Institute for
New Economic Thinking
AT THE OXFORD MARTIN SCHOOL

The Behavioral Foundations of New Economic Thinking

Sanjit Dhami & Eric Beinhocker

9th August 2019

INET Oxford Working Paper No.2019-13

Foundations of Economics Research Theme

The Behavioral Foundations of New Economic Thinking

Sanjit Dhami* **Eric Beinhocker†**

9 August 2019

Abstract

This paper produces an edited version of an interview conducted by Professor Eric Beinhocker, Executive Director, INET Oxford, with Professor Sanjit Dhami of the University of Leicester, author of the Foundations of Behavioral Economic Analysis (Dhami 2016), on 9th May 2019 at the University of Oxford. The questions posed by Professor Beinhocker covered a wide interdisciplinary terrain that ranged from important contributions in behavioral economics and the way forward; inertia in the economics profession to accept the new interdisciplinary research; the scientific method; the interplay of culture, behavior and institutions; the role of norms; and a critique of complexity and agent based models. We are publishing this interview as a working paper to foster debate and reflection among economists and other social scientists regarding some of the "big-picture" questions in the behavioral sciences, and to highlight the powerful role that behavioral economics can play in illuminating a deeper understanding of the economy and other social systems.

Keywords: Behavioral economics; scientific method; culture, preferences, and institutions; rationality; heuristics and biases; other-regarding preferences; complexity and agent based models.

JEL Classification: A12 (Relation of Economics to Other Disciplines); B41 (Economic Methodology); D01 (Microeconomic Behavior: Underlying Principles); D91 (Role and Effects of Psychological, Emotional, Social, and Cognitive Factors on Decision Making); Z1 (Cultural Economics, Economic Sociology, Economic Anthropology).

*Professor of Economics, Department of Economics, University of Leicester, University Road, Leicester, LE1 7RH. Tel: +44-116-2522086. E-mail: sd106@le.ac.uk.

†Executive Director, Institute for New Economic Thinking at the Oxford Martin School, Oxford University. Professor of Public Policy Practice at the Blavatnik School of Government, Oxford University. Address: Manor Road Building, Manor Road Oxford, OX1 3UQ. Tel: +44 (0)1865-288895. Email: info@inet.ox.ac.uk.

1 Introduction

1.1 Eric Beinhocker writes...

Some time ago I saw an advertisement for a new book on behavioral economics, Professor Sanjit Dhami's Foundations of Behavioral Economic Analysis (Dhami 2016) published by Oxford University Press. I thought it might be a useful resource, so I ordered a copy, and it duly arrived landing with a thud on my desk. When I began to peruse its 1,798 pages I was enormously impressed by both its incredibly comprehensive scope and its rigor in its treatment of each topic. Here was just about anything one could possibly want to know about behavioral economics captured in one volume. But I was also impressed at Prof. Dhami's thoughtfulness in taking a critical perspective, highlighting both the great strengths and gains made by the field, but also highlighting weaknesses, areas of current debate, where more work was needed, and where exciting opportunities for research lie. I also appreciated that Prof. Dhami not only read and cited works narrowly in behavioral economics, but drew widely from the behavioral sciences, and connected this work not just with more orthodox economic theories, but also with newer and more heterodox perspectives such as complexity economics.

I thought it would be stimulating for faculty, postdocs and students to invite Prof. Dhami to Oxford, take advantage of his encyclopedic knowledge of the field, and engage him in a critical discussion of these issues. What follows is an edited version of this exchange.

1.2 Sanjit Dhami writes...

A few months ago, Professor Eric Beinhocker, Executive Director INET Oxford, invited me to give a seminar in Oxford, which I accepted and proposed to present a recent paper. He suggested an alternative and unusual format in which he would interview me on a range of questions about the behavioral sciences in front of an Oxford audience that could, at the end of the interview, ask me question. I agreed, but if truth be told, not without some trepidation.

Professor Beinhocker ended up putting together a very important, wide ranging, and challenging set of questions for me to answer. The questions required serious thought and analysis. Following the interview, which was not recorded, Professor Beinhocker suggested that it would be beneficial for the behavioral sciences community, particularly young researchers who need to see which way the entire field is headed, to read the questions and my responses. Hence, a need was felt to bring out the interview in the form of a working paper through INET Oxford. I have taken the opportunity to write up the exchange in the form of a working paper that cross-references the relevant literature. However, in

order to prevent getting bogged down in extensive referencing, I have taken the liberty to cross-reference many of the arguments to my 2016 book (Dhami, 2016), which is now being edited, corrected, and updated in the form of 7 brand new volumes (Dhami, 2019, Volumes 1–7); the interested reader can pick up the relevant references from these and other sources.

2 *Assessment of behavioral economics and exciting developments*

Eric Beinhocker: *What is your overall assessment of behavioral economics today? What is the work that most excites you and where are the areas that more progress is needed?*

Sanjit Dhami: I take a fairly holistic approach to behavioral economics and view it as enriching economics by using insights from psychology, sociobiology, sociology, anthropology, neuroscience, complexity, and machine learning.[1] Without being drawn into a formal definition, behavioral economics is any amendment to neoclassical economics that introduces a more empirically supported picture of human behavior and makes falsifiable predictions based on a rigorous theory, certainly no less rigorous than neoclassical economics itself. Neoclassical theory, as originally envisaged by its creators, arguably had a broader scope than the current version of neoclassical models. For instance, adaptive expectations, other-regarding preferences, and the roles of history and culture were part of the discourse and then somehow disregarded. So that there is no confusion, in what follows, when I refer to neoclassical economics, I refer to the *actual practice of the subject*, and not the initial scope envisaged by its founders.

A behavioral economics approach requires, as essential components, the following. (1) Stringent testing of the predictions of the proposed theory using lab, field, and neuroeconomic evidence. (2) A willingness to abandon or modify the theory if, having exhausted all possible confounds to testing, the evidence is not in conformity with the theory. In this, I draw inspiration from good practice in the natural sciences. In introducing the theory of relativity to readers of The London Times on 28 November 1919, Einstein made three specific predictions and then put it down in black and white that if his theory fails on any of these predictions, it is untenable. So a huge part of the appeal of behavioral economics to me is its method, based on a Popperian approach, which borrows heavily from the natural sciences (Popper, 1934, 1963; Lakatos, 1970). This approach is central to physics, as the Nobel Prize winning physicist Feynman (1965) once said along the following lines: We are trying to prove ourselves wrong as quickly as possible, because only in that way

[1] Throughout the interview, I group experimental economics with behavioral economics because I believe that it is unnatural to separate them.

do we make the fastest progress. A second part of the appeal to me is the extent to which behavioral economics has solved important and outstanding problems in economics in such a short span of time. Let me expand on this theme below.

The dominant view in economic methodology is associated with the name of Milton Friedman, a Nobel Prize winner in economics. In this view, economic models need only be true in an *as if sense*– i.e., even if the assumptions are blatantly incorrect and unrealistic, good predictions make a good model (Friedman, 1953). In the predictions turn out to be correct then, the argument goes, we may legitimately conclude that people act *as if* they conformed with the assumptions of the theory. In its worst, and sadly most common form in economics, Friedman's position has been used as a license to build models that are based on ad-hoc assumptions that are justified by using an 'as if' argument. But the predictions are often never stringently tested. For instance, from the observation that birds fly, we may conclude that they act 'as if' they understand the laws of aerodynamics and the closely related laws of hydrodynamics. One may then stringently test this prediction by leaving a bird underwater. If the inference is correct then the bird should try to swim its way out of the water but most birds (unless they learn through evolution to swim under water like gannets) try to fly under water and drown even more quickly.

The initial research agenda in behavioral economics tried to show that many of the central neoclassical predictions were rejected by the evidence, even in an 'as if' sense. Indeed, this was central to the early work of Kahneman and Tversky in the 1970's in the *heuristics and biases program* (Kahneman and Tversky, 1972; Tversky and Kahneman, 1973, 1974). This work continued into the 1970's and 1980's.[2] If there had been greater acceptance of the scientific method in economics, then this should have immediately set off a major and serious examination of a set of core assumptions in neoclassical economics. But Kahneman-Tversky's work, and often even their very existence, goes unnoticed to this day in most University economics courses. This certainly was true in my own University education. At best, this is a form of self-handicapping in the economics profession that is best understood as a desire to preserve the status-quo.

In the last 2-3 decades, as behavioral economics has matured, the research agenda has moved beyond just engaging in refutations of the neoclassical predictions. Behavioral economics has proposed a set of rigorous, falsifiable, theoretical models in almost all areas of economics, that offer more empirically confirmed predictions than its neoclassical counterparts. This is a major achievement but the best is surely yet to come.

There is lots to excite me in behavioral economics and I find it difficult to choose between different areas, given that I have worked in most of them. Purely for illustration,

[2]For a fascinating and thought provoking account of this work, see Kahneman (2011) and Thaler (2015). For a survey that critically looks at the modern developments and considers the interface with policy, see Dhami, al-Nowaihi and Sunstein (2018).

let me choose two polar situations in economics that cover a fair bit of modern economics-*decision theory* (game against nature under risk, uncertainty, and ambiguity) and *game theory* (strategic interaction against nature and other players).

Decision Theory (Game against nature): Prospect theory has strengthened its reputation for being versatile and competent in dealing with situations of risk, uncertainty and ambiguity (Dhami, 2016, Part I; Dhami, 2019, Vol. 1). The evidence for it mounts from all areas of economics and its predictions are being increasingly confirmed from many sources. For instance, *loss aversion*, the tendency for losses to bite more than equivalent gains, is confirmed with human subjects, capuchin monkeys, and in neuroeconomic studies. It has the same median magnitude, about 2.25, in humans and capuchins suggesting that this is a trait inherited from a common ancestor that lived more than 35 million years ago. Furthermore, the neural firing rates in the domain of loss are also about 2.25 those in gains. While there is substantial heterogeneity of loss aversion that is age, experience, context, and emotion dependent, this brings us closer to some sort of universal laws of human nature, which were absent in neoclassical economics. Successful applications of loss aversion, and prospect theory, continue to proliferate in diverse areas of economics, for instance, in contract theory, theory of goals, principal-agent problems, industrial organization, and labor economics.

It is always difficult to predict where the future developments may come from, but here is a list of some broad problems that need to be resolved:

- Explaining ambiguity aversion has proven to be one of the most challenging problems for behavioral and neoclassical theories, although the behavioral theories (e.g., source dependent preferences) do better. However, neither can explain the finding that in Ellsberg experiments there is ambiguity seeking for low probabilities and ambiguity aversion for high probabilities Dhami (2019, Vol. 1, Sections 4.4, 5.3) . This is a robust empirical finding that is often not highlighted. Fresh thinking is required, perhaps *quantum decision theory* which can explain both these phenomena (al-Nowaihi and Dhami, 2017; Wei et al., 2019).[3]

- We need better theories for human behavior at extreme but economically relevant probabilities. For instance, many people simply ignore events of very low probability, but others do not; Kahneman and Tversky (1979) explicitly recognized this issue. We have proposed *composite prospect theory* to take account of this (Dhami and al-Nowaihi, 2010d), but it needs empirical testing. The problem is that we have very

[3]In quantum decision theory, events need not be distributive (unlike Kolgomorov probability theory). So, under quantum decision theory, the event $X \cap (Y \cup Z)$ need not be equivalent to the event $(X \cap Y) \cup (X \cap Z)$. This along with reasonable auxiliary assumption allows quantum decision theory to explain the relevant paradoxes (al-Nowaihi and Dhami, 2017).

little experimental evidence for what happens at extremely low, but economically relevant probabilities (e.g., a house being destroyed by natural hazards, where probabilities of a huge loss could be in the region of 1 in 0.25 million). It is also difficult to conduct these experiments with real stakes. For a discussion of these issues, see Dhami (2019, Vol. 1, pp. 193-201).

- The interaction between risk preferences and time preferences continues to be an active area of research and will do so for the future too. There could always be some residual risk that a reward promised for the future may not materialize. Hence, we need to incorporate both temporal and risk dimensions into our analysis (e.g., should we combine prospect theory with hyperbolic discounting?). For a discussion, see Dhami (2019, Vol. 1, Sections 2.12, 5.1) and Dhami (2019, Vol. 3, Sections 2.3.3, 4.1.1).

- The domain of existing decision theories is risk, uncertainty and ambiguity. It precludes arguably one of the most important domains, *true uncertainty* or *Knightian uncertainty* (unknown unknowns). This is where the heuristics approach and complexity theory/ agent-based models come into their own. More on this later.

- As in other areas in behavioral economics, there is a need to enrich the models by incorporating insights from the literature on memory, attention, and perception. I suspect this could be a slow process because many of the insights do not lend themselves readily to formal models that economists are used to.

Strategic Interaction (Strategic interaction between players) Game theory forms the basis of modern economic theory. Herb Gintis (2009, 2017) argues that it ought to be the foundation for all social science, but his vision for game theory is quite different from the one that is currently practiced in neoclassical economics. The evidence suggests two broad conclusions (Camerer, 2003; Dhami, 2019, Vol. 4). First, the predictions of neoclassical game theory are routinely rejected by the evidence in many different games/domains, when bounded rationality appears to be a factor and even where it is not. Second, when the predictions are in conformity with the evidence, then competing behavioral models that rely on fewer and more realistic assumptions, are also able to make the same predictions (e.g., in the market entry game; see Dhami, 2019, Vol. 4, Section 2.4.4).

There are a couple of emerging topics that I find interesting and exciting in behavioral game theory.

1. Role of beliefs: In neoclassical game theory, beliefs do not directly enter into the utility function, although Bayesian updating plays an important role in determining actions along the equilibrium path of play. In *psychological game theory*, by contrast,

beliefs directly enter into the utility function, in addition to playing their usual role in neoclassical game theory (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009). Such beliefs help us to directly model *emotions* such as *guilt*, *reciprocity*, and *shame*, which play a central role in human interaction. Indeed, we can use these insights to study the formation of societal norms (Dhami, 2019, Vol. II; Section 5.7).

For instance, we have showed that guilt and reciprocity play a central role in the contribution choices of subjects in public goods games (Dhami, Wei and al-Nowaihi, 2019a). Eliciting beliefs is not straightforward. Merely asking individuals to state their beliefs does not work because this method is subject to the *false consensus effect*. In other words, in guessing the beliefs of others, individuals assign to others, their own beliefs. Hence, we need to use cleverer methods of belief elicitation to ensure that the beliefs are valid– a leading method is the method of induced beliefs. For a discussion of these issues see Dhami (2019, Vol. 4, Sections 2.5.5 and 3.5.3).

Recently Dhami, Arshad and al-Nowaihi (2019) demonstrate the power of these methods to explain the economics of microfinance contracts. This paper is particularly convenient to make a number of important points about neoclassical economic theory. But first let me quickly set out the basics. Microfinance borrowers are individually small and lack any security to offer as collateral, so the formal banking sector refuses to lend to them. In neoclassical economics this should be the end of it. However, Grameen Bank, founded by Mohammed Yunus came along, solved this problem, and won the 2006 Nobel prize. In 2017, estimates suggest that 130 million microfinance borrowers took 114 billion dollars in loans.

Consider a two period model. There are two main kinds of contracts. (1) *Individual liability* (IL) contracts (e.g., Indie borrows money in the first period and is given a second period loan only if she repays the first period loan). (2) *Joint liability* (JL) contracts (Joan and Jim borrow money in the first period but both are given a second period loan only if 'both' repay the first period loan, otherwise both are refused a second period loan).

The basic neoclassical model predicts that JL contracts should not exist because in a joint liability contract Joan and Jim will ignore the positive externalities that they create for each other, so they both will not work hard enough.[4] Yet, the initial business model of the Grameen bank was based on JL contracts; the repayment rate was extremely high, and in 2017 it was 99.6%. It is well known that the Grameen Bank started with offering JL contracts (Grameen 1) but many commercial microfinance banks have, in recent

[4]One can invoke asymmetric information to argue that JL contracts perform a screening role, but it is unclear if this alone can account for the existence of such contracts. Peer pressure and social identity have also been invoked to explain the success of JL contracts. However, the empirical counterparts of these constructs are not specified, and these typically tend to be catch-all terms that try to capture the social connnections between borrowers. For a critical discussion of the literature and the references, see Dhami, Arshad and al-Nowaihi (2019), and Dhami (2019, Vol. 4, Section 3.6).

years, moved towards IL contracts (Grameen 2). Our results offer potential explanations for why this might have happened, while neoclassical models that rely on asymmetric information or moral hazard based explanations typically struggle to explain this switch. The neoclassical model is questionable because it invokes an incorrect model of human motivation. In a two-period model we show that the results are governed by two main emotions, *guilt* and *shame*. Illustrative examples below provide the basic intuition (which is confirmed by the empirical evidence).

Example 1 *(The role of guilt): Irene enters into an IL contract, while Gill enters into a two-person JL contract. Gill receives a "private signal" θ from her partner on the effort level that the partner expects from Gill (the partner's first order beliefs). If Gill is guilt-averse, she might feel guilty from exerting effort $e < \theta$. This requires Gill to use her second order beliefs (about the partner's first order beliefs) which directly enter into Gill's utility function; this allow us to model guilt in the sense of Battigalli and Dufwenberg (2007). All the evidence indicates that guilt is a powerful emotion, so Gill is likely to increase her effort level. Irene's contract is economically independent of any partners, so she might lack the motivation to put in effort that Gill has. This could explain why JL contracts have been so successful.*

Another important institutional factor in microfinance contracts is the *public repayment of loans*. Essentially the loan officer visits a village, say, once a month and gathers around everyone to declare their loan repayments. So defaults or repayment delays become public information, and in closely knit societies they induce shame. Working backwards, borrowers might wish to avoid the situation where public censure of inadequate efforts on their projects might occur. However, standard models in microfinance assume that repayments, even if they are in public, should have no effect because they do not model the emotion of shame. In any case all formal models assume private repayment, even when they seek to explain the empirical results they involve public repayments. The next examples shows how shame works in the model.

Example 2 *(The role of shame): Norma enters into an individual liability (IL) contract under public repayment (P); denote this contract by ILP. Public repayment occurs in front of her social group. Norma receives a signal s of the level of effort others in her social group believe is appropriate for her to exert. This signal reflects 'normative expectations' in the sense of Bicchieri (2006). Norma also observes the effort level of others in her social group ('empirical expectations' in the sense used by Bicchieri, 2006). Assume that the normative and empirical expectations are aligned (which is true of our experimental design), so a necessary condition for social norms is fulfilled (Dhami, 2019, Vol. II; Section 5.7). Norma anticipates disutility from shame by letting down the expectations of her social*

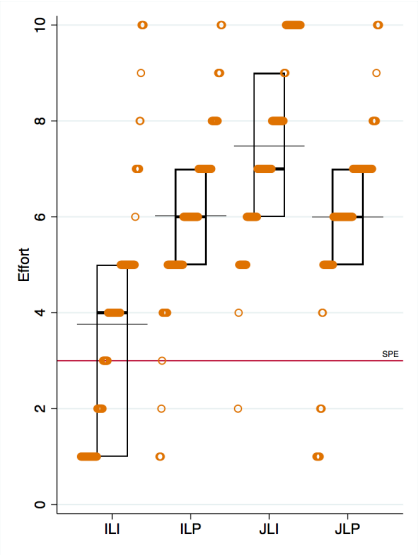

Figure 1: First period effort levels in different contracts/treatments

group in the event that she shirks on her project. This is particularly the case when her group can sanction her. Indeed, in the presence of sanctions and aligned normative and empirical expectations, we have a sufficient condition for the existence of norms (Fehr and Schurtenberger, 2018; Dhami, 2019, Vol. II; Section 5.7). Anticipating this factor, she puts in greater effort and is more likely to repay her loan in public repayment, relative to private repayment (where a third party cannot observe loan repayments).

We put our model to the test using an artefactual field experiment with 400 actual microfinance borrowers in Pakistan. We have a 2×2 design that varies the contracts (individual liability, IL, or joint liability, JL) and the mode of repayment (individual/private, I, or public, P). So, for instance, JLP is a joint liability contract with public repayment and ILI is an individual liability contract with individual repayment. The results on first period effort in the 4 different contracts are summarized in the box and whiskers plot in Figure 1; the distribution of choices for each contract is to be read vertically upwards. The standard neoclassical model predicts an effort level of approximately 3; this is comfortably rejected and there is a great deal of heterogeneity in choices. In the two public repayment treatments (ILP, JLP), the social signal, s (as described in Example 2), is 6 and this is also the actual mean/median effort level, showing the critical role of shame (confirmed separately by regression analysis). The private signal, θ (as described in Example 1), is 6.67 in the JLI treatment (individual repayment) and the mean/median effort is also close to this number, indicating the presence of guilt (confirmed separately by regression analysis).

Thus, guilt and shame both explain why a range of contracts do better than ILI contracts that invoke neither guilt, nor shame. Both IL and JL contracts do well in the

presence of public repayments (ILP and JLP), which indicates the critical role played by shame. Indeed, microfinance institutions have moved away from JL contracts to IL contracts, while maintaining public repayments (the move from Grameen 1 to Grameen 2 noted above). The reason that IL contracts are preferred to JL contracts in recent years is that (1) effort levels are high in such contracts in the presence of public repayments, and (2) they are not as restrictive as JL contracts, which require both partners in a JL contract to repay before any of then gets another loan. Thus, the relevant empirical facts tie in naturally because this is a far more realistic model of human behavior relative to the neoclassical benchmark.

(2) Bounded rationality in strategic thinking. Consider a simple 40 period standard optimization problem in economics in which, in every period, an individual chooses to consume or to save, and in every period, income takes two possible exogenous values–high or low. This is a bread and butter problem in undergraduate and graduate economics courses (except that instead of 40 periods, an infinite number of periods are allowed). In this 40 period problem, an individual needs to make about $2^{40} > 10^{12}$ (one trillion) calculations, i.e., a consumption/savings choice for each of the two levels of incomes in each period to make an optimal intertemporal choice. In this relatively simple problem, the presumption in economics is that the man on the street can do these calculations in his head, in an instant, or indeed people act 'as if' they behaved in this manner!!! All the evidence contradicts this assertion (Dhami, 2019, Volume III). Intertemporal optimization is taught in all Western economics department as the main model describing temporal choice and treated by most of the profession as perfectly acceptable and normal human behavior. This is symptomatic of how unbelievable and unquestioning so many of the core economics models have become. It is even sadder than successive generations of bright students in economics are successfully brainwashed into not questioning the basics and not exposed to reasonably well developed alternatives that exist in behavioral economics.

As noted above, we have already had the bounded rationality revolution in decision theory with the heuristics and biases approach in the 1970's. We need a similar revolution in strategic thinking that incorporates bounded rationality. To be sure, reducing the cognitive burden of calculations and belief formation is central to many of the modern behavioral economics models (e.g., level-k, cognitive hierarchy (CH), quantal response equilibrium (QRE), coarse thinking and categorization, cursed equilibrium, analogy based equilibrium). It remains to be seen if these models go far enough in explaining the relevant evidence. The cognitive requirements in some of these models, particularly the QRE, are still very high in which players play noisy best replies and a fixed point in these noisy best replies is sought–an enormously challenging exercise as the number of players increases.

Do we need to go even further in the direction of players using simple heuristics in strategic situations? I certainly believe that this is long overdue and will perhaps guide

behavioral game theory in the years to come. For instance, could it be that in a strategic situation some players decide to act on the heuristic of always following a moral principle that has a better ex-post chance of being justifiable to them/others, and be willing to take a reduction in material payoffs? Consider Examples A, B, C below that highlight these concerns.

Example A. The very first game that needs to be put up on the board in any course in game theory is the static *prisoner's dilemma game*. This is the most commonly used game in the social sciences and the canonical game theoretic situation that is taught to almost any social science student. Two players (or prisoners) simultaneously choose a strategy– either cooperate, C (the analogue of not snitching on the other), or defect, D (the analogue of snitching on the other). The Pareto optimal payoffs arise if both follow C (in this case they are both let off by the legal system) and the worst possible joint payoffs arise if both follow D (in this case they are both jailed). If one player snitches and the other does not, then the snitch gets a high payoff of 3, while the other gets nothing.

	C	D
C	$2, 2$	$0, 3$
D	$3, 0$	$1, 1$

The unique Nash equilibrium is (D, D); it is in strictly dominant strategies. There is no obvious problem of bounded rationality in this low dimensional and simple problem. Irrespective of the beliefs about the other player, one must never play C the strictly dominated strategy. However, when this game is played once (a static game) we observe that about 60% of the outcomes are (C, C). To see how shocking this result is for any theory that uniquely predicts the play of (D, D), it is like you were asked if you would like to travel in an airplane in a one-off flight and the probability of safely landing at the destination is 40%. Will you take the flight?

How could a game theory course that restricted itself to Nash equilibrium and its refinements hope to go beyond this point? However, not a shred of empirical evidence is ever presented in a typical game theory course (as to why, I defer to a later point in the talk). The outcome (C, C) is problematic for many popular behavioral economics theories as well (e.g., level-k models, where players play a best response to beliefs, so cannot play C). It is also problematic for evolutionary theory; the set of evolutionary stable equilibria is a subset of the set of Nash equilibria, so (D, D) is the only possible evolutionary equilibrium. Of course, group selection theories could result in a play of (C, C), as could other behavioral theories based on a notion of social rationality, such as Kantian rationality, joint intentionality, and team reasoning (more on this below, but also see Gintis (2017) who makes this point very well; see also the guide to further reading in Dhami, 2019, Volume IV).

Ernst Fehr has often said that any reasonable model of other-regarding preferences must first satisfy the experimental evidence (ultimatum, gift exchange, trust, public goods games with punishments and without). Perhaps one could also propose, in a similar vein, that any model of strategic interaction that wishes to be taken seriously must, as a minimum, explain the evidence from the prisoner's dilemma game. In al-Nowaihi and Dhami (2015), we introduce *evidential reasoning* (which has a rich tradition in psychology, and a special case is sometimes referred to as the false consensus in economics) and *evidential equilibrium* into game theory as useful heuristics with which people think about strategic situations. The essence of the idea is this: When uncertain about what other like-minded individuals will do, there is strong evidence that people assign diagnostic significance to their own actions in forming beliefs about what others will do. Further, the default frame for most *conditional reciprocators* is to begin with cooperation and then conditionally reciprocate. We believe that this is the reason that most people choose (C, C) in the PD game.

So perhaps we need a heuristics and biases revolution in game theory too.

At this point, someone trained in neoclassical game theory might protest that it also allows for the cooperative outcome (C, C). However, this requires either an infinitely repeated game (or a game that ends with some probability in each period), or a finitely repeated game in which players have some doubt about the rationality of other players. The explanation of cooperation in this case, elegant in the opinion of some (but elegance is not a scientific criterion), is entirely *instrumental*; people calculate the current benefits and costs of cooperation and if the benefits are high enough, they cooperate. However, people often cooperate because they would like to do so *intrinsically*, not instrumentally. In other words, conditionally cooperative behavior is a part of our preferences. Furthermore, the evidence presented above for the outcome (C, C) arises in the static case where none of the neoclassical game theoretic explanations have any bite.

In some explanations of cooperative play in repeated games, players play mixed strategies. However, the purification agenda used to justify mixed strategies does not have any bite for this class of game. The cooperative results for repeated games hold mainly for public signals received commonly by all the players, but little is known for the more realistic case of private signals. The available results obtain only for private signals that are 'ε close to public signals'; we get no idea of how small ε must be. For the claims in this paragraph, see (Gintis, 2009). It is also noteworthy that the neoclassical solution for finite repeated games uses backward induction, a solution method that repeatedly fails in experiments, for instance, in experiments on the centipede game (Dhami, 2019, Vol. 4, Section 1.2.3).

Finally, one could criticize the experimental evidence from static prisoner's dilemma games on the grounds that experimental subjects mistakenly apply repeated game arguments in a static situation. Indeed, during my talk, Professor Vines raised precisely this

point, citing the well known criticism of Ken Binmore. This is a nice point to illustrate the power and method of experimental evidence. I have said this before in informal conversations with Ernst Fehr that not only has he championed other-regarding preferences in economics, he has also given valuable lessons to the profession on how to choose between competing theories and establish causality by cleverly designed experiments. In this case, Ernst designed experiments to show that the behavior of subjects is very different in static and repeated prisoner's dilemma games, ruling out the possibility that subjects mistakenly apply repeated game heuristics to static choices.

The explanation of cooperation is a central task in the social sciences and the different subdisciplines in it have contrived to offer different and mutually inconsistent explanations. This is the main theme of the book by Gintis (2009). The evidence on human behavior shows that the neoclassical solution is not satisfactory and there is insufficient awareness of this in most game theory courses; indeed, this is hardly mentioned in most game theory courses.

Example B. Do people really optimize in a mathematical sense?

Mathematical optimization lies at the heart of the economic method and undergraduate students are trained very early in the optimization method. However, no empirical evidence is ever cited to justify this choice. In his Nobel Prize speech in 1978, Herbert Simon famously said that no evidence has ever been produced to show that firms produce at the point where marginal revenues equal marginal costs (this is the main result in optimization when applied to firms in economics). It is a bit like a Nobel Prize winner in physics saying in his acceptance speech that no evidence has ever been produced to show that electrons exist. One would imagine that, just like the results from the Kahneman-Tversky heuristics and biases approach, this would have prompted immediate soul searching about basic economic methods. Nothing of this sort happened and just like Kahneman-Tversky, University courses in economics hardly ever mention Herbert Simon's name. The debate is not whether economics is a science or not, but rather if economists understand and abide by the scientific method.

Consider the microfinance example again. The second period is the last period of our game, and there is no future to worry about. However, in the first period, extra diligence in effort makes it more likely that an individual (in individual liability contracts) or a group of individuals (in a joint liability contract) will be able to repay their loan, and so get a second period loan. In other words, the marginal benefit of first period effort is always greater than second period effort. It is then problematic for optimization models, without invoking other auxiliary conditions, possibly ad-hoc, to produce the result that second period effort is equal to or greater than first period effort. However, this is precisely what we get in our empirical results and in particular for the case JLI; see Figure 2. Heuristics such as anchoring on a socially desirable effort level has the potential to explain these results that

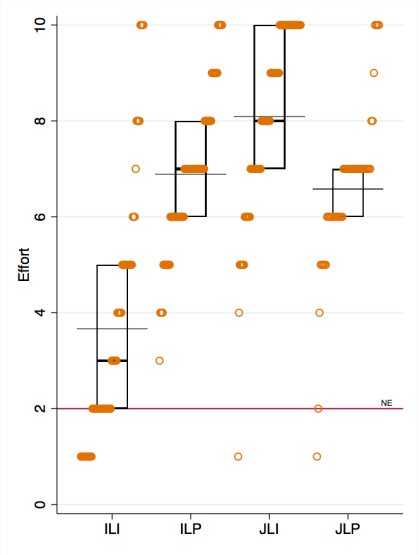

Figure 2: Second period effort levels in different contracts/treatments.

defy the predicted end game effect. By contrast, the end game effects produced in many experimental games typically do not involve the invokation of social norms that underpin the emotion of shame under public repayment or of guilt arising from the private signals of effort from peers.

C. Private rationality or social rationality

Solution concepts in neoclassical game theory rely on *methodological individualism* and *instrumental rationality*; both are present in a Nash equilibrium. I have already noted above that the evidence is not kind to the predictions of neoclassical game theory.

The evidence increasingly indicates that we require a different solution concept that is based on *social rationality* (Gintis, 2017, p. 51; Dhami, 2019, Vol. 4, Section 3.1). This overarching concept includes *team reasoning* but also Tomasello's concept of *joint intentionality* (Tomasello, 2014). Joint intentionality implies that players have a shared common goal and common knowledge of trust in each other to implement the shared goal. In other words, players are able to collaborate, not just coordinate, independently of instrumental reciprocity. A related concept whose implications are being increasingly discussed in behavioral economics is *Kantian rationality* (Roemer, 2015, 2018). In the Kantian protocol, *a player wishes to take the action that he/she would like the other players to take*. Social rationality is then invoked to argue that players believe that other players will also take the same action. This obviously requires players to trust the social rationality of other players. Furthermore, players do not take such actions because they fear reprisals from other players, but rather *because it is the moral thing to do*. This opens a completely new perspective in game theory (including explaining why the strategy (C, C)

is chosen in the prisoner's dilemma game).

Here are a few of the other emerging literatures that excite me in behavioral economics.

- Formalization of heuristics in models of economic theory.
- Formal models of limited attention.
- Formal models of social norms.

Like Herb Gintis I believe that we need to develop formal and rigorous theoretical models to explain the emerging evidence in behavioral economics. I would also encourage and appeal to many experimental economists, who comprise the vast majority of behavioral economists, to be even better versed in the theoretical foundations of the subject. This will quicken progress.

3 Impact of behavioral economics

Eric Beinhocker: *What do you think the impact of behavioral economics has been on the field? Some would note that the field has gone mainstream - most departments have behavioral economists, it is taught in most courses, published in prestigious journals, Nobel prizes have been awarded, etc. But others would argue it has had little impact on the field's core theories and models which still typically rely on simplistic rational actor models?*

Sanjit Dhami: The impact depends, as you note, on the metric one uses. Certainly, if you go by the growth in published peer reviewed papers, freshly minted PhDs in the area, Nobel prizes, Presidents of the AEA, there has been impressive growth.

However, this growth has come from a relatively small base, so the percentage of behavioral economics papers is still relatively small, probably less than 5%, if that. It can sometimes be amusing to observe the simultaneous presence of mutually contradictory behavioral and neoclassical papers in the top journals. But then Fama and Shiller also shared the Nobel prize in economics. This suggests that we are passing through an unsettled time in economics.

The percentage of behavioral economics faculty is small and many top departments have no more than a handful of behavioral economists. The situation is abysmal when it comes to teaching behavioral economics. (1) Not only is the number of courses very small, we have in recent years not observed a dramatic growth either. (2) A large number of the existing behavioral economics courses are too idiosyncratic and based on the instructors' own research or research interests, hardly giving students a chance to develop a rounded understanding of the subject. This is academically perverse. There appear to be very few rigorous, coherent, comprehensive courses in behavioral economics at the moment.

By now, we should have had a far more standardized behavioral economics curriculum across the board that should be able to teach the behavioral analogue of a standard course in microeconomics– only much better because it would be in better conformity with the evidence.

So why are we in a situation where a framework (behavioral economics) that outperforms the existing status-quo (neoclassical economics) in explaining the evidence has not caught up strongly enough in the profession? You might also sometimes wonder at INET Oxford why your own work on complexity, agent based models, and machine learning has such little presence and impact within mainstream economics. It is important to understand the reasons; I outline my views below.

A benign explanation is that there is natural inertia to new ideas in any profession. But this does not explain the continued popularity of expected utility theory and indeed many of the core features of behavioral economics such as exponential discounting, self-regarding preferences, emotionless deliberation, and Nash equilibrium. Luce and Raiffa (1957) commented (p. 35) on "utility theory and rational choice" by writing that: "reported preferences almost never satisfy the axioms". On p 37, they described these refutations as "bolstered by a staggering amount of empirical data". Now we are at a stage where it appears pointless to test expected utility any further (Dhami, 2019, Vol. 1). Similarly, the demonstrated failure of the "as if" position held by economists, in Kahneman and Tversky's work on judgement heuristics in the 1970's, has had very little impact on mainstream economics. By contrast, in behavioral economics, it is beginning to have an increasing impact (Benjamin, 2018; Dhami, 2019, Volume 5; Dhami, al-Nowaihi and Sunstein, 2018).

My preferred explanation for the reluctance to accept the evidence is that many economists need better training in methodology and the philosophy of science. Many of the current homegrown methodological positions in economics, propagated by the leaders in the field are sadly retrogressive and bear no relation to the methodology of science or to the philosophy of science. I deal with these issues in more detail in the introductory chapter of Dhami (2016), so I will be brief here.

It is tragic and frustrating that standard texts in microeconomics, game theory, and contract theory, contain almost no empirical evidence. Many other mainstream applied microeconomics fields do only marginally better. This is in sharp contrast to the natural sciences. Here is my favorite quote on this situation from Herb Gintis (2009, p. xvi):

"*Economic theory has been particularly compromised by its neglect of the facts concerning human behavior... I happened to be reading a popular introductory graduate text on quantum mechanics, as well as a leading graduate text in microeconomics. The physics text began with the anomaly of blackbody radiation,....The text continued, page after page, with new anomalies...and new, partially successful models explaining the anomalies. In*

about 1925, this culminated with Heisenberg's wave mechanics and Schrödinger's equation, which fully unified the field. By contrast, the microeconomics text, despite its beauty, did not contain a single fact in the whole thousand-page volume. Rather the authors built economic theory in axiomatic fashion, making assumptions on the basis of their intuitive plausibility, their incorporation of the "stylized facts" of everyday life, or their appeal to the principles of rational thought....We will see that empirical evidence challenges some of the core assumptions in classical game theory and neoclassical economics."

So how and why has this situation come to be? An important factor is that many otherwise reasonable economists use the "as if" argument to justify almost any assumptions, including completely false assumptions, without bothering to either test them or offer clear tests for them (compare this to Einstein's 3 clear proposals to test his theory of relativity). Many others misunderstand the scientific method. Here are a few anecdotes from personal experience. Many years ago, a reasonably well known economic theorist was appalled that I should suggest testing his model– "but my model is a branch of mathematics, what has empirical testing got to do with it" he protested. A prominent game theorist, when challenged on the evidence for his applied game theory model in a seminar in which I was present, responded that he was only interested in the "mathematical structure of game theory".[5] An editor of a 4* journal once told me that he immediately rejects experimental papers because they are too easy– "theory is hard and more important" he insisted. Beyond these subjective personal anecdotes, there are published accounts from some of the leading economists that give a similar impression and reflect homegrown and entrenched positions within economics. Here are some well-known published examples.

Dekel and Lipman (2010, p.264) write: "Hence the choice of a model will depend on the purpose for which the model is used, the modeler's intuition, and the modeler's subjective judgment of plausibility....One economist may reject another's intuition, and, ultimately, the marketplace of ideas will make some judgments."

Gilboa et al. (2014, p. F. 516) write: "In particular, we agree that: economic models are often viewed differently than models in the other sciences; economic theory seems to value generality and simplicity at the cost of accuracy; models are expected to convey a message much more than to describe a well-defined reality; these models are often akin to observations, or to gedankenexperiments; and the economic theorist is typically not required to clearly specify where his model might be applicable and how."

Rubinstein (2006, p. 882) writes: "As in the case of fables, models in economic theory are derived from observations of the real world, but are not meant to be testable. As in

[5]My long time co-author Ali al-Nowaihi, a mathematician by training, often points out to me that the maths used in most cases in economics is at least 100 years old, or more. I should also point out, in comparison, string theory that many adherents of the "as if" use to justify their method, has not been rejected, and it has produced new maths. To the best of my knowledge economists have not produced new maths, although they have produced many interesting and important applications to economic problems.

the case of fables, models have limited scope. As in the case of a good fable, a good model can have an enormous influence on the real world, not by providing advice or by predicting the future, but rather by influencing culture. Yes, I do think we are simply the tellers of fables, but is that not wonderful?"

The introductory chapter in Dhami (2016) rejects each of these positions, but then it is not hard to do so. These positions should also inform us what anyone who desires to see quick changes in the status-quo in economics is up against.

So, yes, in answer to your question. The field's core theories and models still typically rely on simplistic rational actor models. But we should continue to point out the evidence, and build better models of economic behavior that are based on a more realistic picture of human behavior. We should also continue to robustly challenge the core assumptions in neoclassical economics that are decisively rejected by the bulk of the evidence arising from several decades of research. These include among other things: expected utility theory, exponential time discounting, universal self-regarding preferences, Nash equilibrium, mixed strategy Nash, and refinements, unbounded rationality, unlimited attention, and emotionless deliberation. And we should stay hopeful. My long-time co-author, the mathematician Professor Ali al-Nowaihi, draws an analogy from stochastic social dynamics (see Dhami, 2016, Ch. 16; Dhami, 2019, Vol. 6) to argue that the end for rejected models will come suddenly and unexpectedly like the fall of the Berlin Wall and the collapse of the Soviet Union as we tip into the basin of attraction of a good equilibrium. However, I believe that the end will come more gradually from the next generation of newly minted PhDs if they are trained better in the evidence and in the new behavioral models. I see it a step forward that several student movements in the West now explicitly demand a more empirically relevant economics to be taught in their curriculum.

4 Interdisciplinary work and economics

Eric Beinhocker: *Over the past couple of decades there has been a revolution across the behavioral sciences, including advances in cognitive science, neuroscience, anthropology, social psychology, evolutionary psychology, etc. The picture that this work portrays of Homo sapiens - inductive, heuristic, instinctive, emotional, prosocial, and with strong biochemical underpinnings to decision making and behavior - looks very different from Homo economicus, even the behavioral economics version. Do you think that this interdisciplinary work is being incorporated into economics?*

Sanjit Dhami: This interdisciplinary work looks very different from Homo-economicus because the latter is an inaccurate description of human nature and human motivation, so no surprise there. Unlike what might be implied by the question, the heuristic, emotional, prosocial underpinnings of behavior are already central to behavioral economic theory

and the evidence base. The evidence on biochemical underpinnings is being increasingly incorporated into behavioral economics although we have less in the way of theory here.

The inductive bit is the one that has seen inadequate recognition in behavioral economics. But that is changing now. My new Volume VI on Behavioral Models of Learning, forthcoming in December 2019, devotes a long chapter to these issues. So perhaps I will talk about this more when I take up your question on complexity and agent based models later on. I anticipate that behavioral economics will more fully incorporate the directions that you mention above, in due course. For that we shall need economic models that critically use these new features and made successful predictions in answering important and striking economic questions, which the standard homo economicus model cannot.

Behavioral economists routinely suffer from many neoclassical critiques which often display both ignorance and arrogance, but are ultimately driven by the desire to maintain the status-quo homo-economics model. Perhaps I need to make another reverse criticism here that was driven home to me even more during my recent email exchange with a respected evolutionary biologist in which you along with several prominent social scientists and several Nobel Prize winners in economics were copied-in. Some of the most vociferous critics of the standard homo economicus model, outside economics, do not fully appreciate its scope and criticize it on grounds that are demonstrably incorrect. Economists who practice the homo-economicus model may subscribe to an empirically rejected model but they are not stupid. Many of them are really very clever people and it becomes highly counterproductive to lecture them using an incorrect version of the homo-economicus model.

I was also struck with the realization, in the same exchange, that well meaning critics who should actually be natural allies of this nascent discipline have a limited appreciation of its scope. The common strawman of behavioral economics as a few well meaning nudges and a few undisciplined scattered heuristics does great disservice to this newly emerging field although the popular press is relatively more guilty in this regard. Incidentally the material on these two topics is just 10% of my 1800 page book (Dhami, 2016). What about the remaining 90%? I cannot stress enough how important it is to teach rigorous and comprehensive courses in behavioral economics to the next generation of economists and social scientists. I do my bit from time to time to teach such courses at home and also at the Kiel Institute for the World Economy in Germany. I am really looking forward to a similar course in the Bavarian graduate program this year.

5 Utility theory and rational choice

Eric Beinhocker: *Let's take a specific example - utility theory and rational choice. This of course has been a fundamental construct in economics since the Marginalist revolution in the late 19th century. Behavioral economics of course has pointed out many empirical*

problems with utility and rational choice, at least as neoclassically constructed, and other behavioral scientists have painted a picture of decision making that looks very different than optimizing a utility function. But many economists still defend utility as a "good enough assumption", others such as Herb Gintis have tried to recast rational choice theory to better fit the empirical facts, and still others call for new more cognitively based approaches? What are your views?

Sanjit Dhami: Let us first understand what we mean by utility theory and rational choice. I have just dealt with this topic in a recent paper with Ali al-Nowaihi that is forthcoming in the MIT Handbook of Rationality, which the interested reader could consult for more details (Dhami and al-Nowaihi, 2019). At its most basic level i.e., choice under certainty, rationality implies that individuals have well-defined preferences that satisfy reflexivity, completeness, and transitivity– nothing more nothing less. In particular humans are not assumed to carry a utility function in their heads. So Herb is among those who advocates that economists should not abandon this minimum degree of rationality.

How can we account for evidence that violates well-defined preferences, violates completeness, and violates transitivity. Herb would not deny this evidence, but his response would be that the supposed problems arise because we do not recognize that individuals make decisions that are context, frame, and state dependent. In Herb's view, what appears to us as violations of these features really is our inability to account for the context, frame, and state dependence of preferences. This is a potentially testable proposition but it has not been fully tested because it might be difficult to write down all possible contexts, frames and states in an objectively verifiable manner. The evidence for context, frame, and state dependence of preferences is unassailable, some classic experiments can be found in the early work of Kahneman and Tversky. But I would be surprised if we could account for all the violations by invoking context, frame, and state dependence of preferences in any tractable and parsimonious manner.

There are separate axiomatic developments that give rise to corresponding notions of "utility theory and rational choice" under a range of alternative scenarios, for instance, under risk, uncertainty, and ambiguity; under time discounting; and under strategic interaction. These give rise to utility representations as captured in expected utility theory, exponential time discounting, and Nash equilibrium and its variants. Each of these core features of neoclassical economics is empirically rejected, some worse than the others (see Parts 1-4 in Dhami, 2016; and the first 4 volumes of Dhami, 2019). Herb does defend expected utility theory (EU) in his work but he takes a very broad view of expected utility that borrows everything from prospect theory (reference points, loss aversion, shapes of the utility function), except for non-linear probability weighting. He is right to point out that this is consistent with the von-Neumann–Morgenstern axioms. However, this is neither the way EU is used in economics, nor were these insights discovered by researchers

working on EU. Furthermore, as argued in Dhami (2016, Part 1) and Dhami (2019, Vol. 1), there is very strong evidence on non linear probability weighting that is inconsistent with the von-Neuman-Morgenstern axioms and it cannot be ignored.

Economists do defend EU on the grounds that it is a good approximation. But good approximation to what? If we are interested in EU being a good approximation to actual human behavior, then this assertion is simply incorrect. The predictions of EU are typically quantitatively wrong. Economists often make an argument for the qualitative correctness of EU. In a two horse race between expected utility theory and prospect theory, Dhami and al-Nowaihi (2007, 2010c) show that the quantitative predictions of EU are wrong by a factor of about 100 and the qualitative predictions (comparative static effects of tax rates on tax evasion) are also incorrect. Rabin (2000) showed that any risk aversion over small stakes under EU implies infinite risk aversion over large states. Attempts to discredit this result often end up relying on reference dependence, a central construct in prospect theory.

EU is now buried under a mountain of evidence that rejects it (Dhami, 2019, Vol. 1). Its continuing use in economics as the main decision theory under risk and uncertainty is puzzling and indefensible, especially given that we now have a far superior theory, prospect theory, that is axiomatically founded and is in much better conformity with the evidence. Of course if you take EU out, you are taking out a large chunk of the foundation of modern neoclassical economics (many of the existence results, e.g., existence of a Nash equilibrium; and incentive compatibility conditions in agency theory) and it is possible that the rest of the edifice may not survive. But this should not prevent us from rejecting EU if the evidence does not conform with it, otherwise we will be unable to make progress in economics.

6 Other-regarding preferences

Eric Beinhocker: *Another hot topic is "other regarding preferences". While rational choice doesn't explicitly exclude such preferences, traditionally economists have modelled agents as individualistic and self-regarding. Yet we have a huge body of work now, which you reference in your book, on human prosocial and cooperative instincts and behaviors. How can we incorporate other regarding preferences in core economic theories?*

Sanjit Dhami: You are right in pointing out a common misperception among some economists that rationality implies or is implied by self-regarding preferences (Dhami and al-Nowaihi, 2019). Self-regarding preferences (nothing to do with rationality) have a long intellectual history in economics going back to the classical economists. However, we now have a great deal of evidence on other-regarding preferences and it is already being incorporated into behavioral economics. I should add that this incorporation is often undertaken within an optimization framework that keeps intact many other features of

Bayesian rationality that we know are rejected by the evidence. So any rejection of these models is a joint rejection of other-regarding preferences and Bayesian rationality. This has ultimately to do with how we should build a new economics– from scratch, or by relaxing 1-2 features at a time. Opinions differ on these two choices.

Economists have been very resistant to explaining economic phenomena by altering tastes or preferences. They feared that this would lead to arbitrary preferences and ad-hoc explanations. However, this is not how the behavioral economics literature has progressed. It has been disciplined by the evidence and the general requirement is that any wannabe model of other-regarding preferences must explain the experimental evidence. Ad-hoc models can hardly account for the experimental evidence arising from a range of experimental games. By contrast, given the extent of the evidence that we now have on other-regarding preferences, it is undesirable to place exclusive reliance on self-regarding preferences. Models that allow for a mixture of other-regarding and self-regarding preferences are much more promising (Dhami and al-Nowaihi, 2010a,b).

The initial aim of theoretical models of other-regarding preferences was to explain the evidence arising from experimental games. This gave rise to a set of theoretical models that were important in applications later on. Some of these included models of inequity aversion (Fehr and Schmidt, 1999), belief based reciprocity (Rabin, 1993); combinations of inequity aversion and reciprocity (Falk and Fischbacher, 2006); and others such as type-based reciprocity and those based on a tradeoff between efficiency and inequity (see references in Dhami, 2016 or in Dhami, 2019, Vol II). These models, in turn, gave rise to rich applications that have permeated most of the core areas in economics already. For instance, and purely for illustration of the range, contract theory (Brown et al., 2004, Fehr et al., 2007), theory of the firm (Bartling et al., 2013), principal-agent models (Itoh, 2004; Englmaier and Leider, 2012), and political economy (Dhami and al-Nowaihi, 2010a,b).

I will briefly comment on two further literatures. In the first, there is increasing incorporation of human morality into behavioral economics (Dhami, 2019, Vol. II, Chapters 3, 5). In this literature, individuals exhibit moral behavior *intrinsically* (desired for it's own sake) rather than *instrumentally* (as in repeated games in neoclassical game theory). As is the case with many other areas in behavioral economics, the empirical evidence provided the first breakthroughs in our thinking about morality. There is a separate literature that talks about moral philosophy and ethics, but here I am interested in the experimental evidence on human morality and the theoretical models. This evidence shows that humans behave nothing like the textbook selfish black liars model (see Dhami, 2017 for this terminology). While human morality may be loosened with increased incentives, a significant fraction of the individuals find it hard to lie and many lie only partially, even when full lying could have increased their material payoffs (Fischbacher and Fölmi-Heusi, 2013). In some cases, such as in the lost wallet experiments, when the monetary payoffs from lying

are increased, people may lie less because they must counter the greater guilt arising from a bigger lie (Cohn et al., 2019). For a recent survey of this literature, see Dhami (2017).

The empirical evidence has led to a formal modelling of the moral costs of breaking promises (e.g., Ellingsen and Johannesson, 2004) that tie in nicely with the experimental evidence on why ex-ante communication (e.g., promises, intentions, and threats) alters outcomes in games. A parallel literature has tried to examine human motivation when we tend to signal to ourselves or to others, our moral values (Bénabou and Tirole, 2003, 2006). Morality is often underpinned by emotions such as shame and guilt (Elster, 2011 and Bicchieri, 2006); properly microfounding these emotions requires the machinery of psychological game theory, as in our microfinance paper described above.

The second literature is work done on social identity (Akerlof and Kranton, 2000, 2005). Extremely fruitful empirical studies have continued to proliferate but we need greater incorporation of these insights into our theoretical models; for a recent study that does both in the context of identity and redistribution, see Dhami, Manifold and al-Nowaihi (2019) and the references cited therein.

7 Endogenous preferences and their interaction with the environment

Eric Beinhocker: *A similarly hot topic is endogenous preferences. Again, standard models depict preferences as fixed and exogenous. But we know that they evolve over time, are socially and contextually affected, and there are dynamic interactions between preferences, behaviors, and the environment (institutions, technology, etc.). What is your view on this? Where is the state of the art?*

Sanjit Dhami: A great deal of the literature in behavioral economics shows that preferences are malleable, context dependent, and influenced by social norms, public information campaigns (e.g., the anti-racism drive in UK club football; HMRC "eyes" adverts on tax evasion; public campaigns to reduce teenage pregnancies and smoking in the UK), and private advertising, which is a multi billion pound industry. Preferences can also change during the course of an experiment when learning takes place. Emotional states can influence preferences, for instance, measured loss aversion increases when disgust is induced; projection bias can influence how the current state of the world influences our expectations of future states (hungry shoppers buy more groceries in their weekly shopping); preferences for prosociality are culture-dependent (e.g., results of ultimatum games in different cultures, as in Henrich et al., 2001).

However, there is also a sense in which preferences, particularly where the environment has been stable, also have a fairly stable component. For instance, at least over the not

too long run, most people order the same favorite Italian, Indian, or fast food; watch the same favorite genre of films; read the same favorite classical authors. So, preferences are certainly not completely random and unpredictable either. However, for new products, for which individuals have not had prior experience, there does appear to be a greater scope for forming preferences. As John Kenneth Galbraith and Robert Heilbroner stressed a long time ago, large corporations not only create outputs, they also create the corresponding wants and preferences, implying that consumer sovereignty is not absolute.

Advertising often conveys little information about a product. If preferences were fixed, then we should probably not observe as much advertising expenditure as we do. Of course, we might not even know what our preferences are. Ariely et al. (2003) demonstrate a phenomenon known as *coherent arbitrariness*, in which the initial preferences of individuals are arbitrary/malleable but once they are formed, they are coherent in the sense that they respond to economic variables such as prices, leading to downward sloping demand curves.

There has been a traditional aversion in neoclassical economics to allow for endogenous preferences. It assumes stable, fixed, self-regarding preferences, and it views attempts to change this assumption as ad-hoc tinkering to explain the evidence. Evidence for endogenous preferences is either dismissed outright, or it is argued that the researcher has missed some other relevant factors that result in this empirical confound. This is not borne out by the evidence.

Preferences are influenced by institutions too. Anti-social punishments in public goods games are observed in societies where the rule of law and norms of civic cooperation are low, while pro-social punishments are observed in the opposite case (Herrmann et al. 2008). I take this theme up in a later question in more detail.

If preferences are endogenous, then welfare economics becomes a lot trickier. Should a social planner aim to maximize the utility function of an individual that exists before a social intervention, or the one that is formed after the intervention, or some combination of the two, or neither? Behavioral welfare economics is extremely difficult, unlike the sanitized version that we see in a chapter on welfare economics in microeconomics books; for an introduction to the subject, see Dhami (2016, Part 8).

One problem is that there are relatively few rigorous falsifiable theories of endogenous preference formation that have been tested against the data. Nor are there watertight empirically verified causal mechanisms that tell us how endogenous preferences form. As an illustration of what this literature tries to do, consider Guiso et al. (2008a) who use an overlapping generations model in which trust is transmitted from one generation to the next. Trust, in turn, fosters the desire to trade. As is often the case in this class of models, there are multiple equilibria. In the no-trust no-trade equilibrium, parents have low trust (or even mistrust) which is transmitted to their children who also have low levels of trust, which leads to low trade, and hence low trust. The converse happens in the high-trust

high-trade equilibrium. Empirically verifying such equilibria and establishing causality is not easy.

The state of the art would be to develop the relevant theories and establish causal mechanisms through experimental and field evidence.

8 Co-evolution between behavior, culture, and institutions

Eric Beinhocker: *Another area of growing attention, is the co-evolution between behavior, culture, and institutions. This has historically been territory for sociologists, anthropologists, and political scientists, but there is also a tradition of this in economics (e.g. Ostrom). Do you see behavioral economics contributing to this work? Interesting opportunities?*

Sanjit Dhami: Institutions can be defined as formal constraints on human behavior such as the legal system, rules, and written or unwritten constitutions (North, 1990). By contrast, culture may be defined as informal constraints on human behavior, such as norms and conventions in a society (North, 1990, Alesina and Giuliano, 2015). Cultural values are typically transmitted in ethnic, religious and social groups from one generation to another, often in a relatively unchanged manner. For instance, Luttmer and Singhal (2011) find that the redistribution preferences of second generation immigrants are correlated with the extent of the welfare state in their parents' country of origin. One may also define culture as the set of beliefs and preferences acquired by non-genetic transmission (Bowles and Gintis, 2010). Social learning is one channel of cultural transmission and takes the forms of vertical (parents to children), horizontal (peer to peer), and oblique (nonparental elder to younger) transfer of information.

One of the best known works on the link between culture and institutions in behavioral economics is Henrich et al. (2001). They find that prosociality is positively correlated with: (1) Market integration in the community (predominance of buying/selling for a wage). (2) Cooperation in production. The opposite is taught in most introductory courses in economics, namely, that markets eliminate prosociality. Markets may reduce ethicality, but not prosociality, as I point out in Dhami (2017). There has also been an examination of the link between individualistic preferences (that promote innovation but also less cooperation) and collectivist preferences (conformity with these preferences reduces innovation but there is more cooperation) on the one hand, and growth on the other (Gorodnichenko and Roland, 2018).

Effect of preferences on institutions

Weber's account of the Calvinist doctrine of predestination and capitalist development

in Northern Europe provides a link between culture and development; in al-Nowaihi and Dhami (2015), we link this to evidential reasoning in games. In this context, there is a question in the World Values Survey (WVS) that asks if success is determined by hard work or luck. The percentage of people who believe in the relative importance of hard work is greater in the US relative to Europe. Another similar question in the WVS: whether the respondent believes that the poor could become rich if they tried hard enough? Alesina and Glaeser (2004) use this to explain why the welfare state is more generous in Europe; clearly you are more willing to help the poorer if their income arises from luck rather than an unwillingness to work hard.

Trust is by far the most studied cultural trait in experiments. There are several findings that I can think of. Education is associated with higher trust, and history matters. For instance, Africans whose ancestors were raided during the slave trade display lower trust, relative to those who were not raided (Nunn and Wantchekon, 2011). In general, communities that have been discriminated in the past show lower levels of trust. When the trust game is played, a robust finding is that the ingroup members are trusted more than outgroup members. For instance, Fershtman and Gneezy's (2001) study showed such ingroup effects with descendants of Ashkenazi Jews who have a European origin and Eastern Jews who have an African or Asian origin.

There is also a relationship of the strength of family ties with labor force participation and political participation (Alesina and Giuliano, 2014), and with the extent of family business in society (Bertrand and Schoar, 2006). The relative presence of nuclear families and clans explains different patterns of urbanization in Europe and China (Greif and Tabellini, 2012).

Consider the following fascinating account of how preferences influence institutions and how accidents of history determine current institutions and values. Successive waves of immigrants into the US brought with them their own cultural values that got translated into different institutional arrangements (Fischer, 1989): Puritans from East Anglia believed in education and order, so they introduced laws promoting universal education. Virginia Cavaliars came from the South West of England and their beliefs accepted income inequality, so they promoted low taxes and low government spending, including on education. Quakers came from NW Midlands in England and valued personal freedom, so they promoted laws emphasizing equal rights.

Cultural beliefs can influence institutions and actual economic outcomes. For differences in cultural beliefs of Maghribi and Genoese traders leading to different institutions, the first based on informal legal and enforcement institutions and the latter on formal ones codified in contract law, see Greif (1994). Trust, interpreted as a cultural variable, is positively correlated with several indicators of financial markets, e.g. use of cash, participation in the stock market, and use of bank loans versus loans from friends (Guiso et al.,

2004, 2008a, 2008b).

Effect of institutions on preferences

Institutions can also change preferences. For instance, Carpenter and Seki (2011) find that fishermen who work under shared costs/benefits in their real life job, called the poolers, are also more cooperative when playing the public goods game, relative to the non-poolers who operate on an independent basis in their real life job.

The unification of Germany and the collapse of the Soviet Union provide two interesting examples of the effect of a change in institutions. Did 50 years of East/West Germany lead to different cultural attitudes? Early studies find no difference, suggesting that institutions do not influence culture; but some of this work had a small sample size. However, with a larger sample size, Alesina and Fuchs-Schündeln (2007) find that there are differences between the two peoples in how pro-government they are (Eastern Europeans were more pro-government). However, any differences are closing fast and the authors predict they might be eliminated in 2 generations.

Experience of war can change some preferences but not others. Using a sample of 681 Muslims, Croats, and Serbs in postwar Bosnia–Herzegovina, Whitt and Wilson (2007) identify ingroup favoritism in dictator game experiments, which suggests the effects of war on preferences. Yet, a norm of fairness prevails in the sense that dictators transfer to outgroup members some socially acceptable fraction of their endowment; this component of preferences was not changed by war. People in Burundi who have experienced war related violence share more with neighbors, in a field experiment (Voors et al., 2012).

Herrmann et al. (2008) find that antisocial punishments are higher in societies where the norm of civic cooperation and the rule of law are weak. Fishermen who practice more cooperative modes of fishing (bigger hole sizes while fishing on a common property resource, a lake, allowing young fish to escape) also contribute more in public goods games experiments (Fehr and Leibbrandt, 2011). There is less anti-social punishment (a cultural value) when the punishments are originally decided by a referendum among members (institutions) (Tyran and Feld 2006; Ertan, Page, and Putterman 2010; and Sutter, Haigner, and Kocher 2010)).

Increases in female labor force participation (institutions) gave rise to family size reductions, rise in feminism, and changes in sexual practices (culture) (Bowles, 1998). Repeal of the law of "no shopping on Sundays" (institutions) alters the opportunity cost of participating in Sunday church and reduces church attendances/donations (culture) (Gruber and Hungerman, 2008).

There are also models of the two-way interaction between culture and institutions. Tabellini (2008) considers the interaction between morality (culture) and legal institutions. Better legal institutions reduce the need for private punishments and may also increase the proportion of cooperators. When there are more cooperators, it reduces the load on

the legal system through, say, reduced caseload.

Norms of cooperation may change quickly due to inspired leadership or other institutional changes (Acemoglu and Jackson, 2012). Georgia (as in country) eliminated police corruption within a few days in a dramatic move following the election of the new government in November 2003 on an agenda of fighting corruption. The reforms targeted the ministry of internal affairs and immediately fired a large fraction of the police force, particularly the traffic police which was particularly corrupt and began a new recruitment drive. However, in other cases such norms are hard to change and survive changes in leadership and political parties.

9 Complex adaptive systems

Eric Beinhocker: *Much of our work at INET Oxford is framed by a complex adaptive systems view of the economy. In that frame macro behavior emerges bottom-up from the interactions of agents in networks and institutions, and so a realistic understanding of behavior is fundamental to that agenda. Techniques such as agent-based models enable the incorporation of more behavioral realism into models, but it does require abandoning some core economic ideas (e.g. methodological individualism, equilibrium). What are your thoughts on this agenda and approach?*

Sanjit Dhami: Let me say upfront that I am very excited about the agenda and approach of complex adaptive systems of which agent based models (ABMs) is the leading example in economics. In answer to your question, we should follow the evidence. If it requires abandoning any of the core economic ideas, so be it. How else do we make progress? But this realization requires training and appreciation of methodology and the philosophy of science. As noted earlier, such training is missing among many economists. My long time co-author Professor Ali al-Nowaihi and I are writing a small monograph on this subject based on our Professorial inaugural lectures in 2012, and we have made progress with an initial draft, which is very exciting. So watch this space.

Let me now point a few pros and cons of the complexity approach, as applied to economics. I assume that the reader has a basic knowledge of these concepts; for an introduction to complexity, agent based models, and machine learning, with the relevant references for the claims made below, see Dhami (2019, Vol. 6, Chapter 4) and the references cited therein.[6]

Advantages of complexity theory relative to neoclassical economics

[6]For several other important themes and useful discussions of the complexity approach not covered here, see Beinhocker (2006, 2016), Beinhocker et al., (2019), and Farmer et al., (2019). Gintis (2017) makes an attempt to fuse behavioral insights and an evolutionary complex systems view, though he uses game theory as his primary tool rather than ABMs. Young (1998) uses stochastic social dynamics and low level individual rationality to explain the emergence of social norms and other system-wide properties.

I start with outlining the advantages, but only briefly, because the Oxford audience in this hall is well aware of these arguments. Then, I will focus on the drawbacks, as I see them.

First, as noted above, neoclassical economics has very little to say in situations of true uncertainty, possibly the most common situation for real life decision making. Surprisingly, this basic awareness eludes many in economics and in behavioral economics.

Second, neoclassical economics has chosen to focus solely on equilibrium analysis. This has been most unfortunate because economists have substituted a hard but empirically more relevant question (how do we understand the dynamics of an economy?) for an easier, but not by any means easy, and less empirically relevant question (what happens in equilibrium?). Given that exclusive reliance is put on equilibrium analysis in University courses in economics it is surprising how few examples are ever given to justify such an analysis. The reasons for this are probably best understood in terms of my earlier comments about methodology in economics.

Third, macroeconomic theory has been seduced by the representative agent model, when all the evidence points to the importance of heterogeneity of economic agents. This is not to say that economists do not recognize heterogeneity of economic agents. Indeed many interesting models in economics and in finance rely on simple forms of agent heterogeneity. However, macroeconomics and growth theory have largely, not exclusively, shied away from agent heterogeneity. Perhaps this partly stems from the desire of economists to maintain the assumption of equilibrium analysis (harder to tackle under heterogenous agents) and to value models that can be solved analytically rather than numerically. There is, of course, no scientific basis for the preference for analytically solvable models but historically rigorous economic theory developed at a time when computing resources were far more limited, which favoured the emergence of such models.

Under true uncertainty, non-equilibrium analysis, and agent heterogeneity, complex adaptive systems provide us with a very useful alternative, particularly in macroeconomics.[7] By contrast neoclassical economics does not make predictions in this case. The main benefits of ABMs is that they are able to incorporate unprecedented levels of institutional detail within the confines of the model; agents use believably simple adaptive rules of thumb; and equilibrium is not imposed on the system. By contrast, neoclassical models are too reduced-form to reflect the underlying richness of many phenomena and they do not deal with true uncertainty. In other words, they are far too parsimonious to reasonably capture the empirics of many real world economic phenomena– their parsi-

[7] Under true uncertainty, the heuristics and biases approach already informs us about the way forward in microeconomics. This is not the way that the heuristics and biases approach is viewed in the literature, but it is an approach that I have been trying to push and one that I believe will be increasingly developed in the future.

mony is designed to fit analytical solutions not necessarily to capture all relevant aspects of economic phenomena.

Another attractive feature, at least to my mind, is that fluctuations in ABMs typically arise not from external shocks, but from the endogenous behavior of adaptive agents. This gels well with many observed phenomena that are puzzling for neoclassical economics, such as stock market crashes without any accompanying news about changes in fundamental values. To be sure, there is an earlier and quite well developed literature on self-fulfilling prophecies in macroeconomics that arose in the 1970's and 1980's, but it relied on rational expectations and full economic rationality; these assumptions are not too attractive for a behavioral economist.

Thinking about equilibrium

Equilibrium analysis is very useful to illustrate some of the core economic principles that lie behind demand and supply curves and it often makes sensible predictions. For instance, an enhancement of tastes towards a good raises its price in the short run (e.g., an increase in the price of avocados following its identification as a health superfood). Or an increase in taxes increases the marginal costs of producers and raises the price of a good. However, in the longer run, supply might catch up and prices could actually fall. All this can also be accommodated in equilibrium analysis. However, and this is a point that is often missed, the same might also be predicted by non-equilibrium models.

It is also often not realized that the notion of equilibrium can be adapted to dynamic systems in a constant state of flux, which is also a feature of complex systems. This, for instance, can be found in the work of Peyton Young and colleagues that I survey in Dhami (2016, Chapter 16 titled "Stochastic Social Dynamics). It gives rise to *punctuated equilibria* and other notions of temporal equilibria such as *stochastically stable states* that have been found useful in the study of social norms. It turns out that there is a close connection between Nash equilibria in neoclassical game theory and the social norms predicted by such models under the assumption that individuals engage in low level rationality, i.e., using simple rules of thumb, and simple adaptive rules of learning; for several examples see Dhami (2016, Chapter 16). Unfortunately, this work has had almost no impact in mainstream economics.

Drawbacks of the complexity approach in economics

In your question you point out that *a realistic understanding of human behavior is fundamental to the agenda in complexity based approaches.* I find this only partly true, at best. You are right in the sense that human behavior in ABMs is simple, adaptive, and heuristic. However, the insights from behavioral economics, widely replicated for the last 4 decades at least, have had very little impact and presence in this literature. For me, this literature models humans as having low-level rationality, as in Peyton Young's work, but humans are certainly not behavioral enough in these models. I will have more to say on

this below.

My critical assessment of ABMs below should be taken in a constructive spirit. Nothing benefits an emerging field more than sincere scrutiny from an outsider to the field. I am confident that the criticisms that I raise will be addressed in due course and only make the field stronger.

I find that the learning channels in ABMs do not appear to adequately recognize *behavioral frictions to learning*. In Dhami (2016, Part 5; Dhami, 2019, Vol. 6), I offer many examples: Individuals may be *overconfident* of their own abilities in making economic decisions, thereby reducing the scope for learning; could suffer from *attribution-bias* (blame the circumstances, not their decision rules); suffer from *confirmation-bias* (interpret existing evidence too favorably to support their initial beliefs/models); and suffer from *hindsight bias*, hence, underestimating the variance in their own estimates of volatility and not learning sufficiently. A closer interface between behavioral economics and complexity has the potential to enrich both.

One drawback of ABMs is that they have too many degrees of freedom in the choice of parameters and the initial conditions to fit the data; the postulated behavioral relations are sometimes no better than guesses; and given the complexity of the models, causal relations between variables may be hard to establish, although comparative static effects can easily be simulated over long time periods. The problem of extra degrees of freedom is, of course, not unique to ABMs. It also arises in economic theory when ad-hoc assumptions are used to explain a phenomenon and less than stringent tests are used to evaluate the relevant theory. It also arises in many other areas in social science (e.g., models of climate change) so one has to be mindful of applying the appropriate statistical tests to guard against the problem.

The errors in the specification of the parameters/initial conditions can propagate exponentially through an ABM model. There are two channels for this. First, there are a large number of equations and parameters in ABMs, giving rise to the complex error propagation paths that macroeconomists know well from the large scale macroeconomic models of the 1970's–80's. This, the inability to explain the empirical evidence (stagflation), as well as being subject to the Lucas critique, ultimately led to their demise relative to much more parsimonious models. Second, due to chaotic nonlinear dynamics, even minute changes in the values of the parameters used for simulation may lead to very large changes in the dynamic paths (think, for instance, of the logistic map). Yet, given the degrees of freedom in adjusting the parameter values and initial conditions in ABMs, such error propagation may go unnoticed.

Suppose that the empirical estimates of risk aversion used in an ABM come from an existing empirical study. Empirical estimates of risk aversion are based on an expected utility theory analysis. We now know that most of what we observe as risk aversion is

possibly loss aversion (Novemsky and Kahneman, 2005). However, this would be hard to uncover from an ABM model, which lacks tests of the validity of these individual components. Given the very large number of variables and transmission channels at work, it is difficult to establish cause and effect in ABMs. As such, the results from ABMs can be put on a stronger footing by incorporating insights from behavioral economics, which would have recommended using loss aversion in the first place. It is problematic, however, that by using estimates based on empirically refuted neoclassical models (expected utility in the case of estimates of risk aversion) ABMs still claim to match well the real world data. This will only raise the suspicions of those who believe that ABMs have too many degrees of freedom to make choices.

Which calibration values to use, when there is no consensus on the underlying values? Suppose that one needs to use labour supply elasticities for calibration purposes in ABMs. James Heckman, one of the leaders in this field, once reportedly said that several decades of research has shown that labor supply elasticities lie between $-\infty$ and $+\infty$. So which of the large number of available estimates should an ABM researcher use? Furthermore, a great deal of the data on labor supply elasticities comes from situations where the individual really does not have a choice to vary the hours worked (see Dhami, 2019, Section 3.8), while such choice might have been explicitly modelled in an ABM. Would an ABM researcher happen to be aware of all the nuances about the very large number of calibrated values in his/her model, that only specialists in the field might know about? This suggests the usefulness of teams of researchers with different expertise working on any ABM project.

The calibrated values for ABMs, such as behavioral parameters (risk aversion, marginal propensity to consume and invest, labour supply elasticities) are often taken from real world data. However, as one changes the initial conditions and policies, these behavioral parameters themselves will change. Hence, the well known *Lucas critique* that bedevils many economic models (changes in policies changes the behavioral responses of economic agents) also applies to ABMs. As such ABMs need to develop satisfactory methods of addressing this problem.

One mitigating factor in favour of ABMs is that there is a relatively small set of people working in this area and it is relatively young as a field. It is quite likely that, as the criticisms are taken into account, in due course, this field will continue to gain acceptance and grow within mainstream economics. And it must.

References

Acemoglu, D., Johnson, S., Robinson, J. A. and Yared, P. (2008). Income and Democracy. American Economic Review 98 (3): 808–42.

Akerlof, G.A., and Kranton, R.E. (2000). Economics and identity. Quarterly Journal of Economics. 115(3): 715–753.

Akerlof, G.A., and Kranton, R.E. (2005). Identity and the economics of organizations. Journal of Economic Perspectives. 19(1): 9–32.

al-Nowaihi, A. and Dhami, S. (2015) Evidential equilibria: Heuristics and biases in static games of complete information, in Games 6(4), 637-677.

al-Nowaihi, A. and Dhami, S. (2017) The Ellsberg paradox: A challenge to quantum decision theory? Journal of Mathematical Psychology. 78: 40-50.

Alesina A., and Giuliano, P. (2015) Culture and Institutions. Journal of Economic Literature 53(4): 898–944.

Alesina, Alberto, and Paola Giuliano. (2014). Family Ties. In Philippe Aghion and Steven N. Durlauf, (eds.) Handbook of Economic Growth, Volume 2A. Amsterdam and Boston: Elsevier, North-Holland. pp. 177–216.

Alesina, A., and Glaeser, E. L. (2004). Fighting Poverty in the US and Europe: A World of Difference. Oxford and New York: Oxford University Press.

Alesina, A., and Fuchs-Schündeln, N. (2007). Goodbye Lenin (or Not?): The Effect of Communism on People.American Economic Review 97 (4): 1507–28.

Ariely, D., Lowenstein, G., and Prelec, D. (2003). Coherent Arbitrariness: Stable Demand Curves without Stable Preferences. The Quarterly Journal of Economics, 118(1): 73-104.

Bartling, B., Fehr, E., and Schmidt, K.M. (2013). Use and abuse of authority: a behavioral foundation of the employment relation. Journal of the European Economic Association. 11(4): 711–742.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. American Economic Review 97(2): 170–6.

Battigalli, P., and Dufwenberg, M. (2009). Dynamic psychological games. Journal of Economic Theory. 144(1): 1-35.

Beinhocker, E. D. (2006). The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics. Harvard Business School Press: Boston Massachussets.

Beinhocker, E. D. (2016). The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics, Harvard: Boston MA.

Beinhocker, E. D., Arthur, W. B., Axtell, R., Bednar, Bouchaud, J. J-P, Colander, D., Crockett, M., Farmer, J. D., Hausmann, R., Hommes, C., Kirman, A., Page, S. and Wilson, D. S. (2019). Economics After Neoliberalism: Complexity Economics is Inclusive Economics, Boston Review, March 19

Bénabou, R., and Tirole, J. (2003). Intrinsic and extrinsic motivation. Review of Economic Studies. 70(3): 489–520.

Bénabou, R., and Tirole, J. (2006). Incentives and prosocial behaviour. American Economic Review. 96(5): 1652–1678.

Benjamin, D. J. (2018) Errors in Probabilistic Reasoning and Judgment Biases. USC Dornsife Institute for New Economic Thinking Working Paper No. 18-21.

Bertrand, M., and Schoar, A. (2006). The Role of Family in Family Firms. Journal of Economic Perspectives 20 (2): 73–96.

Bicchieri, C. (2006). The Grammar of society: The nature and dynamics of social norms. New York: Cambridge University Press.

Bowles, Samuel. (1998). Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions. Journal of Economic Literature 36 (1): 75–111.

Bowles, S., and Gintis, H. (2010). A Cooperative Species: Human Reciprocity and Its Evolution. Princeton and Oxford: Princeton University Press.

Brown, M., Falk, A., and Fehr, E. (2004). Relational contracts and the nature of market interactions. Econometrica. 72(3): 747-780.

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

Carpenter, J.P., and Seki, E. (2011). Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay. Economic Inquiry. 49(2): 612–630.

Cohn, A., Maréchal, M. A., Tannenbaum, D., and Zünd, C. L. (2019). Civic honesty around the globe. Science 365(6448): 70–73.

Dekel, E. and Lipman, B. L. (2010) How (Not) to do decision theory. Annual Review of Economics. 2: 257–82.

Dhami, S. (2016). The Foundations of Behavioral Economic Analysis. Oxford: Oxford University Press.

Dhami, S. (2017) Human Ethics and Virtues: Rethinking the Homo-Economicus Model. CESifo Working Paper No. 6836.

Dhami, S. (2019). The Foundations of Behavioral Economic Analysis. Volume I: Behavioral Decision Theory. Oxford: Oxford University Press.

Dhami, S. (2019). The Foundations of Behavioral Economic Analysis. Volume II: Other-Regarding Preferences. Oxford: Oxford University Press.

Dhami, S. (2019). The Foundations of Behavioral Economic Analysis. Volume III: Behavioral Time Discounting. Oxford: Oxford University Press.

Dhami, S. (2019). The Foundations of Behavioral Economic Analysis. Volume IV: Behavioral Game Theory. Oxford: Oxford University Press.

Dhami, S. (2019). The Foundations of Behavioral Economic Analysis. Volume V: Bounded Rationality. Forthcoming. Oxford: Oxford University Press.

Dhami, S. (2019). The Foundations of Behavioral Economic Analysis. Volume VI: Behavioral Models of Learning. Forthcoming. Oxford: Oxford University Press.

Dhami, S. and al-Nowahi, A. (2019) Behavioral economics and rationality. Forthcoming in Handbook of Rationality. MIT Press: Massachusetts.

Dhami, S. and al-Nowaihi, A. (2007) Why do people pay taxes: Expected utility versus prospect theory, Journal of Economic Behaviour and Organization 64: 171–192.

Dhami, S. and al-Nowaihi, A. (2010a) Redistributive policy with heterogenous social preferences of voters. European Economic Review 54(6): 743-759.

Dhami, S. and al-Nowaihi, A. (2010b) The existence of a Condorcet winner when voters have other regarding preferences. Journal of Public Economic Theory 12(5): 897-922.

Dhami, S. and al-Nowaihi, A. (2010c) Optimal income taxation in the presence of tax evasion: Expected utility versus prospect theory, with Ali al-Nowaihi. Journal of Economic Behavior and Organization, 75: 313-337.

Dhami, S., and al-Nowaihi, A. (2010d). A proposal to combine "prospect theory" and "cumulative prospect theory". University of Leicester. Discussion Papers in Economics 10/11.

Dhami, S., and al-Nowaihi, A. (2019). Behavioral economics and rationality. Forthcoming in Handbook of Rationality. MIT Press: Massachusetts.

Dhami, S., al-Nowaihi, A. and Sunstein, C. (2018) Heuristics and Public Policy: Decision Making Under Bounded Rationality. Discussion Paper No. 06/2018, Harvard Law School, John M. Olin Centre for Law, Economics and Business. Forthcoming in Studies in Microeconomics, special issue on Behavioral and Experimental Economics in Practice.

Dhami, S., Arshad, J., and al-Nowaihi, A. (2019). Psychological and Social Motivations in Microfinance Contracts: Theory and Evidence. CESifo Working Paper No. 7773.

Dhami, S., Manifold, E., and al-Nowaihi, A. (2019) Identity and redistribution: Theory and evidence. Mimeo University of Leicester.

Dhami, S., Wei, M., and al-Nowaihi, A. (2019) Public goods games and psychological utility: Theory and evidence. Forthcoming in the special issue on Psychological Game Theory in Journal of Economic Behavior and Organization.

Ellingsen, T., and M. Johannesson (2004): Promises, Threats and Fairness, The Economic Journal, 114, 397–420.

Elster, J. (2011). Norms. in Peter Bearman and Peter Hedström (Eds.). The Oxford Handbook of Analytical Sociology. Oxford University Press: Oxford.

Englmaier, F., and Leider, S. (2012). Contractual and organizational structure with reciprocal agents. American Economic Journal: Microeconomics. 4(2): 146–183.

Ertan, A., Page, T., and Putterman, L. 2009. Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free Rider Problem. European Economic Review 53 (5): 495–511.

Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. Games and Economic Behavior. 54(2): 293-315.

Farmer, J. D., Mealy, P. and Beinhocker, E. D. (2019). A New Generation of Models for Public Policy. Mimeo INET Oxford.

Fehr, E. and Leibbrandt, A. (2011). A field study on cooperativeness and impatience in the Tragedy of the Commons. Journal of Public Economics 95(9–10): 1144–55.

Fehr, E., and Schmidt K.M. (1999). A theory of fairness, competition and cooperation. Quarterly Journal of Economics. 114(3): 817–868.

Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. Nature Human Behavior 2: 458-468.

Fehr, E., Klein, A., and Schmidt, K.M. (2007). Fairness and contract design. Econometrica. 75(1): 121-154.

Fershtman, C., and Gneezy, U. (2001). Discrimination in a Segmented Society: An Experimental Approach.Quarterly Journal of Economics 116 (1): 351–77.

Feynman, R. (1965) The character of physical law. The MIT Press: Massachusetts.

Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies in disguise: an experimental study on cheating. Journal of the European Economic Association. 11(3): 525–547.

Fischer, David Hackett. (1989). Albion's Seed: Four British Folkways in America. Oxford and New York: Oxford University Press.

Friedman, M. (1953). The Methodology of Positive Economics, Chicago: University of Chicago Press.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. Games and Economic Behavior. 1(1): 60-79.

Gilboa, I., Postlewaite, A., Samuelson, L., and Schmeidler, D. (2014) Economic Models as Analogies, Economic Journal, 124, F513-F533.

Gintis, H. (2009). The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences. Princeton: Princeton University Press.

Gintis, H. (2017). Individuality and entanglement: The moral and material bases of social life. Princeton: Princeton University Press.

Gorodnichenko, Y., and Roland, G. (2018). Culture, institutions and democratization. mimeo.

Greif, A. (1994). Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. Journal of Political Economy 102 (5): 912–50.

Greif, A., and Tabellini, G. (2012). The Clan and the City: Sustaining Cooperation in China and Europe. Innocenzo Gasparini Institute for Economic Research Working Paper 445.

Gruber, J., and Hungerman, D. M. (2008). The Church versus the Mall: What Happens When Religion Faces Increased Secular Competition? Quarterly Journal of Economics 123 (2): 831–62.

Guiso, L., Sapienza, P. and Zingales, L. (2004). The Role of Social Capital in Financial Development. American Economic Review 94 (3): 526–56.

Guiso, L., Sapienza, P. and Zingales, L. (2008a). Social Capital as Good Culture.Journal of the European Economic Association 6 (2–3): 295–320.

Guiso, L., Sapienza, P. and Zingales, L. (2008b). Trusting the Stock Market. Journal of Finance 63 (6): 2557–2600.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., et al. (2001). Cooperation, reciprocity and punishment in fifteen small-scale societies. American Economic Review. 91: 73–78.

Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. Science 319(5868): 1362–67.

Itoh, H. (2004). Moral hazard and other-regarding preferences. Japanese Economic Review. 55(1): 18–45.

Kahneman, D. (2011). Thinking fast and slow. New York, NY: Farrar, Strauss, Giroux.

Kahneman, D., and Tversky, A. (1972) Subjective probability: A judgement of representativeness. Cognitive Psychology 3: 430-454.

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. Econometrica. 47(2): 263–291.

Lakatos, I. (1970) Falsification and the methodology of scientific research programmes. in I. Lakatos and A. Musgrave (eds),Criticism and the Growth of Knowledge. Cambridge University Press: Cambridge.

Luce, R.D., and Raiffa, H. (1957). Games and Decisions. John Wiley and Sons, Inc.

Luttmer, E. F. P., and Singhal, M. (2011). Culture, Context, and the Taste for Redistribution.American Economic Journal: Economic Policy 3 (1): 157–79.

North, Douglass C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge and New York: Cambridge University Press.

Novemsky, N., and Kahneman, D. (2005). The boundaries of loss aversion. Journal of Marketing Research. 42(2): 119-128.

Nunn, Nathan, and Leonard Wantchekon. (2011). The Slave Trade and the Origins of Mistrust in Africa. American Economic Review 101 (7): 3221-52.

Popper, K. (1934). Logik der Forschung (Hutchinson & Company published the translation by Karl Popper titled The Logic of Scientific Discovery in 1959).

Popper, K. (1963): Conjectures and refutations: The growth of scientific knowledge. London: Routledge and Kegan Paul.

Rabin, M. (1993). Incorporating fairness into game theory and economics. American Economic Review 83(5): 1281–302.

Rabin. M. (2000) Risk aversion and expected utility theory: a calibration theorem. Econometrica 68(5): 1281-92.

Roemer, J. (2015). Kantian optimization: A microfoundation for cooperation. Journal of Public Economics 127: 45–57.

Roemer, J. (2018). A theory of cooperation in games with an application to market socialism. Mimeo.

Rubinstein, A. (2006). Dilemmas of an economic theorist. Econometrica, 74(4): 865-83.

Sutter, M., Haigner, S. and Kocher, M.G. (2010). Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations. Review of Economic Studies 77 (4): 1540–66.

Tabellini, Guido. (2008). The Scope of Cooperation: Values and Incentives.Quarterly Journal of Economics 123 (3): 905–50.

Thaler, R.H. (2015) Misbehaving: The making of behavioral economics. W. W. Norton and Company: New York.

Tomasello, M. (2014). A natural history of human thinking. Harvard: Harvard University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5 (2) 207-232.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. Science 185, 1124-1130.

Tyran, J-R., and Feld, L. P. (2006). Achieving Compliance When Legal Sanctions Are Non-deterrent.Scandinavian Journal of Economics 108 (1): 135–56.

Voors, Maarten J., Eleonora E. M. Nillesen, Philip Verwimp, Erwin H. Buite, Robert Lensink, and Dean P. Van Soest. (2012). Violence Conflict and Behavior: A Field Experiment in Burundi.American Economic Review 102 (2): 941–64.

Wei, M., al-Nowaihi, A., and Dhami, S. (2019) Quantum decision theory, bounded rationality and the Ellsberg paradox. Forthcoming in Studies in Microeconomics.

Whitt, S., and Wilson, R. K. (2007). The Dictator Game, Fairness and Ethnicity in Postwar Bosnia. American Journal of Political Science 51 (3): 655–68.

Young, P. (1998) Individual strategy and social structure: An evolutionary theory of institutions. Princeton: Princeton University Press.