Routledge
Taylor & Francis Group

Journal of Economic Methodology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjec20

Good and bad justifications of analytical modelling

Robert Sugden

To cite this article: Robert Sugden (13 Nov 2023): Good and bad justifications of analytical modelling, Journal of Economic Methodology, DOI: 10.1080/1350178X.2023.2275584

To link to this article: https://doi.org/10.1080/1350178X.2023.2275584

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 13 Nov 2023.

Submit your article to this journal

Article views: 373

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rjec20

JOURNAL OF ECONOMIC METHODOLOGY
https://doi.org/10.1080/1350178X.2023.2275584

Routledge
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

Good and bad justifications of analytical modelling

Robert Sugden

School of Economics, University of East Anglia, Norwich, UK

ABSTRACT

Gilboa, Postlewaite, Samuelson and Schmeidler (2022, Economic theories and their dueling interpretations. *Journal of Economic Methodology*, 1–20. https://doi.org/10.1080/1350178X.2022.2142270; henceforth GPSS) give a 'sociological' account of various ways in which economists claim to find value in 'analytical' models – i.e. models that investigate formal relationships between concepts without deriving substantive empirical or normative conclusions. In this paper, I argue that some of the claims that GPSS report economists as making are defensible, but that others are used in support of modelling strategies that have little or no scientific value.

ARTICLE HISTORY
Received 14 February 2023
Accepted 23 October 2023

KEYWORDS
Models; economics; modelling; economic methodology

1. Introduction

Itzhak Gilboa, Andrew Postlewaite, Larry Samuelson and David Schmeidler (2022; henceforth GPSS) discuss a range of ways in which economic theorists interpret their work. They are particularly concerned with 'analytical' models – models that investigate formal relationships between concepts without deriving substantive empirical or normative conclusions. Although GPSS are distinguished economic theorists and creators of analytical models, they deliberately avoid making judgements about whether economists' claims about the value of such models are defensible. In this paper, I will argue that some of the claims that GPSS report economists as making are defensible, but that others are used in support of modelling strategies that have little or no scientific value.

GPSS begin with a *tour d'horizon* of economics which leads to the main premise of their paper. That premise is that, among the social sciences, economics is an anomaly in 'embracing a single, unifying conceptual framework'. Specifically: 'Most research in economic theory assumes that each agent maximizes an objective function subject to constraints, and the analyst then focuses on the equilibria defined by the interaction of such agents'. In support of this premise, they point to developments in behavioural economics. GPSS argue that, despite behavioural economists' avowed intention to draw on the ideas and research methods of psychology, most behavioural economic theory retains the maximisation-and-equilibrium framework (pp. 1–2).[1]

If it were true that economics was committed to a single predetermined theoretical framework, there would surely be reason to question whether it was a genuine empirical science rather than a branch of mathematics. GPSS argue that biology is another science with a single theoretical framework (p. 6), but they overlook a fundamental difference between economics and biology. Biology's unifying theory – the Darwinian–Mendelian theory of natural selection – is empirical, but the economic theory of rational choice is not.

CONTACT Robert Sugden r.sugden@uea.ac.uk

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

The principle that each agent maximises *some* objective subject to *some* constraints, if not combined with any hypotheses about what those objectives and constraints actually are, is a purely formal mathematical structure. Economic theories with the maximization-and-equilibrium structure have empirical content only in combination with such hypotheses – for example, the hypothesis that consumers have preferences over alternative bundles of commodities and are constrained by income and prices. But why should economics be *committed to* that particular mathematical structure, rather than merely treating it as one that has often proved useful? As viewed by most economic theorists, that structure is grounded on axioms about mutual consistency among abstract propositions expressing preferences and beliefs.[2] The idea that enquiries into real human behaviour should be constrained by a priori conceptions of rational agency seems inappropriate for an empirical science, but it has obvious attractions to theorists. It implies that there is a domain of knowledge in which theorists can make discoveries that constrain the ways in which empirical researchers can explain their findings, but that those discoveries can be made by purely formal methods. If that is true, empirical economists need to attend to the work of theorists, but theorists do not need to reciprocate that attention.

GPSS are right to describe such a conception of economics as 'anomalous' and 'puzzling'. They suggest that 'economists find value in theories and in specific mathematical results in ways that go beyond their use for prediction or recommendation' and undertake to explore 'other ways in which economists find their theories and results valuable, in the hope of better understanding these puzzles'. In this context, one might have expected economists' 'finding value' in a certain class of theories to mean that these theories really have a value, and that economists have found it. GPSS, one might have thought, will be exploring the nature of this value. But they immediately go on to say that this is *not* their purpose. They will simply report 'the way that economists think about their models'. They will tell us what economists say when they *claim to find* value in their theories, but they will not discuss the validity or otherwise of those claims. Explaining this self-imposed constraint, they say that, of course, each of them 'has views on normative issues relating to the way economic research should be conducted' and that these views are '[not] necessarily shared among the authors', but their paper is not about their own views – it is 'sociology of science' (p. 2).

Readers of a methodological journal might reasonably feel frustrated by this constraint. GPSS are themselves the authors of theories that are not obviously directed at prediction or recommendation. Some of the 'finding value' claims they are reporting are presumably claims that they (severally or collectively) have made about their own theories and believe to be defensible. The remark about views that they may not share suggests that at least some of the authors may believe that some of the claims they report are *not* defensible. It is surprising that GPSS are so reluctant to express their own methodological judgements. Nevertheless, they are participant observers of economic theorising, and their reports should be of great interest to methodologists. As a participant observer myself, I can testify to the truth of their accounts of what economic theorists often say and write when they feel called on to justify their work.

Since GPSS are economic theorists themselves, and since they rarely cite sources in support of their claims about how theorists interpret their work, it is not easy for the reader to separate what GPSS say in their own voices from positions which, as sociologists of science, they attribute to theorists in general. It is correspondingly difficult for me to separate critical comments on GPSS's own arguments from critical comments on methodological positions that they articulate but do not explicitly endorse. However, I will try my best.

In Section 2 of this paper, I consider a claim that GPSS make in their own voices – that (as a matter of sociological fact) economics *does* have a unifying conceptual framework, and that behavioural economic theory has been developed within that framework. I will argue that the conceptual framework that GPSS describe is *what has so far survived* of a larger set of rationality principles that were once seen as essential to economic theory. The conceptual framework of economics is evolving to maintain consistency with observations.

In Sections 3 and 4, I consider two types of analytical modelling that feature in GPSS's arguments. In these sections, my comments are directed at methodological positions that GPSS attribute to defenders of these types of modelling. To avoid misunderstanding, let me emphasise that I am writing a commentary on GPSS's paper. Accordingly, I focus on methodological positions that GPSS have chosen to discuss. Other ways of understanding analytical modelling are beyond the scope of this paper.

Section 3 deals with analytical modelling in 'pure' theory, focusing (as GPSS do) on Robert Aumann's (1987) model of correlated equilibrium. GPSS declare their own belief that Auman's paper 'clearly made a methodological contribution' (p. 5), and consider different interpretations of what that contribution might be. In line with one of these interpretations, I argue that Aumann's model is best understood as 'conceptual exploration' – as an enquiry into the mathematical structure of an abstract conception of rationality and common knowledge. I conclude that this kind of modelling can be justified in its own right, even if it makes no reference to empirical phenomena.

Section 4 deals with a different type of analytical modelling. This type of modelling starts from an observed regularity in the real world which, viewed in the perspective of the unifying conceptual framework, seems anomalous. The modeller's objective is to find auxiliary assumptions, *however unrealistic*, which allow the regularity to be 'explained' within the standard framework. These are the exercises that I characterise as having little or no scientific value. In describing how theorists understand this second type of analytical modelling, GPSS (writing in their own voices) claim that this understanding is 'akin to' my account of models as 'credible worlds' (Sugden, 2000, 2009). In Section 5, I disagree.

2. Behavioural economics and the 'unifying conceptual framework' of economics

According to GPSS's account, behavioural economic theory is located within a unifying conceptual framework which, by implication, existed before the empirical findings of behavioural economics were generally recognised. We are invited to infer that that framework was not (and is not) threatened by behavioural findings.

Similar arguments have been made by other economists. For example, Glenn Harrison (2010) advances a version of the argument in a manifesto for a 'behavioral counter-revolution'. Presenting this manifesto as in the same spirit as Vernon Smith's (2010) challenges to psychologists' interpretation of experimental evidence, Harrison opens with the claim:

> In effect the behavioral revolution in recent years has served to remind the economics profession that it simply forgot to answer many of the questions that it should have been answering all along with our traditional tools. In other words, we did not need a behavioral revolution: we just needed to do our job better as experimental economists. (2010, p. 49)

He ends with:

> There is no question that economics has been improved by being reminded of the heterogeneity of economic behavior, and the fact that much of the behavior observed does not fit well with existing models. If one can get past the marketing hype of the behavioral economics tradition, then this contribution can and should be applauded. However, now begins the more serious task of restating, re-applying, and extending the tools of traditional economics. (2010, p. 56)

In his characteristic take-no-hostages style, Harrison seems to be distinguishing, as GPSS do, between some fundamental theoretical framework ('traditional economics') and particular models that have been developed within that framework. He rejects as 'marketing hype' claims by psychologists to have disconfirmed properties of the fundamental framework. Psychologists deserve mild applause for pointing out the limitations of some particular economic models, but those economists who had upheld traditional economics against psychologists' challenges had been right all along (even if they had forgotten why). The really serious task is to create models within the traditional framework that are consistent with the new findings.

But all this is using hindsight to rewrite the history of the 'behavioural revolution'. I will focus on one controversy in which I have participated – the controversy about the interpretation of observed disparities between willingness-to-pay (WTP) and willingness-to-accept (WTA) valuations – but I could equally well have used other controversies (for example, about the interpretation of observed violations of the independence axiom of expected utility theory). The WTP–WTA disparity first came to light in responses to contingent valuation surveys (e.g. Bishop & Heberlein, 1979). As viewed by theorists at the time, the preferences reported by survey respondents violated one of the most fundamental principles of rational choice theory – the existence of preference orderings over potential objects of choice. Behavioural economists (let me say 'we', since I was one of them) interpreted this observation as suggesting that human decision makers have asymmetric attitudes to gains and losses.

In the early stages of the controversy, our more orthodox opponents used two lines of argument to try to reconcile the disparity with the received theory of rational choice. One was to point out that contingent valuation surveys have no direct incentives for truthful reporting of preferences, and that if respondents anticipate how survey results will be used, there may be incentives for *mis*reporting. The second was to point out that standard consumer theory has the qualitative implication that, if income effects are normal, WTA is greater than WTP. We responded to these challenges by running incentivised experiments which elicited WTP and WTA valuations of ordinary consumption goods and which built in controls for income effects (e.g. Knetsch & Sinden, 1984). The disparity remained. We also used theoretical analysis to show that observed WTP–WTA disparities were far larger than those that could be induced by income effects in the standard theory (Horowitz & McConnell, 2003; Sugden, 1999). To explain the disparity and a range of related effects, we proposed models of *reference-dependent* preferences (Munro & Sugden, 2003; Tversky & Kahneman, 1991). In these models, preference is a trinary rather than binary relation: an individual prefers one consumption bundle x to another bundle y, viewed in relation to some reference point r.

Clearly, these models are *developments of* previously orthodox forms of rational choice theory. That should not be surprising, since standard consumer theory had worked reasonably well in explaining and predicting behaviour in many different settings. But the fact remains that the behavioural models dispensed with an assumption (the existence of context-independent preferences) that had previously been seen as just as much a part of the basic conceptual framework of economics as the concepts of maximisation and equilibrium. If economists now regard reference-dependent preferences as compatible with a unifying conceptual framework, that is because the framework itself has changed. It has changed through the work of researchers who took the initially suggestive evidence seriously, proposed new hypotheses to explain it, and tested those hypotheses in controlled experiments.

3. Analytical models: the case of pure theory

Although GPSS offer a general account of the role of models in economics, they give particular emphasis to models that are not directed towards prediction or recommendation. They suggest that by understanding why economists find value in such models, we may be able to understand 'why economists are so blasé about observations that their assumptions are far-fetched, and so unperturbed by refutations of their theories' (p. 2).

GPSS propose a classification of models as *positive*, *normative* or *analytical*. They do not give an explicit definition of an analytical model, but instead give a physical example. An architect's maquette of a town square is an analytical model if it is intended 'to test the feasibility of a possible square' (p. 4). By implication, an analytical model tests the feasibility of what it tries to represent. The type of modelling that GPSS have in mind seems similar to what Daniel Hausman (1992, p. 221) classifies as 'conceptual exploration'.

GPPS give two initial examples of analytical models in economics. The first is Aumann's (1987) model of correlated equilibrium. GPSS interpret Aumann as using this model 'to demonstrate that

the notion of the common knowledge of rationality could be incorporated within standard techniques'. The second example refers to a family of models which 'explain apparent anomalies as outcomes of Bayesian Nash equilibria', thereby 'prov[ing] the consistency of the assumptions [i.e. Bayesian Nash equilibrium] with some stylized facts' (p. 5).

There is a significant difference between these two examples. The 'apparent anomalies' of the second example are stylized facts, i.e. general tendences in the behaviour of economic agents in the real world. Analytical modelling is being used to show the feasibility of a certain type of explanation of those facts. In contrast, common knowledge of rationality is not a stylized fact about the real world. It is not even a *possible* fact about the real world; it is a theoretical concept. The aim of Aumann's modelling is to test whether this concept is coherent. This is a research problem in the domain of pure theory. In the remainder of this section, I will consider the value of this type of analytical modelling. I will return to GPSS's second example in Section 4.

At the time Aumann was constructing his model, there was broad agreement among economic theorists that rationality in individual decision-making was best represented by subjective (or 'Bayesian') expected utility theory. In addition, theorists had arrived at a coherent understanding of common knowledge (within a set of individuals) of an 'event', defined as a set of states of nature. In this conception, each individual has a given 'information partition' of the set of all states of nature; what the individual knows, conditional on any state, is the element of their information partition that contains that state. This set-up allows a definition of common knowledge, conditional on a state of nature, in terms of intersections between individuals' information partitions.[3] But game theory also had the concept of *common knowledge of rationality* (CKR). Intuitively, there is CKR if each player of a game is rational in the sense of subjective expected utility, knows that this true of every other player, knows that every other player knows this, and so on. It was not obvious that this construct was coherent. A further open question concerned the status of Nash equilibrium. Most applications of game theory implicitly assumed that the strategies chosen by the players in a game would constitute a Nash equilibrium, but Nash equilibrium was not an implication of CKR.

Aumann's model is an attempt to resolve these issues by widening the concept of a 'state of nature' to that of a 'state of the world', the specification of which includes which strategy each player chooses at that state. This has the apparently paradoxical feature that the alternative strategies available to a player, as viewed by that player in the perspective of standard decision theory, are simultaneously acts (i.e. options that might be chosen, with consequences that depend on which event occurs) and events in themselves. Nevertheless, Aumann creates a mathematically consistent model in which CKR is well-defined and in which CKR, so defined, implies that players' chosen strategies constitute a 'correlated equilibrium' – a generalisation of Nash equilibrium. He has taken three abstract theoretical frameworks – Bayesian decision theory, information theory and game-theoretic equilibrium – and unified them in a single formal structure.

Unless I have missed something, Aumann makes only one reference to empirical properties of the real world in the whole paper, and that reference is an indirect one. He devotes one paragraph to pointing out that one of the key assumptions of his model – that players have common priors over states of the world – is 'pervasive in the enormous literature on rational expectations, trading in securities, bargaining under incomplete information, […], bankruptcy, what have you' (pp. 12–13). Notice that Aumann is not claiming that this assumption is a reasonable approximation to the truth. He is still working in the domain of theory, but he can perhaps be read as expressing the belief that he is contributing to a branch of economics that has explanatory power in the real world. What I take to be his principal justification of the common prior assumption is that it 'expresses the view that probabilities should be based on information; that people with different information may legitimately entertain different probabilities, but there is no rational basis for people who have always been fed precisely the same information to do so' (pp. 13–14).

What Aumann is doing, I suggest, is creating a mathematical structure to represent a *conception of rationality*. I use the term 'conception' in the sense that John Rawls (1971, pp. 3–6) uses when he distinguishes between concepts and conceptions. In Rawls's sense, the *concept* of rationality

corresponds roughly with the content of a dictionary definition of the word 'rationality'. (The *Concise Oxford English Dictionary* defines 'rational' as 'based on or in accordance with reason, able to think sensibly or logically'). A *conception* of rationality is a tighter specification of a particular way of thinking about rationality. On this reading, Aumann's model is a mathematical structure with an associated interpretation, but that interpretation is conceptual rather than empirical. Much subsequent debate about this model has focused on whether the model is philosophically coherent as a conception of rationality. For example, critics have challenged Aumann's argument that, given the same information, rational agents necessarily form the same subjective beliefs (e.g. Gul, 1988). Robin Cubitt and I have shown that, although Aumann's method of representing common knowledge of rationality is consistent if 'rationality' is identified with subjective expected utility, it leads to logical contradictions if it is used in combination with some other apparently reasonable forms of received decision theory (Cubitt & Sugden, 2014).

Whatever conclusions one draws from these debates, it is surely uncontroversial that game theorists routinely assume CKR in *some* form, and that it is therefore important to ask what that form is, whether it is coherent, and how far it corresponds with everyday understandings of the concept of rationality. Since many economic explanations of real-world phenomena use the tools of game theory, there is reason to think that answers to these questions might ultimately contribute to empirical science. But that is not the immediate purpose of models such as Aumann's. GPSS are right to say (in their own voices) that models that are constructed with the aim of conceptual exploration can be justified in their own right, while also producing 'building blocks' for analyses that are directed at prediction (pp. 3–4).

4. Analytical models: explaining apparent anomalies

I now turn to GPSS's second example of analytical modelling:

> Theorists who explain apparent anomalies as outcomes of Bayesian Nash equilibria need not believe that such equilibria, or even common knowledge of rationality, are very plausible. If asked, 'Why do you propose this explanation, then?,' they might say, 'Well, I believe that there is some value in testing whether the standard assumptions are compatible with the phenomenon at hand'. […] Proving that there exists a formal mathematical structure within which all agents are rational, have commonly known beliefs, and exhibit a certain behavior pattern is an exercise that proves the consistency of the assumptions with some stylized facts. (p. 5)

This is a peculiar perspective on economic modelling. GPSS are describing the work of a representative theorist who is aware of some regularity in the real world (the 'stylized facts'). Taken at face value, this phenomenon is inconsistent with assumptions that are deeply embedded in what GPSS call the 'unifying conceptual framework' of economics. The theorist's proposed explanation is a formal mathematical structure that is consistent with the standard framework, but which may include any number of other assumptions whose truth value is unknown. These assumptions may be particular specifications of agents' preferences and beliefs. Or they may be unverified assumptions about the environment in which the phenomenon is observed. (In one of GPSS's examples, the phenomenon to be explained is individuals' concern for their social status.[4] The proposed explanation – which GPSS rather questionably describe as 'surprisingly simple' – is a multi-generational game-theoretic model of bequests and status-seeking marriages which makes many specific assumptions about the society it represents.)

Notice that, although GPSS are discussing models that theorists propose as explanations of empirical phenomena, their account of the purpose of these models says nothing about their value *as explanations*. As represented by GPSS, the theorist's interest is not in the phenomenon itself, but in the 'standard assumptions' of a pre-existing theoretical framework. The aim of the modelling exercise is to arrive at conclusions *about those assumptions*. Faced with the criticism that the model is an implausible explanation of the phenomenon, GPSS's theorist replies that he doesn't care about that: what matters is that the standard assumptions have been shown to be logically

consistent with a description of the phenomenon. This account of modelling is crystalised in GPSS's declaration that '[t]he main feature of an analytical model seems to us a formal problem of the type: Is there a model in a given class that can exhibit a given type of observations?'[5] (p. 5)

GPSS express the same thought when they refer to the model of social status:

> [A]n argument that familiar models cannot account for the fact that people appear to envy the economic well-being of others gives way to the observation that parents may be concerned about the welfare of their children, who will face a tournament in which desirable mates go to the wealthy. (p. 6)

As I read this passage, GPSS are imagining a theorist who is committed to some set of standard assumptions. The theorist is challenged by someone who cites evidence that people envy the economic well-being of others and claims that this is inconsistent with those assumptions. The theorist produces the model and thereby wins the argument; the critic has to 'give way'. It seems that the critic is not allowed to ask whether the model is a good or bad explanation of the actual evidence.

GPSS might reasonably point out that the passages I have quoted so far in this section are part of their discussion of *analytical* models, and that it is a matter of definition that analytical models are used to address issues that are internal to theory. But then what are GPPS's theorists doing when they propose these models *as explanations of* empirical observations? My sense of methodological unease is heightened by GPSS's account of how 'the same researcher might prefer different interpretations [of their own model] in different contexts or at different times'.[6] GPSS describe a practice that (here I have to agree with them) is not uncommon among theorists:

> [O]ne may suggest a model with a descriptive interpretation in mind, but, when facing an aggressive audience, one might take a step back and rather than promoting the model as an explanation of a real-life phenomenon, present it as a 'proof of concept' or 'merely an exercise' in testing the scope of the standard paradigm. (p. 7)

Think what is going on here. A theorist has created a model which he customarily presents as a proposed explanation of some empirical phenomenon. But now he is facing an audience that has sufficient knowledge of the evidence about this phenomenon, or of other related phenomena within the explanatory scope of the model, to raise pertinent questions about the plausibility of that explanation. (It is revealing that, as viewed by theorists, reporting such knowledge can be seen as aggressive rather than informative.) The theorist's response is to make a temporary change in the way he 'promotes' his model (another revealing term), before reverting to the original one when facing less knowledgeable or less critical audiences. This surely amounts to acting in bad faith.

The truth is that GPSS are describing a type of analytical modelling whose intent is not explanatory in any serious scientific sense. It is a branch of pure theory which treats items of empirical evidence rather as if they were the basis for examination questions in an advanced course in economic theory. It as if the examinee is told to assume the truth of some empirical proposition and then, using the tools of standard theory, to create a model that generates that proposition as an implication; marks are to be awarded for theoretical virtuosity and technical rigour. By implication, GPSS's theorists claim that this kind of modelling has scientific value. But what (according to those theorists) is that value?

GPSS do not give an explicit answer to this crucial question. Given GPSS's emphasis on the importance that economists attach to their 'unifying conceptual framework', the answer that best fits their account seems to be that 'show[ing] that seemingly anomalous results can be encompassed within the standard theory, even if it is difficult to defend the resulting models as realistic' is evidence in support of that framework (p. 6). But, presented in the context of empirical science, that argument seems methodologically unsound. The test of the scientific value of an overarching theoretical framework should be its ability to generate *true* or *sound* explanations of observations. However one construes 'truth' or 'soundness', that value can be assessed only through an activity that GPSS's analytical modellers seem to disdain – namely, looking at the evidence and following it where it leads.

What I have called 'bad faith' in analytical modelling is not only ethically unworthy; it is also an obstacle to scientific progress. The early years of behavioural economics, described in Section 2,

provide an example. Those of us who proposed psychologically-based explanations of experimentally observed anomalies such as the WTP–WTA disparity and violations of the independence axiom were in contention with theorists who proposed a variety of different models in which those anomalies were consistent with what was then the unified conceptual framework of economics. The prevailing view among economists was that these models provided grounds for scepticism about the behavioural approach. It was possible to run new experiments to test for the anomalies we thought we had already found, but with additional controls to screen out the mechanisms which, according to our opponents, might be causing those anomalies. (Recall Knetsch and Sinden's tests of the WTP–WTA disparity with controls for income effects.) In at least some cases, the experimenters running these tests had the prior expectation that screening out the relevant mechanism was unlikely to have a significant effect, but believed that they still had a duty to investigate it. That is how good science progresses.

But suppose such an experiment has been run and the opponent's hypothesis has been rejected. Suppose the opponent then says: 'I'm not at all surprised about your findings. When I proposed that mechanism as an explanation of your original observations, I knew it was implausible. My model was merely a theoretical exercise to show that your original observations were logically consistent with the standard theory. Now I'll try to find an implausible explanation of your new results, and you can test that.' This is *not* good science. If all that can be claimed for a model is that it is a theoretical exercise, empirical scientists should not be expected to treat its results as hypotheses that deserve to be tested.

5. Credible worlds and potential explanations

GPSS (p.5) claim that their interpretation of analytical modelling is 'akin to' my account of models as 'credible worlds' (Sugden, 2000, 2009). I think there is a fundamental misunderstanding here.

My account starts from the premise that 'model-building in economics has serious intent only if it is ultimately directed towards telling us something about the real world'. I recognise that many theoretical models in economics are 'abstract and unrealistic' and 'lead to no clearly testable hypotheses'. Disassociating myself from 'those economic theorists who, off the record at seminars and conferences, admit that they are only playing a game with other theorists', I look for the distinguishing characteristics of models that are genuinely informative. My strategy is to study two famous models that are abstract and unrealistic and that do not lead to testable hypotheses, but which almost all economists regard as in some way informative about the real world. My object is 'to discover just what these models do tell us about the world, and how they do it' (Sugden, 2000, pp. 1–2).

One of my examples is Thomas Schelling's (1978) 'checkerboard' model of racial segregation – a model that has featured in many subsequent methodological discussions. (My other example is George Akerlof's 'market for lemons'.) It is hard to find any feature of Schelling's modelling strategy that matches GPSS's account of analytical models.

On any reading of Schelling, he is not positioning his work in relation to standard economic theory – or to any other theory. He begins from familiar observations of spontaneously emerging patterns of binary social segregation, such as between whites and blacks in American residential areas or between men and women at 1960s cocktail parties. No then-current theory treats those patterns as anomalous; in so far as his contemporaries think about the issue at all, they think of them as evidence of strong preferences for like-with-like association. Schelling's hunch is that sharp segregation can be (or is? – see later) the result of a mechanism of 'sorting and mixing' among spatially located individuals who have some freedom to move and who are mildly averse to (or perhaps merely embarrassed by) being in a small minority in their locality. He develops a dynamic model of this mechanism using dimes and pennies on a checkerboard. This model is not in the spirit of maximisation-and-equilibrium. It is now generally admired as a pioneering work of agent-based modelling and evolutionary economics, but Schelling does not present it as an experiment in theory: it is a model of a 'social mechanism' (Schelling, 2006, pp. 235–248).

I interpret Schelling's checkerboard model as a *credible world*. A credible world is a description of a self-contained small world that the modeller has created. There is no claim that the model is an approximation to or abstraction from any reality, but it is credible in the sense of truthlikeness or verisimilitude: it is realistic in the same sense that a novel can be both fictional and realistic. The workings of Schelling's model produce patterns of dimes and pennies that are *similar to* observed patterns of social segregation. On my reading, Schelling is claiming that his model explains those observations. That claim is based on what I argue to be a scientifically defensible inductive inference. If this is what Schellling is doing, it is not analytical modelling in GPSS's sense: it is empirical through and through.[7]

Several methodologists have challenged my interpretation of the checkerboard model, arguing that Schelling is presenting it only as a *potential* explanation of observed patterns of segregation (e.g. Aydinonat, 2007; Grüne-Yanoff, 2009; Mäki, 2009). On this latter view, Schelling is showing that the conventional explanation of segregation, i.e. as the direct result of strong like-with-like preferences, *might be* false. In other words, Schelling's model is a test of the feasibility of explaining observed segregation without assuming such preferences.

If the issue is how best to read Schelling's text, there is evidence for both interpretations. But if Schelling is presenting his model only as a potential explanation, his understanding of 'potential' must be stronger than the concept of logical possibility that is implicit in GPSS's account of analytical modelling. It has empirical content: he is claiming that his explanation *may be* true. To put this another way, he is inviting his readers to entertain the possibility that it *is* true, and perhaps to act on this possibility (for example, by carrying out further research to investigate *whether* it is true). What makes me so sure that Schelling *must* mean this? Any fair reader of Schelling's text can sense that he is *interested in* the social behaviour he observes. He is curious about why it occurs. He *wants* to explain it. In contrast, the interest and curiosity of GPSS's analytical modellers seems to be directed at formal structures. A pure theorist might see Schelling's model as a (perhaps regrettably informal) way of proving a theorem about an interesting mathematical structure, but that is surely not how Schelling sees it. For Schelling, it is an attempt to understand the real world of housing markets and cocktail parties.

As I have argued in this paper, using Aumann's model of correlated equilibrium as an illustration, pure theory has a proper place in economics. That is true even if economics is ultimately understood as an as empirical science. But for theorists who want to propose explanations of concrete real-world phenomena, I commend Schelling's approach to economics. Intentionally or not, GPSS have revealed the poverty of some of the arguments by which economic theorists excuse themselves from taking an interest in the phenomena they purport to explain.

Notes

1. Throughout the present paper, unspecified page references are to Gilboa et al. (2022).
2. Theorists sometimes try to justify these axioms on the grounds that agents who violate them are vulnerable to 'money pumps' (or 'Dutch books'). But the money pump argument is not empirical; it is a formal theorem within the standard theoretical framework. It refers only to agents who satisfy some of the standard axioms but consistently contravene others (see Cubitt & Sugden, 2001).
3. This definition is fundamentally different from the one proposed by Lewis (1969) in the first formal analysis of common knowledge. Cubitt and Sugden (2014) argue that some of the paradoxical properties of Aumann's model result from using the modern definition rather than Lewis's.
4. The model in GPSS's example is that of Cole et al. (1992).
5. In a footnote, GPSS confirm that, in their account of this type of model, '[f]ormally speaking, a test of consistency allows for rather fanciful assumptions as long as they are within the paradigm', but add the qualification that 'the practice of economic theory imposes additional restrictions of plausibility, which make the "mere test of consistency" closer to the notion of "credibility"' (p. 16, footnote 10). It is still striking that the issue of whether a proposed explanation is plausible is relegated to a footnote and that credibility is treated as a kind of optional extra.
6. GPSS cite Leonard Savage's response to the Allais Paradox as an example of this practice. This is unfair. From the outset, Savage's project was to find normative foundations for statistical theory. He was surprised to find that his

own intuitively reasonable preferences contravened one of his axioms, and gave serious thought to the normative implications of this (Dietrich et al., 2021).

7. I cannot see the basis for GPSS's statement that 'Sugden focuses on the role of logical coherence in the judgment of "credibility"' (p. 5). Logical coherence is a minimal requirement of *any* argument in science or philosophy, and hence of any model. A model world is credible, and thereby a suitable case from which to draw inductive inferences, '[to] the extent to which we can understand [it] as a description of how the world *could be*' (Sugden, 2000, p. 24, emphasis in original).

Acknowledgements

In writing this paper, I have benefitted from discussions with Yam Maayan.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Robert Sugden is Professor of Economics at the University of East Anglia. His research uses a combination of theoretical, experimental and philosophical methods to investigate issues in behavioural economics, normative economics, choice under uncertainty, the foundations of decision and game theory, the methodology of economics, and the evolution of social conventions. His current work aims to reconcile behavioural and normative economics, using principles of opportunity and mutual advantage rather than welfare.

References

Aumann, R. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, *55*(1), 1–18. https://doi.org/10.2307/1911154

Aydinonat, N. E. (2007). Models, conjectures and exploration: An analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology*, *14*(4), 429–454. https://doi.org/10.1080/13501780701718680

Bishop, R., & Heberlein, T. (1979). Measuring values of extramarket goods: Are indirect measures biased? *American Journal of Agricultural Economics*, *61*(5), 926–930. https://doi.org/10.2307/3180348

Cole, H., Mailath, G., & Postlewaite, A. (1992). Social norms, savings behavior, and growth. *Journal of Political Economy*, *100*(6), 1092–1125. https://doi.org/10.1086/261855

Cubitt, R., & Sugden, R. (2001). On money pumps. *Games and Economic Behavior*, *37*(1), 121–160. https://doi.org/10.1006/game.2000.0834

Cubitt, R., & Sugden, R. (2014). Common reasoning in games: A Lewisian analysis of common knowledge of rationality. *Economics and Philosophy*, *30*(3), 285–329. https://doi.org/10.1017/S0266267114000339

Dietrich, F., Staras, A., & Sugden, R. (2021). Savage's response to Allais as Broomean reasoning. *Journal of Economic Methodology*, *28*(2), 143–164. https://doi.org/10.1080/1350178X.2020.1857424

Gilboa, I., Postlewaite, A., Samuelson, L., & Schmeidler, D. (2022). Economic theories and their dueling interpretations. *Journal of Economic Methodology*, 1–20. https://doi.org/10.1080/1350178X.2022.2142270

Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis*, *70*(1), 81–99. https://doi.org/10.1007/s10670-008-9138-6

Gul, F. (1988). A comment on Aumann's Bayesian view. *Econometrica*, *66*(4), 923–927. https://doi.org/10.2307/2999578

Harrison, G. (2010). The behavioral counter-revolution. *Journal of Economic Behavior & Organization*, *73*(1), 49–57. https://doi.org/10.1016/j.jebo.2008.11.007

Hausman, D. (1992). *The inexact and separate science of economics*. Cambridge University Press.

Horowitz, J., & McConnell, K. E. (2003). Willingness to accept, willingness to pay and the income effect. *Journal of Economic Behavior & Organization*, *51*(4), 537–545. https://doi.org/10.1016/S0167-2681(02)00216-0

Knetsch, J., & Sinden, J. A. (1984). Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *The Quarterly Journal of Economics*, *99*(3), 507–521. https://doi.org/10.2307/1885962

Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press.

Mäki, U. (2009). MISSing the world. Models as isolations and credible surrogate systems. *Erkenntnis*, *70*(1), 29–43. https://doi.org/10.1007/s10670-008-9135-9

Munro, A., & Sugden, R. (2003). On the theory of reference-dependent preferences. *Journal of Economic Behavior & Organization*, *50*(4), 407–428. https://doi.org/10.1016/S0167-2681(02)00033-1

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Schelling, T. (1978). *Micromotives and macrobehavior*. Norton.

Schelling, T. (2006). Dynamic models of segregation. In T. Schelling (Ed.), *Strategies of Commitment and Other Essays* (pp. 249–310). Harvard University Press. First Published in *Journal of Mathematical Sociology* 1 (1971).

Smith, V. (2010). Theory and experiment: what are the questions? *Journal of Economic Behavior & Organization*, *73*(1), 3–15. https://doi.org/10.1016/j.jebo.2009.02.008

Sugden, R. (1999). Alternatives to the neoclassical theory of choice. In I. Bateman, & K. Willis (Eds.), *Valuing environmental preferences: Theory and practice of the contingent valuation method in the US, EC and developing countries* (pp. 153–181). Oxford University Press.

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, *7*(1), 1–31. https://doi.org/10.1080/135017800362220

Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis*, *70*(1), 3–27. https://doi.org/10.1007/s10670-008-9134-x

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956