interpreted as altruistic punishment because it is costly for the respondent and is driven by sense of unfairness. In fact, as explained by Eckel and Gintis (2010), the respondent is less willing to sacrifice endowment through rejection, when the respondent is a computer and cannot be "blamed" of unfairness. A stylized fact that emerged from a cross cultural research study with small scale societies is that altruistic punishment is widespread, but the level of heterogeneity is significant: "In five societies, the Tsimane, the Shuar, Isanga village, Yasawa, and the Samburu, less than 15% of the population were willing to reject 10% offers. In contrast, over 60% of the samples in four populations rejected such offers" (Henrich et al. 2006: 1769). Western subjects, and non-students' samples in particular, stand on the right extreme of the distribution, maximizing across societies the minimum offer they are willing to accept (Henrich, Heine and Norenzayan, 2010). A very interesting stylized fact (Henrich et al. 2005) is that average offers in Ultimatum Game, which (over)incorporate expected rejection, are correlated with Market Integration (the extent to which members of society engage into market transactions) and with Payoff to Cooperation (the extent to which household have independent income or depend on cooperating with others).

Another important piece of evidence is from public good games. In this setup, participants can either contribute or not to the public good: there is a social dilemma in the sense that contributing to the public good provides less incentive than free riding, but individuals gain if everybody contributes. Evidence by Fehr and Gachter (2000) is striking. First of all, they implement a protocol which is carefully eliminating reputation formation and direct reciprocity mechanisms, through anonymity and random group composition without repetition. The participants contribute very little in baseline condition, and contribution rate declines across repetitions. At the opposite, when altruistic punishment is allowed, contribution rate increases significantly, regardless if the treatment is implemented before or after the baseline, and punishment is immediately effective as can be seen from the jump from the last to the first round of the conditions, on both directions. Incidentally, this also rules out problems of comprehension. Moreover, there is no decreasing trend of contribution under punishment treatment. Finally, punishment intensifies if the subject contributes less than the average contribution, which suggests the existence of social norms.

3.3 Social norms

The importance of social norms has been recognized throughout social sciences, from anthropology to sociology, but is barely mentioned in standard economic theory. However,

the socialization of norms explains how human societies were able to "domesticate" certain human instincts and accomplish large scale organization (Henrich et al. 2010). Norms are based on shared beliefs on how individuals are ought to behave in certain situations (Fehr and Fischbacher, 2004a).

In general social norms belongs to the set of collective patterns of behaviour. It is customary to distinguish between the group of habits, collective custom, or legal injunction, which are typically followed irrespectively of what other people do, and the norms which are based on social expectation, i.e. beliefs on what other people do or expect (Bicchieri, 2017). One further distinction by Bicchieri is the one between *descriptive norms*, such as driving on the right side, which is causally determined by the empirical expectation on what other people do, and the social norms (e.g. protected sex, female genital mutilation etc.), which are grounded on both empirical expectation and *normative expectation*, i.e. second order beliefs on what other people (in the reference group) believe other ought to conform with (2017: p. 35).

The socialization of norms is typically transmitted through generations, but it is also a process of social learning.

Norms guide the behaviour of human beings, and our brain interprets certain contextual and environmental cues as signal to apply internalized norms to specific situation. Some examples are provided by Gintis et al. (2003) in the review of experiments with small scale societies, where they document that measured behaviour mimics everyday patterns of interaction. "The Orma immediately recognized that the public goods game was similar to the *harambee*, a locally initiated contribution that households make when a community decides to construct a road or school" and "Among the Au and Gnau, many proposers offered more than half the pie, and many of these ''hyperfair'' offers were rejected! This reflects the Melanesian culture of status-seeking through gift giving. Making a large gift is a bid for social dominance in everyday life in these societies, and rejecting the gift is a rejection of being subordinate" (Gintis et al. 2003: 159).

Norms can be seen as constitutive and not regulatory (Ostrom, 2000; Gintis, 2007; Fehr and Gintis, 2007), in the sense that they are aims in themselves and not instrumental to accomplished certain goal (as it would be under models of reciprocity).

The emergence and maintenance of social norms, besides relying on mechanisms of social learning, are again related with the presence of strong reciprocators. This can be seen in the propensity by human beings to punish deviants from social norms. This altruistic punishment is different from the one described in the previous subsection, which consists in punishing counterparts (*second party punishment*). In this case, the literature uses the definition of Third Party Punishment (TPP). TPP is the willingness to bear a cost to punish agents who deviates from social norms, even though the transaction that involves this deviation is not directly affecting the punisher. In everyday life, it would capture someone facing an unknown person because it threw garbage on the street.

Experimentally, TPP can be added to many standard interactions. Fehr and Fischbacher (2004b) introduce TPP in games where a norm of collaboration exists, such as the Prisoners Dilemma. Typically, the third party can spend a part of her endowment to reduce the payoff of participants to a PD, without being directly involved in the transaction. The game allows third party to punish both cooperators and defectors, without inducing framing of cooperation explicitly. However, the norm emerges spontaneously since one of five punishes mutual defectors, whereas almost nobody punishes cooperators. A defector is punished in 45.8% of the cases if it is matched with a cooperator.

Very interesting evidence comes out from the combination of TPP with games where a distributional norm is involved. This is typically the case in the Dictator Game. In the DG with TPP, the third party may reduce the payoff of the proponent. The decisions of the third party are collected through the *strategy method*: in this way, it is possible to record the willingness to punish the dictator under any possible scenario, and then the action corresponding to the actual scenario is implemented. Fehr and Fischbacher (2004b) document the presence of strong punishment if the dictator shares less than 50%, and some marginal *anti-social* punishment when more than 50% is shared. TPP is less than second party punishment, where the altruistic punisher is directly involved in the transaction, but not at the egalitarian level (50%) where no difference emerges.

The DG-TPP has also been implemented in small scale societies by Henrich et al. (2005; 2006; 2010) and Marlowe et al. (2008). One way to analyse these games is the Minimum Acceptable Offer (MAO): this is the minimum offer from a dictator that a TPP is willing to accept, restraining from punishing. Henrich et al. (2006) show that although individual level variation is partly explained by socio-demographic characteristics, a significant amount of between groups variation (around 38.2% of variation) is not explained by individual level determinants, which suggests a role of culture, and his correlated with altruism, as measured through the offer in the Dictator Game (the coefficient from the weighted regression is 0.23, statistically significant).

Marlowe et al. (2008) show that MAO is positively correlated with the rank of societies in terms of local group population (from field data) and ethnic group population (from the Ethnologue database of world languages on-line). Community Size is correlated with MAO also in Henrich et al. (2010).

Finally, since social norms are based on shared beliefs, Bicchieri (2006) suggests that two different types of expectations are important in explaining the adherence to a norm, what we believe others expect us to do (*normative expectations*), and what we think others will do (*empirical expectations*). In their experiments using Dictator Game and manipulating expectations through reported information on other participants' beliefs and behaviour, Bicchieri and Xiao (2009) show that empirical expectations are relatively more important in predicting Dictators' choices, especially when the two expectations run into conflicts.

3.4 Trust

We close this section by discussing the experimental evidence on the propensity to trust. At the microeconomic level, trust has been defined as the lubricant of economic transactions (Arrow, 1974): in a world of incomplete contracts, the shared beliefs on the trustworthiness of a counterpart lower down transaction costs facilitating the writing of contracts and sustaining markets. At the macro level, it has been shown that trust positively correlates with economic performance, measured in various ways (Zak and Knack, 2001; LaPorta et al 1997), although existing empirical evidence is based on questionable instruments and other empirical problems (Durlauf, 2002). The idea that trust can be beneficial to economic performance goes back to the original literature on social capital, where individual relationships are valuable resources (Coleman, 1988) and some general propensity to trust and cooperate increase efficiency of the market economy (Putnam et al., 1993) and the lack thereof constrains backward societies (Banfield, 1967).

Part of the empirical literature on trust uses survey based measures, such as the question in the World Value Survey "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" On the one hand, this information is recovered through representative samples of the population, but the quality of the information may be subject to critiques, since it is not clear what is been measured (Cardenas et al. 2013).

Experimental evidence on trust is provided through the investment game by Berg et al. (1995). This is an interaction between two parties, which took their decisions sequentially. The first mover (*trustor*) decides how much of her initial endowment is willing to transfer to the second mover (*trustee*). The transfer generates increasing resources and is socially beneficial, usually following the rule that every transferred token is tripled by the experimenter. The counterpart can decide on the allocation of the resources, her choice being purely distributive.

This is played under complete information and if backward induction holds, the relevant solution (*subgame perfect Nash Equilibrium*) is unique and includes as equilibrium strategies zero transfer and zero sharing.

In the implementation of this game, the trustee's moves are taken as measures of trustworthiness, and trustor's moves as trust. The most common protocol includes the strategy method to recover the trustee moves, different potential transfer levels for the trustor, and different levels of anonymity in the interaction (Johnson and Mislin, 2011).

The importance of this measure of trust is that it is behaviour-related and not just belief-related (Fehr, 2009). In fact, it concerns some transfer of resources to the disposal of another person, typically a stranger. While it can be argued that survey questions force the respondent to imagine how she would act in a real-life situation involving trust, experiments have the advantage of explaining precisely all the contingencies and thus provide a cleaner

measurement (Fehr, 2009), and of course the use of performance related payments reduces reported noise (Camerer and Hogarth, 1999).

Experimental measures of trust appear slightly correlated with some socio-demographic characteristics, like age (Belli et al. 2012) and gender (Buchan et al. 2008), but the regional and cultural variation (Johnson and Mislin, 2011; Cárdenas and Carpenter, 2008) suggests that the main determinants of trust stand elsewhere.

Although the transfer of resources at the disposal of other people in expectation of some return is by all means an investment, risk preferences do not appear as the main determinant of trust behaviour. Bohnet et al. (2008) propose an experimental protocol including a trust game, a risky decision problem where the counterpart is just a lottery and a third task replicating the trust game but where the principal and agents' payoffs depend on a random mechanism (*risky dictator game*). The trust game is dichotomic for both the trustor and the trustee and similarly for the lotteries: in this way, the only differences are the strategic element for the trust game and the presence of another person's payoff (which may trigger other regarding preferences), for the risky dictator game. For each of the tasks, the authors elicit the Minimum Acceptable Probability (MAP), which is the minimum acceptable likelihood to gain when facing risky choices, which is then compared with the real probability and determines the final decision to face or not the risk. Comparing MAPs across the three tasks they capture the presence/absence of betrayal aversion, which is the preference to avoid a risk related with other person's lack of trustworthiness, social preferences (when another person payoff is involved) and of course a standard measure of risk attitude (by comparing the MAP with the real probability in the decision task).

They are able to show that *betrayal aversion* is a very relevant phenomenon, since MAPs in the trust game significantly exceed MAP in the risky dictator game. Betrayal aversion is also positive in all six countries where experiments were conducted (Brazil, China, Oman, Switzerland, Turkey and the US).

Cross-cultural evidence on trust seems to be partly explained by formal and informal institutional settings, at least with regards to shared beliefs, while preferences may be partly exogenous (Fehr, 2009). In particular, the discussion on strong reciprocators and social norms in previous paragraphs is consistent with the experimental evidence on trust:

- Allowing reputation formation increases trust (Berg et al. 1995; Boero et al. 2009; Charness et al. 2011; Dubois et al. 2012);
- Third party punishment increases trust (Charness et al. 2008);
- Reduction in social distance increases trust, as shown by the effect of unbinding communication in Bracht and Feltovich (2009);
- Fair procedures, such as consultative voting on the preferred outcome, increase trust in Bogliacino, Jiménez and Grimalda (2015).

4. Concluding Remarks

In this paper, we provide a selective review of the literature based on RCTs that is relevant to evolutionary and structural dynamics approach in economics. We discuss the methodological strengths and weaknesses of the evidence provided by RCTs, the evidence on cognitive limits, and the evidence on homo oeconomicus as microfoundation, which we think should be abandoned in favour of human being as strong reciprocator equipped with social norms. Taking stock of this literature is a very hard task, but the tentative conclusions we draw are the following ones.

At the methodological level, we claim that field data, combined with cleaned and controlled experiments and computer based ABMs are the way forward to provide evidence on both micro decisions and aggregate properties of complex systems.

On the best way to answer Dosi and Virgillito's quest for external consistency for macro ABMs or generally for input in terms of empirical parameters to incorporate in such models, we think that the open issue is related with the concept of *emerging property*, which outlines an anti-reductionist view of reality. A set of entities at a certain level owes its existence to lower-level entities but also presents a set of states / properties / regularity of its own that can be studied independently.

The relationship between the micro and macro level upon which the concept of emerging effects hinges is captured more analytically by that of supervenience: (1) a higher-level structure depends ontologically on the lower level one, that is, the former could not exist if it did not exist the latter; (2) distinctions and variations within the higher-level phenomena \ structures - different performances, different fitness, different function - are necessarily based on differences on the lower level, but no the opposite: different individual configurations can determine the same macro phenomenon.

In the social sciences, after a period of ostracism - due to stiff interpretations of 'methodological individualism tenets: a social phenomenon is explained only if it is or in principle can be "reduced" to properties attitudes and actions of individuals - the concept of emergent effects is now accepted if based on a non-metaphysical concept of causation. If so it could be also useful for empirical research. The concept of emergent effects must therefore be rooted in individual action – but it must be shown that, in certain situations, individual actors have less weight than the interdependence of actions. The concept of interdependence is therefore pivotal and needs to be formalized. The presence of emerging effects requires a concept of social structure different from the neoclassical economic theory's notion.

References

Adorno, TW, KR Popper, R Dahrendorf, J Habermas, H Albert and H Pilot (1973) La disputa del positivismo en la sociología alemana. Barcelona: Ediciones Grijalbo.

Andreoni, J and J Miller (2002) Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. Econometrica, 70(2): 737-753

Angrist, JD and JS Pischke (2010) The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. Journal of Economic Perspective, 24(2): 3-30

Angrist, JD, Imbens, GW and DB Rubin (1996) Identification of Causal Effects using Instrumental Variables. Journal of the American Statistical Association, 91(434): 444-455

Arrow, K (1974) The limits of organization. New York: Norton.

Ashraf, N, CF Camerer, G Loewenstein (2005) Adam Smith, Behavioral Economist. Journal of Economic Perspective, 19(3): 131-145

Aumann, R (1995) Backward induction and common knowledge of rationality. Games and Economic Behavior 8, 6–19.

Axelrod, R and WD Hamilton (1981) The evolution of cooperation. Science, 211(4489): 1390-1396

Babcock, L., & Loewenstein, G. (1997). Explaining Bargaining Impasse: The Role of Self-Serving Biases. The Journal of Economic Perspectives, 11(1), 109-126. doi: citeulike-article-id:6656163

Banfield, E. C. (1967) The moral basis of a backward society, Simon & Shuster.

Barr, M., Mullainathan, S., & Shafir, E. (2013). Behaviorally Informed Regulation. In E. Shafir (Ed.), The Behavioural Foundations of Public Policy (pp. 441-461). Princeton, NJ: Princeton University Press.

Belli, S, R Rogers, and J Lau (2012) Adult and adolescent social reciprocity: Experimental data from the Trust Game. Journal of Adolescence, 35, 1341-1349.

Berg, J, Dickhaut, J and K McCabe (1995) Trust, Reciprocity, and Social History. Games and Economic Behavior, 10, 122-142.

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? History of Economic Ideas, 18, 133–166

Bicchieri, C and E Xiao (2009) Do the Right Thing: But Only If Others Do So. Journal of Behavioral Decision Making, 22: 191-208.

Bicchieri, C. (2017) Norms in the Wild. How to Diagnose, Measure, and Change Social Norms. Oxford University Press.

Bicchieri, C. (2006). The Grammar of Society: the Nature and Dynamics of Social Norms, New York: Cambridge University Press.

Binmore, K and A Shaked (2010) Experimental economics: Where next? Journal of Economic Behavior and Organization, 73: 87-100

Boero, R., Bravo, G., Castellani, M. and Squazzoni, F. (2009). Reputational cues in repeated games. The Journal of Socio-Economics, 38, 871-877

Bogliacino F, L Jiménez and G Grimalda (2015) Consultative, Democracy and Trust, DOCUMENTOS DE TRABAJO - ESCUELA DE ECONOMÍA 012696, UN - RCE - CID.

Bogliacino, F and P Ortoleva (2015) The Behavior of Other as a Reference Point, DOCUMENTOS DE TRABAJO - ESCUELA DE ECONOMÍA 013611, UN - RCE - CID.

Bogliacino, F, C Codagnone and GA Veltri (2015) The Behavioural Turn in Consumer Policy: Perspectives and Clarifications, Intereconomics, 50(2): 108-114

Bogliacino, F. C. Codagnone, GA Veltri, A Chakravarty, P. Ortoleva, G. Gaskell, A. Ivchenko, F. Lupiáñez-Villanueva, F Mureddu, C. Rudisill (2015) Pathos & ethos: Emotions and willingness to pay for tobacco products. PLoS ONE DOI: 10.1371/journal.pone.0139542

Bohnet, I, F Greig, B Herrmann, and R Zeckhauser (2008) Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. American Economic Review, 98, 294–310.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. The American Economic Review, 90(1), 166-193.

Boyd, R, H Gintis, S Bowles and PJ Richerson (2003) The evolution of altruistic punishment. PNAS, 100(6): 3531-3535

Bracht, J. and Feltovich, N. (2009).Whatever you say, your reputation precedes you: observation and cheap talk in the trust game. Journal of Public Economics, 93(9-10), 1036-1044.

Buchan, N, R Croson, and S Solnick (2008) Trust and gender: An examination of behavior and beliefs in the Investment Game. Journal of Economic Behavior & Organization, 68, 466-476.

Camerer, C. (1995). Individual decision making. In A. Kagel (Ed.), The Handbook of Experimental Economics (pp. 587- 704). Princeton, NJ: Princeton University Press.

Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton: Princeton University Press.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2002). Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games. Journal of Economic Theory, 104(1), 137-188. doi: http://dx.doi.org/10.1006/jeth.2002.2927

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A Cognitive Hierarchy Model of Games*. The Quarterly Journal of Economics, 119(3), 861-898. doi: 10.1162/0033553041502225

Camerer, C., & Ho, T. (1999). Experience-weighted Attraction Learning in Normal Form Games. Econometrica, 67(4), 827-874. doi: 10.1111/1468-0262.00054

Camerer, C, and R Hogarth (1999) The effects of financial incentives in experiments: a review and capital-labor-production framework. Journal of Risk and Uncertainty 19, 7–42.

Camerer, C., & Lowenstein, G. (2003). Behavioral economics: Past, present, future In C. Camerer, G. Lowenstein & M. Rabin (Eds.), Advances in Behavioural Economics (pp. 3-52). Princeton: Princeton University Press.

Campbell, D. (1969). Reforms as experiments. American Psychologist, 24(4), 409-429.

Cardenas, JC and J Carpenter (2008) Behavioural Development Economics: Lessons from Field Labs in the Developing World, Journal of Development Studies, 44(3), 311-338

Cárdenas, JC, A Chong, and H Ñopo (2013) Stated social behavior and revealed actions: Evidence from six Latin American countries, Journal of Development Economics, 104(C), p. 16-33

Cerulo, K. A. (2010). Mining the intersections of cognitive sociology and neuroscience. Poetics, 38(2), 115-132. doi: http://dx.doi.org/10.1016/j.poetic.2009.11.005

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. The Quarterly Journal of Economics, 117(3), 817-869.

Charness, G., Cobo-Reyes, R. & Jiménez, N. (2008). An investment game with third-party intervention. Journal of Economic Behavior & Organization, 68, 18-28.

Charness, G., Du, N. & Yang, C. (2011). Trust and trustworthiness reputations in an investment game. Games and Economic Behavior, 72, 361-375.

Chetty, R. (2015). Behavioral Economics and Public Policy: A Pragmatic Perspective. The American Economic Review, 105(5), 1-33. doi: 10.1257/aer.p20151108

Chong, J.-K., Ho, T.-H., & Camerer, C. (2016). A generalized cognitive hierarchy model of games. Games and Economic Behavior, 99, 257-274. doi: http://dx.doi.org/10.1016/j.geb.2016.08.007

Christ, CF (1994) The Cowles Commission's Contribution to Econometrics at Chicago, 1939-1995. Journal of Economic Literature, XXXII, 30-59

Cohen, M., & Bacdayan, P. (1994). Organizational Routines Are Stored As Procedural Memory: Evidence from a Laboratory Study. Organization Science, 5(4), 554-568.

Coleman, J. (1988). Social capital in the creation of human capital. The American Journal of Sociology, 94, s95-s120

Cooper, D, Kagel, J (2013) Other regarding preferences: A selective survey of experimental results. In The Handbook of Experimental Economics; Kagel, J.H., Roth, A., Eds.; Princeton University Press: Princeton, NJ, USA; Volume 2.

Crawford, V. (1995). Adaptive Dynamics in Coordination Games. Econometrica, 63(1), 103-143.

Crawford, V. P. (2013). Boundedly Rational versus Optimization-Based Models of Strategic Thinking and Learning in Games. [10.1257/jel.51.2.512]. Journal of Economic Literature, 51(2), 512-527.

Crawford, V., Costa-Gomes, M., & Iriberri, N. (2013). Structural Models of Non-equilibrium Strategic Thinking: Theory, Evidence, and Applications. Journal of Economic Literature, 51(1), 5-62.

Cronbach, LJ, S.R. Ambron, S.M. Dornbusch, R.D. Hess, R.C. Hornik , D.C. Phillips, D.E. Walker, SS Weiner: Toward reform of program evaluation, San Francisco 1980, Jossey-Bass.

Cyert, R., & March, J. (1963). A Behavioral Theory of the Firm. New York: Prentice Hall.

DiMaggio, P. (2002). Why cognitive (and cultural) sociology needs cognitive psychology. In K. Cerulo (Ed.), Culture in Mind: Toward a Sociology of Culture and Cognition (pp. 274–282). New York/London: Routledge. doi: 10.2307/2138254

Dosi, G., Marengo, L., & Fagiolo, G. (2005). Learning in evolutionary environments. In K. Dopfer (Ed.), The Evolutionary Foundations of Economics Cambridge Cambridge University Press.

Dosi, G and ME Virgillito (2017) In order to stand up you must keep cycling: change and coordination in complex evolving economies. Structural Change and Economic Dynamics, this issue

Dubois, D., Willinger, M. and Blayac, T. (2012). Does players' identification affect trust and reciprocity in the lab? Journal of Economic Psychology, 33, 303-317.

Durlauf, SN (2002) On the empirics of social capital. Economic Journal, 112: 459-479

Eckel, C and H Gintis (2010) Blaming the messenger: Notes on the current state of experimental economics. Journal of Economic Behavior and Organization, 73: 109-119

Falk, A and U Fischbacher (2006) A theory of reciprocity. Games and economic behavior, 54: 293- 315

Fehr, E (2009) On the economics and biology of trust. Journal of the European Economic Association, 7(2–3):235–266

Fehr, E and H Gintis (2007) Human Motivation and Social Cooperation: Experimental and Analytical Foundations. Annual Review of Sociology, 33: 43-64

Fehr, E and KM Schmidt (1999) A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114, 817–868

Fehr, E and S Gachter (2000) Cooperation and punishment in public good experiments. American Economic Review, 90(4): 980-994

Fehr, E and S Gachter (2002) Altruistic punishment in humans. Nature, 415: 137-140

Fehr, E and U Fischbacher (2003) The nature of human altruism. Nature, 425: 785-791

Fehr, E and U Fischbacher (2004a) Third-party punishment and social norms. Evolution and Human Behavior 25: 63–87

Fehr, E and U Fischbacher (2004b) Social norms and human cooperation. TRENDS in Cognitive Sciences, 8(4): 185-190

Fehr, E, G Kirchsteiger and A Riedl (1998) Gift exchange and reciprocity in competitive experimental markets. European Economic Review, 42: 1-34

Forsythe, R, JL Horowitz, NE Savin, M Sefton (1994) Fairness in simple bargaining experiments. Games and Economic Behavior, 6, 347-369

Fowler, JH (2005) Altruistic punishment and the origin of cooperation. PNAS, 102(19): 7047-7049

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time Discounting and Time Preference: A Critical Review. [10.1257/002205102320161311]. Journal of Economic Literature, 40(2), 351-401.

Frerichs, S. (2011). False Promises? A Sociological Critique of the Behavioural Turn in Law and Economics. Journal of Consumer Policy, 34(3), 289-314

Friedman, D and S Sunder (1994) Experimental Methods. A primer for economists. Cambridge University Press.

Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases". European Review of Social Psychology, 2(1), 83-115. doi: 10.1080/14792779143000033

Gigerenzer, G. (2010). Personal Reflections on Theory and Psychology. Theory & Psychology, 20(6), 733-743. doi: 10.1177/0959354310378184

Gigerenzer, G. (2015). On the Supposed Evidence for Libertarian Paternalism. Review of Philosophy and Psychology, 6(3), 361-383. doi: 10.1007/s13164-015-0248-1

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. The Annual Review of Psychology, 62, 451–482.

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

Gintis, H (2007) A framework for the unification of the behavioral sciences. Behavioral and Brain Sciences, 30, 1-61

Gintis, H, S Bowles, R Boyd, and E Fehr (2003) Explaining altruistic behavior in humans. Evolution and Human Behavior 24: 153–172

Gneezy, U (2005) Deception: The role of consequences. American Economic Review 95(1): 384–94

Goldstein, D., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. Psychological Review, 109, 75-90.

Goldstein, W., & Weber, E. (1995). Content and its discontents: The use of knowledge in decision making. In J. Busemeyer, R. Hastie & D. Medin (Eds.), Decision making from a cognitive perspective. The psychology of learning and motivation (Vol. 32, pp. 83-136). New York Academic Press.

Guala, F (2005) The Methodology of Experimental Economics. Cambridge University Press

Guth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization 3, 367–388.

Guth, W., Tietz, R., 1990. Ultimatum bargaining behavior: A survey and comparison of experimental results. Journal of Economic Psychology 11, 417–449.

Hamilton, WD (1964) The Genetical Evolution of Social Behaviour. I. Journal of Theoretical Biology, 7: 1-16

Harrison, G (1982) Theory and Misbehavior of First Price Auctions: Reply. American Economic Review, 82:5, 1426–43

Harrison, GW and JA List (2004) Field Experiments. Journal of Economic Literature, Vol. 42, No. 4, pp. 1009-1055

Hayes. JR and HA Simon. (1974) Understanding Written Problem Instructions," in LW Gregg (Ed) Knowledge and Cognition, Hillsdale, NJ: Erlbaum.

Heckman, JJ (2000) Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective. Quarterly Journal of Economics, Vol. 115, No. 1, pp. 45-97

Heckman, JJ (2008) Econometric Causality. International Statistical Review, 78(1): 1-27

Heckman, JJ (2010) Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. Journal of Economic Literature, 48(2): 356-398

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

Henrich, J, Heine, SJ and A Norenzayan (2010) The weirdest people in the world? Behavioral and Brain Sciences, 33: 1-23

Henrich, J, J Ensminger, R McElreath, A Barr, C Barrett, A Bolyanatz, JC Cardenas, M Gurven, E Gwako, N Henrich, C Lesorogol, F Marlowe, D Tracer, and J Ziker (2006) Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. Science, 327: 1480-1484

Henrich, J, R Boyd, S Bowles, C Camerer, E Fehr, H Gintis, R McElreath, A Barr, J Ensminger, N Henrich, K Hill, F Gil-White, M Gurven, F Marlowe, J Patton and D Tracer (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behavioral and Brain Sciences, 28: 795-815

Henrich, J, R McElreath, A Barr, J Ensminger, C Barrett, A Bolyanatz, JC Cardenas, M Gurven, E Gwako, N Henrich, C Lesorogol, F Marlowe, D Tracer, and J Ziker (2006) Costly Punishment Across Human Societies. Science, 312: 1767-1770

Hobbes, T (2011) Leviatán. First edition, 1651; This edition: Buenos Aires, Lozada 2011

Hollis, M., & Sugden, R. (1993). Rationality in Action. Mind, 102(405), 1-35.

Hume, D (1793) A Treatise of Human Nature. Penguin, London

Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. Journal of Economic Methodology, 23(1), 1-25.

Johnson, N and A. Mislin (2011) Trust Games: a Meta-Analysis. Journal of Economic Psychology, 32, 865-889.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. American Economic Review, 93(5), 1449-1475.

Kahneman, D. (2011). Thinking fast and slow. London: Penguin Books

Kahneman, D. (1996). Comment [on Plott (1996)]. In K. Arrow, E. Colombatto, M. Perlman & C. Schmidt (Eds.), The rational foundations of economic behaviour (pp. 251–254). Basingstoke: Macmillan and International Economic Association.

Kahneman, D., & Frederick, S. T. G., D. Griffin and D. Kahneman (eds.) . New York: . (2002). Representativeness revisited: Attribution substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), Heuristics of intuitive judgment: Extensions and applications. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica 47(2), 263–292.

Kahneman, D., & Tversky, A. (1996). On the Reality of Cognitive Illusion. Psychological Review, 103, 582-591.

Kahneman, D., Knetsch, J., & Thaler, R. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. The Journal of Economic Perspectives, 5(1), 193–206.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). Judgment under uncertainty: Heuristics and biases. Cambridge, UK: Cambridge University Press

Kreps, DM (1988) Notes on the Theory of Choices. Boulder, Colorado: Westview Press

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting*. The Quarterly Journal of Economics, 112(2), 443-478. doi: 10.1162/003355397555253

LaPorta, R, F Lopez-de-Silane, A Shleifer, and RW Vishny (1997) Trust in Large Organizations. American Economic Review, 87, 333–338.

Leamer, E (1983) Let's Take the Con Out of Econometrics. American Economic Review, 73(1): 31–43.

Lee, DS and T Lemieux (2010) Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48: 281-355

Lewis, M (2016) The Undoing Project: A Friendship That Changed Our Minds. New York, WW Norton & C.

Loewenstein, G. (2000). Emotions in Economic Theory and Economic Behavior. [10.1257/aer.90.2.426]. American Economic Review, 90(2), 426-432.

Loewenstein, G., & Prelec, D. (1992). Anomalies in Intertemporal Choice: Evidence and Interpretation. Quarterly Journal of Economics, 107(2), 573-597.

Marlowe, FW, J Colette, A Barr, C Barrett, A Bolyanatz, JC Cardenas, J Ensminger, M Gurven, E Gwako, J Henrich, N Henrich, C Lesorogol, R McElreath and D Tracer (2008) More 'altruistic' punishment in larger societies. Proc. R. Soc. B (2008) 275, 587–590

Marshall, A (2009) Principles of Economics. First Edition 1890, This Edition: Orlando, Fla. Signalman Publishing

McCloskey, D (2005) The Trouble with Mathematics and Statistics in Economics. History of Economic Ideas XIII (3): 85-102

McCloskey, D (1983) The Rhetoric of Economics, Journal of Economic Literature 31: 482-517

Milgrom, P., & Roberts, J. (1991). Adaptive and Sophisticated Learning in Normal Form Games. Games and Economic Behavior, 3(1), 82-100.

Mill, JS (1970) A system of logic. Rationative and inductive: being a connected view of the principles of evidence and the methods of scientific investigation. London, Routledge (First Edition, 1843)

Neisser, U. (1967). Cognitive psychology. New York Meredith.

Neisser, U. (1976). Cognition and reality. San Francisco Freeman.

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

Newell, A., & Simon, H. (1972). Human Problem Solving. New York: Prentice Hall.

Nowak, MA and K Sigmund (1998) Evolution of indirect reciprocity by image scoring. Nature, 393: 573-577

O'Donoghue, T., & Rabin, M. (1999). Doing It Now or Later. The American Economic Review, 89(1), 103-124.

Ostrom, E (2000) Collective Action and the Evolution of Social Norms. Journal of Economic Perspectives, 14(3): 137–158

Parsons, T (1937) The Structure of Social Action, New York: McGraw-Hill Book Co.

Payne, J., Bettman, J., Coupey, E., & Johnson, E. (1992). A Constructive Process View of Decision Making: Multiple Strategies in Judgment and Choice. Acta Psychologica, 80, 107-141.

Popper, K (1963) Conjectures and Refutations: The Growth of Scientific Knowledge. London: Routledge & Kegan Paul

Putnam, RD, R Leonardi, RY Nanetti (1994) Making democracy work: Civic traditions in modern Italy. Princeton University Press.

Quine, WVO (1951) The two dogmas of empiricism. The Philosophical Review 60: 20–43

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. American Economic Review, 83(5), 1281-1302.

Rebonato, R. (2014). A Critical Assessment of Libertarian Paternalism. Journal of Consumer Policy, 37(3), 357-396. doi: 10.1007/s10603-014-9265-1

Rosenbaum PR and DB Rubin (1983) The central role of the propensity score in observational studies for causal effects. Biometrika. 70:41–55.

Rosenzweig, MR, and KI Wolpin (2000) Natural 'Natural Experiments' in Economics. Journal of Economic Literature, 38(4): 827–74.

Rubin, DB (1974) Estimating causal effects of treatments in randomized and non randomized studies. Journal of Educational Psychology, 66(5), 688-701.

Rubinstein, A. (1998). Modeling Bounded Rationality. Cambridge and London: MIT Press.

Scazzieri, R (2017) Structural Dynamics and Evolutionary Change. Structural Change and Economic Dynamics. [This issue]

Selten, R and R Stoecker (1986) End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach. Journal of Economic Behavior and Organization, 7(1): 47–70

Shadish, WR, TD Cook, and DT Campbell (2002) Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Boston, MA, Houghton Miffl in Company

Shafir, E, P Diamond and A Tversky (1997) Money Illusion. Quarterly Journal of Economics, 112(2): 341-374

Shah, A., & Oppenheimer, D. (2008). Heuristics made easy: an effort-reduction framework. Psychological Bullettin, 137, 207–222.

Simon, H. (1955). A Behavioral Model of Rational Choice. Quarterly Journal of Economics, 69(1), 99–118.

Simon, H. (1990). Invariants of human behavior. The Annual Review of Psychology, 41, 1-19.

Slovic, P. (1995). The construction of preferences. American Psychologist, 50, 364-371.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. (2002). The Affect Heuristic. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgement (pp. 397-420). New York: Cambridge University Press.

Smith, A (1759) The Theory of Moral Sentiments. D. D. Raphael and A. L. Macfie, eds. (1981) Liberty Fund: Indianapolis.

Smith, VL (1976) Experimental Economics: Induced Value Theory. The American Economic Review, Vol. 66, No. 2, Papers and Proceedings of the Eighty-eighth Annual Meeting of the American Economic Association: 274-279

Smith, VL (1982) Microeconomic Systems as an Experimental Science. American Economic Review, Vol. 72, No. 5, pp. 923-955

Smith, VL (1989) Theory, Experiments and Economics. Journal of Economic Perspectives, 3(1): 151-169

Smith, VL (1994) Economics in the Laboratory. Journal of Economic Perspectives, 8(1): 113–1

Smith, VL (2003). Constructivist and Ecological Rationality in Economics. The American Economic Review, 93(3), 465-508

Smith, VL (2010) Theory and experiment: What are the questions? Journal of Economic Behaviour and Organization, 73: 3-15

Sobel, J. (2005). Interdependent Preferences and Reciprocity.. Journal of Economic Literature, 43(2), 392-436.

Streeck, W. (2010). Does 'Behavioural economics' offer an alternative to the neoclassical paradigm? Socio-Economic Review, 8(2), 387-397.

Thaler, R. (1980). Toward a positive theory of consumer choice. Journal of Economic Behavior and Organization, 1, 36-60.

Thaler, R. (2016). Behavioral Economics: Past, Present, and Future. American Economic Review, 106(7), 1577–1600.

Thaler, RH (1988). Anomalies: The Ultimatum Game. Journal of Economic Perspectives 2: 195–206

Thaler, RH (1999) Mental Accounting Matters. Journal of Behavioural Decision Making, 12: 183-206

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychological Review, 79, 281-199.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124–1131

Tversky, A., and Kahneman, D. (1986). Rational choice and the framing of decisions. The Journal of Business, 59(4), S251-S278

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics 106, 1039-1061.

Wilson, DS and E Sober (1994) Reintroducing group selection to the human behavioural sciences. Behavioral and Brain Science 17, 585–654

Woodford, M. (1990). Learning to Believe in Sunspots. Econometrica, 58(2), 277-307. doi: 10.2307/2938205

Wrong, DH (1961) The Oversocialized Conception of Man in Modern Sociology. American Sociological Review, Vol. 26, No. 2, pp. 183-193

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35, 151–175.

Zajonc, R. B. (1980) Feeling and Thinking. American Psychologist, 35(2), 151-175

Zak, PJ, and S Knack (2001) Trust and Growth. Economic Journal, 111, 295–321

Zerubavel, E. (1997). Social Mindscapes. An Invitation to Cognitive Sociology. Cambridge, Mass.: Harvard University Press.

Appendix. Causality and Validity

Addressing the demands posed by the criteria to achieve internal validity is the very essence of RCTs. Understanding the extent to which this methodology allows for causal identification between two variables goes beyond the pure interest on the experiment as a tool, since most of the current empirical research in applied microeconomics relies on a theoretical apparatus (quasi-experiment), whose *gold standard* is the RCT.

Angrist and Pischke (2010) provide a good historical reconstruction of what they call the "credibility revolution" in econometrics. In a nutshell, the dissatisfaction with structural econometrics and the use of "whimsical" assumptions (Leamer, 1983) gave rise to a new methodological toolkit, which covers most of the applied work (with some field-specific exceptions). As at the time the replicability of econometric results was nil, the robustness of standard econometrics results was challenged, for little consensus was achieved. A second problem was the use of exclusion restrictions (following the typical approach of the Cowles Commission; Heckman, 2000; Christ, 1994) that appear arbitrary at least. The so-called revolution puts most of the emphasis on research design, and in particular, the identification of some source of discontinuity or exogenous variation in the condition of expositions, e.g. natural phenomena, some peculiar pattern of application of a law, and sometimes a direct ex ante specific choice of implementation. This is the basic tenet beyond the quasi-experimental apparatus, including instrumental variable (Angrist et al. 1996), Regression Discontinuity Design (Lee and Lemieux, 2010), Propensity Score Matching (Rosenbaum and Rubin, 1983), and "natural experiments" (e.g. the use of twins, or the sex of the new-born; Rosentzweig and Volpin, 2010).

In reality, this apparatus goes back to the experimentalist approach to policy evaluation developed in the 1960s by Campbell and his circle (Campbell, 1969), who clarified most of the theoretical coordinates, upon which the credibility revolution and modern policy evaluation toolkit are based.[13]

For the sake of the present discussion, we will briefly explain the problem of internal validity, i.e. the degree of truth of a causal statement, then we will discuss its limits, and finally move to the issue of inference, related with what Shadish et al (2002) defined as external and construct validity.

In economics, causality is expressed in the form of controlled variation: given a certain theoretical relationship between a variable y and a set of variable $\boldsymbol{x}=\{x_1, \ldots, x_n\}$**,** causality

[13] There is an important political message behind Campbell's program. In an epoch in which planned economy was believed to be a credible competitor to the dominating power (the US), the rhetoric of *laissez faire* was precautionary abandoned, in favour of the more concrete *programming*. The philosophy that stands behind Campbell was a Popperian "reforms-as-experiments" (Bogliacino et al., 2015), with an emphasis on gradual change, cumulative understanding and rational choice among policy alternatives. Of course, the overall approach presumes some sort of *structural* invariance, in the sense that society represents the lab and not itself an object of change. This goes back to the famous sociological debate between Popper and Adorno on the nature of scientific knowledge (Adorno et al. 1973).

between x_i and y, requires to variate the former and detect the effect on the latter, keeping everything else constant. This is the meaning of a *ceteris paribus* clause, as popularized in economics by Marshall (2009).

This goes back to the concept of successionist causation as developed by Hume (1793), where causation is something external to observation, a characteristic of *being conjoint*, which cannot be experienced, but only inferred from the repeated observation of some specific cause and some specific effect. It is also related to Stuart Mills' (1970) method of difference that could be considered the first intuition behind the development of experimentalism.

In other words, causality is a covariation in the studied effect following the variation of a cause, where alternative plausible explanations are excluded. Whereas temporal succession or statistical association may be technical problems of measurement that can be solved under certain degree of confidence, the key question is how to exclude plausible alternative stories. Refutation of alternative conjectures may appear a relative easier task (Popper, 1963); in reality, alternative stories may not be finite in number, nor we can consider a rejection to be a death of a hypothesis given that, in social science, complete causal channels are rarely specified (i.e. you cannot reject the causal description between a light switch and the light just because the bulb may be burned; Shadish et al. 2002) and there is as usual a problem of *correspondence* between theoretical construct and empirical variable (see *infra* on the concept of construct validity).

The structural econometrics approach tries to operationalize this problem in the following way. First of all, the definition of the problem is theory-driven and requires to specify the set of variables that are in the information set of the econometrician, distinguishing them from unobservable ones,[14] and then to isolate the variable of interest. Secondly, identification of the relevant parameter requires computing the exclusion restrictions. As explained by Heckman (2000) we have a many-to-one mapping from models to data, and the identification requires computing the minimum number of restrictions that allows the invertibility of such a relationship. The third step is of course the estimation of the parameter from the data.

In the credibility revolution approach, grounded on Rubin (1974), this idea is expressed in terms of alternative potential outcomes for units. Assume an independent variable which is either zero or one, as in the standard treatment effect literature. In simple words, one could think at the outcome in presence of treatment and the outcome in absence of treatment for each unit j. The causal contribution is the difference between the two, but this poses an obvious problem: since we are either observing units under status one or status zero of the independent variable, this generates a missing data problem (*evaluation problem*). Of course, one can compute the average outcome of all those units which have been treated and the average outcome of those which have not been, but the difference between the two groups is

[14] In the terminology of the Cowles Commission, the distinction was the one between "*systematic observable* variables that are ingredients of economic theory […] and *nonsystematic unobservable* variables that represent random disturbance to equations" (Christ, 1994: 34, emphasis in the original).

not the required causal impact of treatment, unless one can be sure that what would have occurred in absence of treatment on those that are in status one is on average equal to what did happen to those in status zero. Typically, this is not the case with observational data because people self-select through their decision rules into status (*selection problem*). RCTs "solve" the selection problem through the random assignment of the treatment status, accomplishing the objective of balancing out observable and unobservable variables on subjects in status one and zero, thus keeping "everything else" constant, on average. The rest of the modern toolkit of quasi-experimental research tries to accomplish the same result, for example, searching for a discontinuity under which a certain subset of units is assigned to status one or zero of the dependent variable *as if* the status was randomized (see Figure 5 in Lee and Lemieux, 2010).

The first implication of this methodology is that we want to keep environment under control. We are not denying the role of complex and evolving context, but we must go gradually, changing one condition at the time, to understand what is going on. As suggested by Guala (2005) models and experiments are very similar animals (the former are thought experiments), and they simply address different concerns. What we claim in this paper is that experiments can feed models, for example in terms of individual decision-making, but should not be asked for what they cannot provide. *Controlled* conditions mean precisely that we want the environment to be "transparent", and not change while we manipulate a certain variable, whereas complex evolutionary dynamics and simulations are typically characterized by "black box" features, which is the price to deal with complexity.

The causal interpretation requires *stable unit treatment value assumption* (SUTVA), which is equivalent to ruling out that the status one or zero can affect subjects in the alternative condition. In the interpretation, one should be conscious that this is a *policy invariance argument*. Alternative *potential* outcomes are assumed unaffected by assignment mechanisms, i.e. excluding social interaction, general equilibrium effect and contagion (Heckman, 2008). This obviously constrains the kind of causal inference we can aim at.

The above discussion defines the overall domain under which internal validity holds. What about the inference that can be done from experiments? It can be of legitimate concern that the quest for internal validity may have turned design-based studies into narrow research (Angrist and Pischke, 2010). The question can be framed in the following way, as presented by Guala (2005): given that, we are able to support a claim under certain environment, to what extent we are able to make inference outside that specific environment?

One could further split the problem in two different sub-problems: the representation of empirical evidence in terms of categories on which theories are predicated (*construct validity*) and the generalizability of the results outside the specific sample at hand (*external validity*). This is the formulation of the general theory of Shadish et al. (2002).

Whereas construct validity is self-explanatory, and corresponds for example to the problem of the extent to which measuring first party moves through the investment game of Berg et

al. (1995) can be considered a clean identification of trust, external validity usually extends beyond standard representativeness of the sample. Inference from local experimental evidence should be analysed as the possibility to generalize from specific UTOS (unit, treatment, observation and settings) to general ones (Cronbach et al. 1980). Most of the discussion on field experiment in the introduction from Harrison and List (2004) are responses to specific questions that can be traced back to the issue of external validity.

There is no blueprint to achieve external validity (as in the case of internal validity). Again, following Guala (2005) the best practice is to go step-by-step, discussing the potential threat to validity and looking for controlled variation to detect potential effects. This is the reason why we will discuss comparative evidence across samples and designs.