

Mechanism in behavioural economics

Michael Joffe
Imperial College London

Department of Epidemiology & Biostatistics
St Mary's Campus
Norfolk Place
London W2 1PG

m.joffe@imperial.ac.uk

+44 20 7609 1470

Orcid: 0000-0001-6907-6183

Abstract

Behavioural economics promises to bring economics closer to being evidence based. However, its ability to do this depends on aligning its account of mechanism with that operating in the real world. This in turn may depend on a methodological issue: whether the findings of behavioural economics are used to modify or extend standard theory, or to contribute towards replacing it where required – respectively the *incremental* and *selective replacement* strategies. I focus on the incremental approach, in terms of its implied causal mechanism. Two stages are involved, corresponding to the prediction of standard theory and to a separate component that aligns it with actual observations. In behavioural economics, one possible interpretation of the language of ‘biases’ is such a two-stage approach. More explicitly, Rabin advocates it in the form of PEEMs (Portable Extensions of Existing Models). A more direct, one-stage approach may have some advantages, at least for some research topics.

Key words

Behavioural economics
Biases
Economic methodology
Causal mechanism
Evidence-based economics
PEEMs (Portable Extensions of Existing Models)

Jel classification

B41; D91; G41

‘... it is time to fully embrace what I would call evidence-based economics. This should not be a hard sell. Economists use the most sophisticated statistical techniques of any social science, have access to increasingly large and rich datasets, and have embraced numerous new methods from experiments (both lab and field) to brain imaging to machine learning. Furthermore, economics has become an increasingly empirical discipline. ... behavioral economics is simply one part of the growing importance of empirical work in economics ...’

Richard Thaler (2016)

1. Evidence-based economics and causal mechanisms

1.1. *The idea of evidence-based economics*

It is undoubtedly true that evidence plays a major role in modern economics. Statistical evidence is particularly influential; the variety and scope of datasets is rapidly increasing, and there has been a great improvement in the methods of statistical analysis and causal inference over recent decades. Much of this is on important practical issues, such as the influence of class size on educational attainment, and of educational investment on future earnings. Other types of evidence can also make important contributions, e.g. comparative economic history, as used for example in institutional economics (Acemoglu & Robinson, 2012). In principle, these methods can provide evidence not only of a descriptive type (*what* happens) and of a causal type (in the sense of what factors *bring about* the economic phenomena we see), but also mechanistic evidence – *how* these causal processes work.

However, for ‘economics’ to be evidence-based, more would be needed: the basic account of how the economy works – economic theory – would need to be *founded* on evidence (Joffe, 2013a; 2014). To what extent is this happening? Evidence-based economics in this sense has a rival: ‘standard theory’ that is constructed on the basis of axioms and assumptions, not derived from evidence. In practice, a great deal of empirical work is devoted to testing a hypothesis that is tightly specified by a traditional model, rather than being generated from evidence about the real world, or even flexible enough to allow the data to speak freely (Juselius, 2011). If economics were evidence-based, there would be a stronger link *from evidence into theory* – our account of how the economy works would either be directly derived from evidence of diverse types by means of generalization and explanation, or involve hypotheses generated after systematic consideration of the available evidence of all types (Joffe, 2013a; 2014). This would more closely align theory with the real world, supplementing and/or replacing those elements of traditional theory that conflict with the evidence.

When empirical observations do not coincide with traditional theory, their relationship can take any of three forms. The empirical findings can be used to criticise standard theory – what Rabin (2013a) calls ‘flaw-finders’ – which could be described as *reactive mode*. Secondly, one could regard the findings of behavioural economists as evidence that traditional rational choice theory needs to be modified or extended – what may be called *incremental mode*. Thirdly, one could take these findings as evidence that rational choice theory is wrong, at least in certain respects, and therefore needs to be replaced – *selective replacement mode*. Leaving aside reactive mode, which on its own does not generate any new

perspective, we are left with incremental and selective replacement modes. Both are represented in behavioural economics. It may well be that incremental mode is appropriate in some cases, and selective replacement mode in others.

Behavioural economics has now become an important part of mainstream economics. It is tempting to believe that its rise will ensure that economic theory will become ever closer to actual reality. I take this to be identical to the call for evidence-based economics. Thaler may be right in saying that ‘this should not be a hard sell’, but there is an important yet relatively unexplored methodological question to be addressed in relation to the direction that should be taken. This relates to the mechanisms that bring about the phenomena in the actual economy: whether or not our theories provide accurate mechanistic descriptions. The aim of this paper is to explore the implications of incremental mode in terms of its mechanistic adequacy. Detailed discussion of the second strategy is left for future research.

1.2. The methodology of behavioural economics

Behavioural economics involves at least four methodological issues:

1. *The appropriate source(s) of knowledge.* An important tradition in economic theorizing has been an emphasis on *a priori* principles, derived from introspection or intuition, intended to provide a logical (rather than psychological) foundation. An alternative is the view proposed by Thaler in the above quotation: knowledge should be derived from evidence. There is no *necessary* contradiction between these, as it is possible to support a role for both. However, behavioural economics has its roots in the apparent conflict faced in the late 1960s by Amos Tversky between his experimental findings and the theories of von Neumann, Morgenstern and Savage that continued the traditional emphasis on strict rationality (Heukelom, 2014, p. 132).

His solution to this problem was to assign a specific role to each. Formal theories of rational decision making were still considered to provide the ‘correct’ answer. At the same time, the empirical research showed that in practice, ‘people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations’ (Tversky & Kahneman, 1974, p. 1124).

2. *The role of normativity.* The implication was that rationality retained its normative role as the criterion for economic behaviour, while allowing a descriptive role for the new findings. This made it possible for the behavioural economists to maintain a connection with formal economic theories of rational decision making, as well as the closely-associated measurement theory, thereby becoming potentially influential among economists (Angner & Loewenstein, 2012). This in turn facilitated Thaler in publicizing the new behavioural economics among economists in his regular column in the *Journal of Economic Perspectives* (e.g. Thaler, 1987).

The emphasis on psychology as error implied that the task was to identify the conditions under which these mechanisms lead to systematic mistakes (Angner & Loewenstein, 2012; Heukelom, 2014, p. 132). Consequently, economic outcomes should be inferior, or at least not superior, to those that result from the type of rational deliberation that features in the theories of von Neumann, Morgenstern and Savage.

3. *The reducibility of economic phenomena to behaviour.* An implicit assumption is often made that success in explaining economic behaviour would automatically result in a good account of the economy and how it works. But a reduction of the economy to the individual level is not necessarily valid. Here I outline how this issue has been addressed in the previous literature, focusing primarily on experimental economics and on economic methodology. The relationship between economic behaviour and economic causation more broadly is discussed in more detail below, in section 3.2.

In parallel with the burgeoning behavioural economics research programme, experimental economists (notably Vernon Smith) were proposing a different interpretation of rather similar empirical results. Smith distinguished between individuals' short-term behaviour, which is often far from the rational ideal, and the long-term result, brought about by market forces – 'the incentive properties of markets'. 'Initial choices may reflect all manner of beliefs and expectations, but if these choices are not sustainable in a market clearing or a noncooperative equilibrium, subjects adapt their expectations and behavior until they attain such an equilibrium' (Smith, 1989, p. 166). This view emphasizes 'the role of time in driving the market to higher degrees of rationality' (Heukelom, 2014, p. 134). It could be seen as a contrast between the *ex ante* behaviour of individuals, and the *ex post* consequence of the market mechanism.

This perspective is related to the view of Hausman (1992, Chapter 11), who recommended that supply and demand explanations should be seen as the result of causal forces acting through time, rather than as a static equilibrium. Evolutionary economics similarly emphasizes consequences, e.g. differential survival, interacting with behaviour. This contrasts with the view of Friedman (1953), for whom it was immaterial whether outcomes resulted from strictly rational behaviour or from the later consequences of decisions – for example, firms' behaviour might end up resembling the textbook account either because they made calculations that approximated the equations of standard theory, or because those that did not do so tended to falter and disappear.

Instead of the normative/descriptive division of behavioural economics, therefore, the perspective of experimental economics emphasized that rationality still provided an empirically accurate account of the economy, but this was located in the market mechanism rather than in the behaviour of individuals. The *interaction* of sellers and buyers (supply and demand) was considered to bring about a rational outcome, despite the possible irrationality of the individual participants. The implication is that the operation of market forces is sufficient to explain the observed stability of e.g. product markets, making redundant the assumption of rationality in the extreme sense of von Neumann, Morgenstern and Savage. Thus, the behaviour of the economy is not necessarily reducible to a simple summation of the economic behaviour of all the individuals in the population (see also Herfeld, 2018).

4. *Accurate characterization of the causal mechanism.* A fourth methodological issue has perhaps received less attention than the other three: the extent to which a theory represents the actual causal processes involved in economic behaviour. A causal process involves both a mechanism and the phenomenon that it produces (Russo & Williamson, 2007; Joffe, 2013b). One implication is that for any candidate theory to be considered realistic, (a) it must involve the existence of *some* mechanism; and (b) evidence should be available indicating that it is likely to be the *correct* mechanism. In the early stages, (a) may be present without (b), when

it may be considered to be a ‘how-possibly’ rather than a ‘how-actually’ explanation (Craver & Darden, 2013).

Furthermore, the theory should provide a satisfactory explanation of the features of the phenomenon itself. As Kahneman (1996) and others have emphasized, causal mechanisms need to be able to account for behaviour that conforms with strictly rational criteria, as well as behaviour that might be judged highly irrational. A causal explanation is required for all phenomena, not only those deemed “psychological” because they fail the rationality test. In addition, the mechanism ideally needs to explain the quantitative features of the phenomenon, including whether it is relatively invariant, or whether there is substantial variation according to context.

1.3. Two-stage and direct approaches to causation

It is clear that some behavioural economists see their task in terms of uncovering the real causal mechanisms that underlie the phenomena that they study. For example, Camerer and Loewenstein (2004, p. 4) have stated: ‘the ultimate test of a theory is the accuracy with which it identifies the actual causes of behavior’. However, this is not necessarily true of the discipline as a whole. In particular, there is a *prima facie* case that the language of ‘biases’ indicates, at least sometimes, a combination of behaviour deemed to conform with traditional standard theory, together with a second implicit process that is required to explain observed behaviour; it is an open question whether both of these correspond to actually-existing causal mechanisms, as this is not routinely discussed in the behavioural economics literature.

This may be called a ‘two-stage’ account – although it is not essential that the two stages in fact occur sequentially – they could for example act simultaneously. The second process would correspond not to the phenomenon itself, but to the *deviation* of behavioural evidence from traditional theory. This has been termed the ‘Bernoulli repair program’, indicating Bernoulli’s attempt to rescue expected-value maximization as a normative criterion after the St Petersburg Paradox was discovered (Berg & Gigerenzer, 2010). Such a two-stage account could be a good representation of the real world if both processes are mechanistically accurate. However, in some cases the first stage, the rational component, might be conceived of in ‘as if’ terms, i.e. it might not correspond with the actually-occurring mechanism. If the first stage does not correspond to any real-world process, then the second stage representing the deviation would not describe any real causal process either. From a causal viewpoint, the result would be a *double error*.

In contrast to two-stage thinking, the direct approach proposes that the phenomenon itself, not the deviation, is brought about by the action of a mechanism – always bearing in mind that causation in the human realm is typically complicated and may in fact involve multiple mechanisms and their inter-relation. It implies that *there is no necessity to start from standard theory*. Instead, the description of each phenomenon provides the starting point for research on the causal explanation (see the examples in section 3.1). In section 2, I explore the two-stage approach to behavioural economics. I focus especially on the work of Rabin, who has set out this perspective with great clarity. Then in section 3, I consider the implications of this issue, for (a) the possible advantages of a direct or one-stage approach, (b) how behavioural economics relates to explaining the behaviour of the economy, and (c) whether other sub-disciplines of economics are affected.

2. Behavioural economics and the two-stage approach

2.1. The ‘benchmark and biases’ perspective in behavioural economics

One influential viewpoint is that the aim of behavioural economics is to uncover the biases that lead people’s behaviour to diverge from the traditional assumptions of rationality, perfect information, perfect foresight, etc – i.e. *Homo economicus*. As Thaler (2015, 7) has written ‘We need an enriched approach to doing economic research, one that acknowledges the existence and relevance of Humans. The good news is that we do not need to throw away everything that we know about how economies and markets work. Theories based on the assumption that everyone is an Econ [*Homo economicus*] should not be discarded. They remain useful as starting points for more realistic models.’ This may be called the ‘benchmark and biases’ approach, and its relationship to standard economic theory is incremental: its aim is to take standard theory and to add an adjustment factor.

The language of ‘biases’ implies that there are two distinct stages. However, not all of Thaler’s work follows this two-stage schema. One of the first examples in his *Misbehaving* is that people routinely buy a particularly large portion on a Sunday for dinner on Tuesday if they are hungry when shopping; and then on Tuesday evening they will finish that huge meal because they have already paid for it (Thaler, 2015). This could be regarded as a step towards a direct rather than a two-stage account: description of behaviour that is in need of a causal explanation – even though Thaler chooses to phrase it as something an Econ would not do. But it is a trivial task to recover the observation from the way it is described.

It is understandable that the findings of behavioural economics should historically have been presented in a negative way, to show that some received ideas about economic behaviour are wrong. The old traditions are still powerful, and by emphasising what is wrong with them, behavioural economics has arguably been able to attract more attention than would otherwise have been possible with a more direct, positive message. Reactive mode served an important purpose.

In addition, the focus on ‘biases’ has provided a tractable method of studying how human economic behaviour really operates, just as optical illusions provided powerful insights into visual perception. The two-stage approach – incremental mode – has been fruitful, even when it has neglected mechanistic accuracy. But when the postulated stages do not correspond to real causal mechanisms, this approach has serious limitations. The trend towards evidence-based economics implies that it will only be possible to shift the thinking of the majority of economists in the long term if there is a positive alternative. Reactive and incremental modes will need to give way to selective replacement mode.

2.2. Portable Extensions of Existing Models

In contrast to the selective replacement possibility, some authors advocate retaining as much of traditional economics as possible, i.e. incremental mode. In this paper, I focus on the viewpoint of Rabin (2013a). He advocates the development of what he terms ‘PEEMs’: Portable Extensions of Existing Models. The idea is to improve mainstream economics, specifically its psychological realism, while ‘maintaining its conventional techniques and goals: formal theoretical and empirical analysis using tractable models, with a focus on

prediction and estimation'. However, he is clear that this approach cannot encompass all of behavioural economics; as he says, 'most great research improving psychological realism is outside' this framework.

The justification for starting with existing models is that they 'are rarely devoid of truth and insight'. An additional component is added to the traditional model, to better match reality, while preserving many of the original theory's predictions. This is done 'by formulating a modification that embeds it as parameter values with the new psychological assumptions as alternative parameter values'.

The issue for the present discussion is whether or not the traditional starting point provides a psychologically accurate mechanism. Rabin's primary concern is not with mechanistic accuracy, but rather with quantitative adjustments that remedy the poor performance of some traditional models. The same is true of a longer paper that focuses on bounded rationality (Rabin, 2013b).

In Rabin's schema, the new parameter may take a value such as 0 or 1 in the original neoclassical model, taken as a baseline. An augmented model is then developed in which the parameter is free to take a different value, and the task of empirical research is to estimate its true value. Thus, it is not a question of adding a new assumption, rather it is 'unburying' an existing assumption in the original model that was hitherto invisible. The approach is made portable by using the same independent variables as in existing research, or proposing measurable new variables. It is explicitly in incremental mode, and framed in terms of bias, that is, of deviation from the standard model: 'If measured present bias varies between 0 and 1, then always assuming one particular $\beta < 1$ is an imperfect model – but always assuming $\beta = 1$ is a worse one'. This method has the advantage that one can easily measure the increase in explanatory power that is attributable to the inclusion of the additional variable.

In some ways, this methodology represents an important departure from aspects of the traditional approach to theory. Unlike Vernon Smith's view, it denies that the psychological factors identified in behavioural economics will simply be obliterated by the market mechanism: 'PEEMs can replace a certain type of market mysticism that markets will wipe out "behavioral" phenomena with the same sort of serious market analysis that predominates economics based on more familiar assumptions'.

The approach also challenges the view that regards the unique solution generated by standard theory as scientifically necessary: 'PEEMs can also help to put aside habitual dismissiveness that "anything can happen" once one modifies existing assumptions'. However, the focus is still on 'prediction and estimation', with the underlying expectation that the economy will obey the type of laws (equations) that featured in nineteenth-century physics, and which have been imitated in mathematical economics. Although a deep-rooted feature of modern economics, this belief is open to challenge by an alternative view that emphasises causation, and particularly that the economy is characterised by multiple causal factors that inter-relate in complicated ways, producing an open-ended and therefore inherently unpredictable reality – Knightian uncertainty. In this alternative view, systematic correspondence with the real-life forces that operate in the economy is emphasised, and it is acknowledged that precise predictions of the future are not necessarily possible – causality and realisticness are

prioritised over prediction. Mathematical and statistical modelling are then embedded in a broad, causal account (Joffe, 2017a).¹

Within the context of statistical modelling, Rabin (2013a) makes some important points, which would still be valid in this ‘embedded’ context. Two important properties of a theory are identified. One is ‘power’, the ability to ‘rule enough things out to make it valuable to science’. The other is ‘scope’: ‘how broad is the set of situations to which it applies?’ As he says, ‘theories have most explanatory power when they are general in their applicability across contexts but *specific* in their predictions within each context’.

He also distinguishes PEEMs from PHEEMs (*post hoc* extensions of existing models), which appear in many reports based on specific datasets. He advocates better reporting before settling on a particular framework to extend an existing model, because PHEEMs are not subject to the a priori restrictions of PEEMs. This would be analogous to the avoidance of ‘specification mining’ in the context of statistical analysis.

In addition, he cautions against the assumption that the value of the new behavioural parameter is constant. ‘When parameter estimates are highly variable across contexts, it is important to make this salient so that researchers know further improvements are needed.’ This is required to try and ensure that PEEMs work across settings, not just in a particular domain or dataset.

Finally, Rabin emphasises the need to consider a theory in relation to other theories. The traditional way this has been recognised is the failure to consider alternative explanations, when one’s favoured hypothesis performs well (sufficiency bias). Traditional economists have been tempted to try and explain the findings of behavioural economics in the terms of standard theory: ‘Examples where what looked like concern for fairness could also be consistent with pure self-interest were greeted with great excitement. But since it was clearly possible that what looked like a concern for fairness was in fact a concern for fairness, greater enthusiasm for instead seeking out examples where the two *were* distinguished would have been a healthier response.’ To this, he adds ‘insufficiency bias’, which is the mirror image: a theory should not be rejected because data are inconsistent with it, when no alternative theory is able to explain those data. Rather, the question should be whether the lack of fit with the data results from the *distinguishing* features of the theory.

2.3. Examples of PEEMs

Rabin gives several examples of PEEMs. Perhaps the simplest is the type of model that modifies traditional expected utility theory to allow for non-linear probability preferences. They embed the same risk-free cardinal preferences as expected utility, differing only in the way that the probabilities enter the risk preferences.

In the context of behavioural game theory, various modifications to the theory have been made, to improve its correspondence with data from experimental games. This is because the

¹ These features make economics similar to the scientific study of other open-ended, unpredictable systems, such as biology and geology. The future course of biological evolution is rarely predictable, and it is impossible to predict the timing (as contrasted with the location) of earthquakes or volcanic eruptions.

traditional theory has long been known to perform poorly, even with simple games. For example, McKelvey and Palfrey (1996) extend the standard Nash equilibrium model by introducing the idea that there are many different possible ‘types’ of players. This has two implications. First, it explains the statistical variation in behaviour that is always observed in experiments, even if all agents have private information about their type and always adopt pure strategies. Secondly, and more importantly, players will respond to anticipated errors of opponents in ways that can lead to systematic deviations from Nash equilibrium.

In this model, players choose probabilistically, with better alternatives being chosen more often than worse ones. The Nash equilibrium model is a special case of this model. The new model can therefore be seen as a generalization of the traditional model. Each player has a set of strategies; the expected utility of each of these is determined by the probability distribution of the strategies of the other players. The resulting ‘Quantal Response Function’ thus corresponds to the Best Response Correspondence in game theory, and is a generalization of it. It can provide a unique equilibrium solution, that differs from the Nash equilibrium.

This is achieved by introducing a free parameter, λ , which can take any non-negative value, and can be thought of as measuring how careful a player’s decision is, or alternatively, how experienced the player is. When $\lambda = 0$, behaviour is essentially random, and as λ approaches infinity, the model of behaviour approaches the best response model. On the face of it, this is a causally plausible account. The human diversity is empirically realistic, and probabilistic modelling is appropriate for processes that can be described as stochastically causal. However, a closer examination exposes some issues. The two ‘stages’ here are (i) traditional rationality, and (ii) a randomizing factor that makes the outcome fuzzy. The question is, how plausible are these *from a mechanistic viewpoint*? Do they have any psychological meaning? No argument is presented, either by McKelvey and Palfrey or by Rabin, to justify these as actually-existing causal processes.

In addition, this theory does not provide a stable account of the quantitative features of the phenomenon: the theory is unclear what determines λ , and estimates of λ have been found to vary considerably from game to game – variation that is unexplained. Essentially, it is a fudge factor that assesses the departure of behaviour from the assumed rationality – the discrepancy from a model that is assumed to be structurally correct even if empirically poor. It is explicitly an alternative to abandoning Nash equilibrium entirely, and using psychological theories of behaviour instead – and in the paper, this possible alternative is phrased in terms not of explaining the *actual behaviour*, but of explaining ‘*systematic deviations* from Nash equilibrium’.

An alternative model has been developed by Eyster and Rabin (2005), who suggest what they call a ‘cursed equilibrium’. Each player correctly predicts the distribution of other players’ actions, but they fail to understand how these actions are related to these other players’ information. This can lead to engaging in a trade that would not occur if they correctly assessed others’ information, as in Akerlof’s study of the used car market, where sellers’ willingness to sell should have been taken as indicating that the car is a ‘lemon’, and therefore not worth buying. Another example is the winner’s curse in the context of auctions, where the winner should have realised that other possible bidders may have stopped bidding because they knew the item was not as valuable as the winner had thought. The underlying idea is ‘a single psychological principle – the underappreciation of the informational content of other people’s behavior’.

The parameter that describes the degree to which a player under-appreciates the private information of other players is the probability χ , where $\chi = 0$ corresponds to the fully rational Bayesian Nash equilibrium and $\chi = 1$ corresponds to the case where each player assumes no connection whatever between other players' actions and their types. Any value of χ in the range (0, 0.6) provides a better fit than the Bayesian Nash equilibrium for the experiments analysed in Eyster and Rabin (2005), but in other contexts it makes unrealistic predictions, or is not well defined, or lacks robustness. The authors see their method as complementing, not replacing, other approaches such as quantal response functions.

In this case, defective understanding of others' motives does appear to meet the requirement of psychological plausibility; it could be regarded as an aspect of bounded rationality. But again, the model does not perform well empirically.

The incremental approach lends itself to iterative application. For example, in a laboratory experiment on first- and second-price independent-value auctions, some findings were incompatible both with Bayes-Nash equilibrium and cursed equilibrium, leading the authors to construct a model combining cursedness with an underweighting of the opportunity costs of higher bids, together with substantial bidder heterogeneity (Turocy & Cason, 2015). Thus, there is a danger, as Rabin (2013a) warned, of *post hoc*, situation-specific model specification, the equivalent of over-fitting in statistical analysis.

The case of present bias is 'surely the most successful PEEM to date' according to Rabin (2013a). Following Strotz (1956), Laibson (1997) describes the wide variety of highly-prevalent ways that people choose to commit themselves, thus protecting them from future temptation. He points out that this phenomenon had previously received little attention from economists, which he attributes to the fact that commitment is only necessary for consumers whose preferences are dynamically inconsistent, which might have been problematic for most economists. However, he points out that there is 'a substantial body of evidence' for dynamic inconsistency: 'Research on animal and human behavior has led psychologists to conclude that discount functions are approximately hyperbolic (Ainslie, 1992)'.

This is in sharp contrast to the standard exponential discount function, that is associated with rational, dynamically consistent, behaviour. With dynamic inconsistency, preferences at date t are inconsistent with preferences at date $t+1$. The marginal rate of substitution between periods $t+1$ and $t+2$ is different for the decision maker at time $t+1$ than at time t . Laibson develops a model based on the idea that an individual is a sequence of temporal 'selves' making choices in a dynamic game, and looks for subgame perfect strategies in that game. The model is *quasi*-hyperbolic rather than truly hyperbolic, because it is set up to retain much of the analytical tractability of exponential discounting, while capturing the key qualitative feature of discounting with true hyperbolas.

The implied first stage is, as before, the standard strictly rational account, and the second stage is now not a separate parameter, but rather the coexistence of multiple selves – the assumed first stage is multiplied rather than supplemented. From a causal viewpoint, the issue is the psychological plausibility of the idea of multiple selves with differing time horizons. This is not the only hypothesis that postulates that the mind or brain is segmented. Fodor (1983) proposed that the mind, apart from its higher-level cognitive processes, is modular. Cosmides and Tooby (1994a) have argued that the mind is composed of domain-specific

computational modules that evolved to solve specific problems in the evolutionary past.² The idea of selves with different timescales could also be compared with the different types of memory that have been described, and with Kahneman's 'fast' and 'slow' systems. It would be possible to argue that Laibson's model corresponds to some compartmentalization of this kind, although to my knowledge this has not been attempted.

Another perspective could be obtained by following Rabin's emphasis on the importance of considering hypotheses in relation to other theories. There may be more causally plausible ideas, that make tomorrow more different from today than is true of a-year-and-day compared with a-year-from-today.

The question remains as to how well the Laibson model corresponds to the mechanism that brings about the observed pattern of discounting.³ Many models have sought to reproduce a hyperbolic type of discounting that lead to 'present-biased preferences' (O'Donoghue and Rabin, 1999, especially footnote 2), and some may do this better than others. More evidence is required to address this issue.

In all these examples, their presentation emphasizes the mathematical aspects of the theory at the expense of discussion of mechanistic plausibility. From a causal viewpoint, they are *ad hoc*. In particular, the routine reliance on traditional strict rationality as the first of the two stages implies the need to supply evidence supporting the occurrence of the assumed rational calculation as an actual psychological mechanism.

A further consideration is the question of the demands being made of the agent. The addition of extra parameters to the already complicated calculation that is assumed to be taking place implies mental capacities that are likely to be super-human. In the empirical literature, these theories tend to be judged not by genuine out-of-sample prediction, but only by goodness-of-fit measures such as R-squared, or the Akaike or Schwarz Information Criteria (Berg & Gigerenzer, 2010).

3. Some implications

The analysis so far has suggested that from a mechanistic viewpoint, there may be problems with the conventional two-stage approach to behavioural economics. In this section, I consider the implications (i) for behavioural economics: whether there is an alternative to starting from standard theory; (ii) for economics as a whole: to what extent a satisfactory account of the economic behaviour of individuals contributes to the achievement of a satisfactory account of the behaviour of the economy; and (iii) for economic methodology: whether the two-stage issue is specific to behavioural economics, or whether it is more widespread.

3.1. The direct or one-stage approach

² However, the idea of modularity has been criticized (e.g. Panksepp & Panksepp, 2000; Uttal, 2003; Buller, 2005).

³ In addition, some studies have sometimes cast doubt on the interpretation that present bias is best represented in hyperbolic (or *quasi*-hyperbolic) form. See e.g. Rubinstein (2003) and Andersen et al. (2014).

There are many possible ways to develop and revise economic theory, i.e. our account of the phenomena that occur in the economy and the causal processes that bring them about. It has become normal practice in economics to take traditional standard theory as the starting point. As economists, we often assume that this is natural, and maybe there is discomfort when that method is not used.

However, there is no compelling methodological reason why this should be the only way to proceed. The starting point in the natural sciences is often a theory based on evidence, which can be expressed in mathematical, diagrammatic and/or verbal/causal terms. It clearly needs to be internally consistent, but does not have to be derived from logical/mathematical criteria based on introspection. The alternative is what I have called the direct approach, in which no *a priori* decision is made concerning rationality or optimization, and theory is generated from evidence of multiple different kinds (Joffe, 2013a, 2014), as is standard practice in natural sciences such as biology and geology (Joffe, 2017a). This could be a promising methodology, if we wish to move towards evidence-based economics (Joffe, 2014, 2017a).

There are also empirical grounds for questioning the two-stage approach. If strict rationality, corresponding to the first stage, is the normative criterion, then it should generate a superior outcome when operating on its own, as compared with routine two-stage behaviour – the psychological biases should lead to ‘economically significant losses’ in terms of earnings, happiness, health, beliefs and/or life expectancy (Berg & Gigerenzer, 2010). This cannot be a once-for-all judgement covering all behavioural phenomena – it needs to be considered on a case-by-case basis. The implication is that research on each topic in behavioural economics should routinely present evidence demonstrating the loss in welfare that results from the second stage. This is not standard practice.

In fact, where out-of-sample testing has been carried out, it does not necessarily confirm the superiority of traditional strict rationality (e.g. DeMiguel et al, 2009), and sometimes indicates that expected utility violators and time-inconsistent decision makers perform better (Berg et al, 2010a, Berg et al, 2010b).

There are already some examples where the account of economic behaviour is based directly on the available evidence, without taking standard economic theory as a starting point. For example, Abeler et al. (2016) examine people’s preferences for truth-telling. Traditional economic theory assumes that individuals lie when they derive a material payoff for doing so. Abeler et al. examine 72 experimental studies – from psychology and sociology as well as from economics – and observe that people lie surprisingly little. They then test 22 models that have previously been suggested, most of which do not derive from the assumption of *Homo economicus*, or from a reaction or incremental change to it. They find that the model most likely to be true is one that combines multiple types of costs of lying, especially a desire to be honest and a desire to appear honest. Even if the motivation for the study was to disprove a traditional story, the study itself takes a fresh starting point.

Similarly, Rand et al. (2016) study altruism, in a meta-analysis of 22 studies. They ask, are humans intuitively altruistic, or does altruism require self-control? By promoting intuition relative to deliberation in a Dictator Game setting, they find that altruism is intuitive in women but not in men. Their starting point is the Social Heuristics Hypothesis, not traditional economic theory.

A more systematic research programme with a direct, single-stage approach has been carried out by the psychologist Gigerenzer and his co-workers. They study the heuristics that people actually use in practice (Gigerenzer et al., 2000; Gigerenzer et al., 2011). They have observed that people rarely perform complex calculations, instead tending to use ‘fast and frugal’ heuristics (rules of thumb) (Gigerenzer & Goldstein, 1996). For example, the recognition heuristic allows people with little specialized knowledge to make judgements, e.g. on the relative sizes of cities, based on whether or not they have heard of them, whether they have a famous football team, etc. In some cases, this has enabled non-experts to out-perform experts, for example in predicting the outcome of sports tournaments. Also, people do not typically integrate multiple types of information as if they were carrying out an analogue of multiple regression analysis,

A further element in this conception is that these heuristics exploit the information structure of the environment, thus embodying ‘ecological’ rather than logical rationality. They have developed in the course of natural selection, to solve adaptive problems in our evolutionary history (Cosmides & Tooby, 1994b). This perspective opens up the possibility of making links between economic behaviour and its neurobiological or neuroeconomic correlates, as well as the way that neuro/psychological abilities and propensities develop in each individual (Joffe, 2013b).

The direct, one-stage method may provide a useful way of developing economic theory. This does not depend on the validity of the examples given here. Gigerenzer’s fast and frugal heuristics may or may not represent the way forward, or they may apply in some cases and not others. Judgement over the relative merits of the direct and the two-stage approaches needs to be made according to the evidence, on a case-by-case basis.

3.2. The limits of an exclusively behavioural focus

It is clear that the behaviour of the economy depends both on individuals’ behaviour, *ex ante*, and the consequences that later emerge, *ex post*. To determine their relative importance, and the manner in which they interrelate, evidence is required.

Previously, the position of behavioural economics was contrasted with that of experimental economics, especially Vernon Smith’s view that emphasizes how market forces generate a ‘rational’ outcome, even when individuals make behavioural errors. Smith emphasizes the role of time, and the stabilizing property necessarily involves a *causal* process involving human interaction, e.g. between buyer and seller. It also may be that authors who have doubted the practical importance of behavioural factors in finance (e.g. Ross, 2005, Chapter 4) are expressing a similar view.

This stabilizing process does not occur in all situations. Where buyers and sellers interact *in a market*, their initial positions may be far apart, but negotiation leads to an agreement based on the price mechanism. However, in other circumstances the economic outcome is simply the sum total of the decisions of individuals, and if this is subject to bias or irrationality, it will not be corrected subsequently. Examples include decisions on saving for retirement, and on how much insurance to buy. Any systematic bias in the behaviour of individuals will be reflected in the aggregate outcome.

Another type of market interaction that generates systematic consequences not reducible to the actions of individuals is the phenomenon of bubbles. Again, important insights have been achieved in experimental economics, notably by Vernon Smith and colleagues. They have found that endogenous boom-bust cycles occur regularly and robustly in the experimental situation, even when full information on the fundamental value of assets is provided.

Bubbles occur because ‘momentum’ traders interpret a recent asset price rise as a trend that will continue into the future. Their magnitude is greatly influenced by the available quantity of cash (liquidity), leading to a rapid increase in price; this then returns to the fundamental value over time. Trend extrapolation, i.e. that ‘demand and supply are dependent in part on the price change, or derivative, of the asset price’, drives the bubble; it ends when liquidity becomes insufficient for it to continue. Both the trend extrapolation of momentum traders and the impact of excess liquidity are ‘absent in classical price theory’ (Caginalp et al, 2001).

Bubbles have continued to occur in the internet age, with rapid and accurate information now available, and in the experimental situation they are not prevented when full information on asset values is provided. The bubble phenomenon is therefore behavioural, not information-related. It depends on ‘uncertainty about the actions of other traders’, not about the expected payout, and its magnitude is reduced by the experience of repeated trading with the same group of people (Caginalp et al, 2001).

These two well-established phenomena, the price mechanism and the occurrence of bubbles, both result from particular types of human interaction. In the case of markets, buyers and sellers interact to produce movement towards a position of equilibrium, in the sense that further trade is not advantageous to the participants. With bubbles, it is interaction between rival traders that generates the phenomenon. In both cases, the system is driven by causal forces acting through time, rather than an abstract tendency to static equilibrium.

Individual behaviour and interaction thus both have a role. One way in which they inter-relate is that individuals anticipate the likely result, e.g. a seller may set prices that promise to correspond to the amount that potential buyers will be willing to spend, in order to attract them. On the other hand, a seller may set a greatly inflated price, to set a reference point that will influence the buyer to offer a higher price (as with tourist markets in many low- and middle-income countries).

In addition, these systems are driven by behavioural tendencies. The market mechanism is driven by the fact that people tend to respond to incentives. Bubbles are driven by trend extrapolation, the human tendency to expect a past trend to continue, despite the lack of a guarantee that this will occur. Both also involve the quantity of money available, e.g. the wealth of the buyer and the liquidity available for asset purchases, respectively.

Both the market mechanism and bubbles can be described and modelled in terms of the concept of feedback. The price mechanism is a balancing⁴ (‘negative’) feedback system (Stermann, 2000). Momentum trading is an example of reinforcing (‘positive’) feedback. This

⁴ The terms ‘balancing’ and ‘reinforcing’ feedback are preferable to use of the more familiar ‘negative’ and ‘positive’ feedback, respectively, because the latter pair are often misused. For example, reinforcing (positive) feedback may be wrongly termed negative because its consequences are harmful (e.g. in Acemoglu & Robinson, 2012).

is widely understood (e.g. by Shiller, 2005), but not always made explicit – e.g. Caginalp et al (2001) refer to a ‘self-feeding mechanism’, but do not use the word ‘feedback’.

These two types of system share the characteristic that they lead to consequences that are not intended by any individual agent; they are *structured unintended consequences*. As a result, the behaviour of the economic system cannot be reduced to the sum total of individual behaviours in these situations.

Other systems concepts, less well recognised, explain many other phenomena that regularly occur in the economy. The fluctuating prices typical of real estate are traceable to balancing feedback with delay (Sterman, 2000, pp. 698–707). The phenomenon of increasing returns, which can lead to path dependence and technological lock-in, is an example of reinforcing feedback (Arthur, 1994; Sterman, 2000, pp. 349–364 & 387–406). Another manifestation of reinforcing feedback is competition between capitalist firms, an arms race which has been responsible for much of the economic growth since the industrial revolution (Joffe, 2011, 2012).

Still other types of system property, grouped under the heading of complexity, underlie such phenomena as the power law distribution of income, as well as volatility in some types of market. For example, prices of shares and other financial assets, and also certain other items (e.g. commodities such as cotton), display a complex pattern of volatility. The behavioural tendency towards trend extrapolation is a major contributing factor here too (Mandelbrot, 1963; Mandelbrot & Hudson, 2008).

3.3. Two-stage thinking elsewhere in economics

The two-stage approach is not necessarily confined to behavioural economics. In some other branches of economics, traditional theory is also frequently the starting point. This means that the task of the researcher is often seen as explaining how an observed phenomenon deviates from standard theory, rather than explaining the phenomenon itself.

This raises the issue of possible double error: if the first stage does not correspond to a real causal mechanism, then the second stage cannot represent a real causal process either. Whether or not this is true will vary from topic to topic, and must be judged on a case-by-case basis. This cannot be attempted here, and is left for further research.

Two-stage theorizing may be widespread in economics, judging from the terminology that is used. Just as the term “bias” in behavioural economics implies a deviation from something regarded as standard, expected, optimal, normal or natural, much of the language of economics also indicates a divergence. A *prima facie* case can therefore be made that this involves two-stage thinking.

One example is the Solow *residual* (‘a measure of our ignorance’ (Abramovitz 1956)), which refers to the observed GDP growth that is greater than what factor endowments can explain. Another is *imperfect* competition, indicating a departure from the assumed ideal type of perfect competition. A third is market *failure*, signifying a situation where the market fails to generate the expected efficient allocation of goods or services. A fourth is *rent*, in the sense of a reward that departs from (exceeds) the assumed perfect-competition equilibrium. Other candidates include *externalities* – occurring when a third party not directly involved in a

transaction is impacted by it – and *nonlinearity*, implying that the change of the output is not proportional to the change in the input. These terms all refer to something that is *not* happening as expected, and may or may not correspond to an actually-occurring mechanism.

A further possibility is the use of the term ‘puzzle’ to refer to a disjunction between evidence from the real world and a theoretical prediction or expectation. For example in international economics, it has repeatedly been observed that people prefer to trade and to hold shares within their own country (the puzzle of home bias in trade and the puzzle of home bias in equity portfolios, respectively). There is also a strong tendency for investment to be funded by savings within the same country (the Feldstein-Horioka puzzle) (Obstfeld & Rogoff, 2000). In each case, the puzzle relates to the discrepancy between the theory and the reality it seeks to explain. But unlike in the natural sciences, where such observations would be taken as starting points for developing better theory, in traditional economics there is no suggestion that they indicate the need for a fundamentally different approach to theory.

A particularly famous puzzle in international economics is the ‘Lucas puzzle’ (or ‘paradox’). In a classic paper, Lucas (1990) contrasted the theoretical prediction that all capital in rich countries should flow to poor ones, which have less capital and therefore higher returns, with his observation that this did not in fact occur. The explanations that he provided were able to account for the lack of rich-to-poor country flows, but were unable to explain the massive poor-to-rich country flows that started to occur later in the 1990s, notably from China to the US. This phenomenon is actually quite straightforward to explain, and indeed is widely understood, e.g. by high-quality financial journalists (*The Economist* 2014a; 2014b). And yet the academic literature seeks to explain not the phenomenon itself, but rather its deviation from what is regarded as ‘typical’, i.e. standard theory. A widely-supported story is that excess saving occurs because the Chinese financial system is weak; if it were more developed then there would be less saving/more borrowing (Prasad et al, 2007; Sandri 2010; Song et al, 2011; Alfaro et al, 2014) – despite the evidence that in East Asia (including China) ‘more financial development leads to *higher* saving’ (Chinn & Ito, 2008, emphasis in the original). A clear case of double error. Close examination reveals a fundamental problem with the causal direction that is assumed in the academic literature, because of the perceived need to adhere to the standard model as a starting point (Joffe, 2017b).

4. Conclusion

Behavioural economics has made important contributions to our understanding of the way that economic decisions are made. Its relationship with more traditional forms of theory, especially that which involves rational choice theory, can take two different directions. One current tendency is for behavioural factors to be seen as modifying existing standard models. From the perspective of causal mechanism, this implies that two distinct processes exist. The implication is that for both the suggested causal processes, their mechanistic adequacy should routinely be considered in behavioural economics research. In particular, if the postulated first stage, derived from traditional theory based on *Homo economicus*, does not correspond to any real-world causal mechanism, then the second-stage causal process cannot have any real-world existence. This would mean that both stages fail to represent any actual causal processes that operate in the real world – a situation of double error.

An alternative is a direct approach, which abandons the perceived need to start from standard theory. Rather than requiring two processes, one for the stage corresponding to traditional theory and one corresponding to the discrepancy from that theory, it postulates that only one causal process exists. Research then seeks to uncover the mechanism underlying this single process, not for both components of a two-stage process. This may provide a more promising basis for generating a coherent theory of economic behaviour, which is currently lacking. The advantages and disadvantages of each approach need to be decided by the evidence, for each research topic on a case-by-case basis.

Funding details

No financial support was obtained for this work.

Disclosure statement

I declare that I have no conflict of interest relating to this work.

Biographical note

Michael Joffe trained in biomedical sciences in Cambridge, and worked as an epidemiologist at Imperial College before becoming an economist. He has published on evidence-based economics, growth and the capitalist firm, and economic methodology. He has also published on philosophy of causation, including on evidence discovery and causal inference.

Acknowledgements

I would like to thank Alex Rosenberg and Helena Cronin for helpful comments on an earlier draft.

References

- Abeler, J., Nosenzo, D., & Raymond, C. (2016). Preferences for truth-telling (Discussion Paper No 2016-13). The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham. <http://econpapers.repec.org/paper/notnotcdx/2016-13.htm> [14 July 2018]
- Abramovitz, M. (1956). Resource and output trends in the United States since 1870. *American Economic Review* 46: 5–23.
- Acemoglu, D., Robinson, J.A. (2012). *Why nations fail: the origins of power, prosperity and poverty*. London: Profile Books Ltd.
- Ainslie, G. W. (1992). *Picoeconomics*. Cambridge, UK: Cambridge University Press.
- Alfaro, L., Kalemli-Ozcan, S., Volosovych, V. (2014). Sovereigns, upstream capital flows and global imbalances. *Journal of the European Economic Association* 12 (5): 1240–1284.
- Andersen, S., Harrison, G. W., Lau, M., & Rutström, E. E. (2014). Discounting behavior: a reconsideration. *European Economic Review*, 71, 15–33.
- Angner, E., Loewenstein, G. (2012). Behavioral economics. In Uskali Mäki, (ed.) *Handbook of the Philosophy of Science: Philosophy of Economics*, Elsevier.
- Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. Ann Arbor, MI: University of Michigan Press.
- Berg, N., Gigerenzer, G. (2010). As-if behavioral economics: neoclassical economics in disguise? MPRA Paper No. 26586. https://mpra.ub.uni-muenchen.de/26586/1/MPRA_paper_26586.pdf [14 July 2018]
- Berg, N., Eckel, C., Johnson, C. (2010a). Inconsistency Pays?: Time-inconsistent subjects and EU violators earn more. MPRA Paper No. 26589. https://mpra.ub.uni-muenchen.de/26589/1/MPRA_paper_26589.pdf [14 July 2018]
- Berg, N., Biele, G., Gigerenzer, G. (2010b). Does consistency predict accuracy of beliefs?: Economists surveyed about PSA. MPRA Paper No. 26590. https://mpra.ub.uni-muenchen.de/26590/1/MPRA_paper_26590.pdf [14 July 2018]
- Buller, D.J. (2005). Evolutionary psychology: the emperor's new paradigm. *Trends in Cognitive Sciences* 9(6): 277–283.
- Caginalp, G., Porter, D., Smith, V. (2001). Financial bubbles: excess cash, momentum, and incomplete information. *The Journal of Psychology and Financial Markets* 2(2): 80–99.
- Camerer, C.F., Loewenstein, G. (2004). Behavioral economics: past, present, future. In Camerer, C.F., Loewenstein, G., Rabin, M. (eds.) *Advances in behavioral economics*. Princeton: Princeton University Press.

Chinn, M., Ito, H. (2008). Global current account imbalances: American fiscal policy versus East Asian savings. *Review of International Economics* 16: 479–498.

Cosmides, L., Tooby, J. (1994a). Origins of domain specificity: the evolution of functional organization. In L.A. Hirschfeld and S.A. Gelman, eds., *Mapping the mind: domain specificity in cognition and culture*. Cambridge: Cambridge University Press. Reprinted in R. Cummins and D.D. Cummins, eds., *Minds, Brains, and Computers*. Oxford: Blackwell, 2000, 523–543.

Cosmides, L., Tooby, J. (1994b). Better than rational: evolutionary psychology and the invisible hand. *The American Economic Review* 84(2): 327–332.

Craver, C.F., Darden, L. (2013). *In search of mechanisms: discoveries across the life sciences*. Chicago: Chicago University Press.

DeMiguel, V., Garlappi, L., Uppal, R. (2009). Optimal versus naive diversification: how inefficient is the 1/N strategy? *The Review of Financial Studies* 22(5): 1915–1953.

The Economist (2014a). On cloud nine trillion, 29 March 2014, page 65.
<https://www.economist.com/news/china/21599806-our-asia-economics-editor-takes-his-leave-less-worried-many-his-peers-about-frailties> [14 July 2018]

The Economist (2014b). Up, up and away, 29 March 2014, page 75
<http://www.economist.com/news/finance-and-economics/21599774-higher-inflation-may-be-needed-leave-extra-low-interest-rates-behind-up-up-and> [14 July 2018]

Eyster, E., & Rabin, M. (2005). Cursed Equilibrium. *Econometrica*, 73(5), 1623–1672.

Fodor, J.A. (1983). *Modularity of mind: an essay on faculty psychology*. Cambridge, MA: MIT Press.

Friedman, M. (1953). The methodology of positive economics, reprinted in Hausman D. M. (Ed.). (1994). *The philosophy of economics: an anthology*, 2nd edition, Cambridge: Cambridge University Press.

Gigerenzer, G., Goldstein, D.G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review* 103(4): 650–669.

Gigerenzer, G., Todd, P. M., ABC Research Group. (2000). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). *Heuristics*. New York, NY: Oxford University Press

Hausman, D. (1992). *Essays on philosophy and economic methodology*. Cambridge: Cambridge University Press.

- Herfeld, C. (2018). Explaining patterns, not details: reevaluating rational choice models in light of their explananda. *Journal of Economic Methodology*, 25(2), 179–209.
- Heukelom, F. (2014). *Behavioral economics: a history*. Cambridge: Cambridge University Press.
- Joffe, M. (2011). The root cause of economic growth under capitalism. *Cambridge Journal of Economics*, 35, 873–896.
- Joffe, M. (2012). Cost competition between firms: system dynamics and other approaches. Presented at the 30th International Conference of the System Dynamics Society, St Gallen, 2012. <https://www.systemdynamics.org/assets/conferences/2012/proceed/papers/P1283.pdf> [14 July 2018]
- Joffe, M. (2013a). Teaching evidence-based economics. Royal Economic Society *Newsletter*. <http://www.res.org.uk/view/art6Oct13Features.html> [14 July 2018]
- Joffe, M. (2013b). The concept of causation in biology. *Erkenntnis*, 78, 179–197.
- Joffe, M. (2014). Can economics be evidence-based? Royal Economic Society *Newsletter*. <http://www.res.org.uk/view/art4aApr14Features.html> [14 July 2018]
- Joffe, M. (2017a). Causal theories, models and evidence in economics – some reflections from the natural sciences. *Cogent Economics & Finance*, 5, 1280983. <http://www.tandfonline.com/doi/pdf/10.1080/23322039.2017.1280983?needAccess=true> [14 July 2018]
- Joffe, M. (2017b). Why does capital flow from poor to rich countries? – the real puzzle. *Real-World Economics Review*, issue 81, <http://www.paecon.net/PAEReview/issue81/Joffe81.pdf> [14 July 2018]
- Juselius, K. (2011). Time to reject the privileging of economic theory over empirical evidence? A reply to Lawson. *Cambridge Journal of Economics* 35: 423–436.
- Kahneman, D. (1996). Comment [on Plott (1996)]. In K. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The rational foundations of economic behaviour* (pp. 251–254). Basingstoke: Macmillan and International Economic Association.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2), 443–477.
- Lucas, R.E., Jr. (1990). Why doesn't capital flow from rich to poor countries? *American Economic Review* 80(2): 92–96.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36, 394–419.
- Mandelbrot, B., Hudson, R. L. (2008). *The (mis)behavior of markets: a fractal view of risk, ruin, and reward*. New York, NY: Basic Books.

- McKelvey, R. D., & Palfrey, T. R. (1996). A statistical theory of equilibrium in games. *Japanese Economic Review*, 47(2), 186–209.
- Obstfeld, M., Rogoff, K. (2000). The six major puzzles in international macroeconomics: is there a common cause? *NBER Macroeconomics Annual* 15: 339–412.
- O'Donoghue, T., & Rabin, M. (1999). Doing It Now or Later. *American Economic Review*, 89(1), 103–124.
- Panksepp, J., Panksepp, J. (2000). The seven sins of evolutionary psychology. *Evolution and Cognition* 6(2): 108–131.
- Prasad, E., Rajan, R., Subramaniam, A. (2007). The paradox of capital. *Finance and Development (IMF)*, 44(1). <http://www.imf.org/external/pubs/ft/fandd/2007/03/prasad.htm> [14 July 2018]
- Rabin, M. (2013a). An approach to incorporating psychology into economics. *American Economic Review*, 103(3), 617–622.
- Rabin, M. (2013b). Incorporating limited rationality into economics. *Journal of Economic Literature*, 51(2), 528–543.
- Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V., & Barcelo, H. (2016). Social heuristics and social roles: intuition favors altruism for women but not for men. *Journal of Experimental Psychology: General*, 145, 389–396. <http://dx.doi.org/10.1037/xge0000154> [14 July 2018]
- Ross, S. A. (2005). *Neoclassical finance*. Princeton: Princeton University Press.
- Rubinstein, A. (2003). ‘Economics and psychology’? The case of hyperbolic discounting. *International Economic Review*, 44(4), 1207–1216. [doi:10.1111/1468-2354.t01-1-00106](https://doi.org/10.1111/1468-2354.t01-1-00106)
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21, 157–170.
- Sandri, D. (2010). Growth and capital flows with risky entrepreneurship. IMF Working Paper WP/10/37. <https://www.imf.org/external/pubs/ft/wp/2010/wp1037.pdf> [14 July 2018]
- Shiller, R. J. (2005). *Irrational exuberance* (2nd ed.). Princeton, NJ: Princeton University Press.
- Smith, V. (1989). Theory, experiment and economics. *The Journal of Economic Perspectives* 3(1): 151–169.
- Song, Z., Storesletten, K., Zilibotti, F. (2011). Growing like China. *American Economic Review* 101(1): 196–233. <http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.101.1.196> [14 July 2018]

- Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Boston, MA: Irwin McGraw-Hill.
- Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23(3), 165–80.
- Thaler, R. 1987. Anomalies: the January effect. *Journal of Economic Perspectives* 1(1): 197–201.
- Thaler, R. H. (2015). *Misbehaving: The making of behavioural economics*. New York, NY: W. W. Norton & Company.
- Thaler, R. H. (2016). Behavioral economics: past, present, and future. *American Economic Review*, 106(7), 1577–1600.
- Turocy, T. L., & Cason, T. N. (2015, December). *Bidding in first-price and second-price independent-value auctions: a laboratory experiment*. CBESS Discussion Paper 15–23. <https://www.uea.ac.uk/documents/166500/0/CBESS+15-23.pdf/2452beab-0056-49cd-a80b-7b847b1ad97f> [14 July 2018]
- Tversky, A., Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131.
- Uttal, W.R. (2003). *The new phrenology: the limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT Press.