Mentalism versus behaviourism in economics: a philosophy-of-science perspective*

Franz Dietrich
CNRS & University of East Anglia

Christian List
London School of Economics

18 November 2012

Abstract

Behaviourism is the view that preferences, beliefs, and other mental states in social-scientific theories are nothing but constructs re-describing people's behavioural dispositions. Mentalism is the view that they capture real phenomena, no less existent than the unobservable entities and properties in the natural sciences, such as electrons and electromagnetic fields. While behaviourism has long gone out of fashion in psychology and linguistics, it remains influential in economics, especially in 'revealed preference' theory. We aim to (i) clear up some common confusions about the two views, (ii) situate the debate in a historical context, and (iii) defend a mentalist approach to economics. Setting aside normative concerns about behaviourism, we show that mentalism is in line with best scientific practice even if economics is treated as a purely positive science of economic behaviour. We distinguish mentalism from, and reject, the radical neuroeconomic view that behaviour should be explained in terms of people's brain processes, as distinct from their mental states.

1 Introduction

Economic theory seeks to explain the social and economic behaviour of human (and sometimes other) agents.[1] It usually does so by (i) ascribing, at least in an 'as if' mode,

*This paper was presented at the LSE Choice Group workshop on 'Rationalizability and Choice', July 2011, the D-TEA workshop, Paris, July 2012, and the EIPE seminar, Rotterdam, September 2012. We are grateful to the participants and especially Nick Baigent, Walter Bossert, Richard Bradley, Mikaël Cozic, Eddie Dekel, Ido Erev, Itzhak Gilboa, Conrad Heilmann, Johannes Himmelreich, Marco Mariotti, Friederike Mengel, Clemens Puppe, Larry Samuelson, David Schmeidler, Asli Selim, Daniel Stoljar, Kotaro Suzumura, and Peter Wakker for comments and discussion.

[1]We here focus on micro-economic theory. Other agents to which the theory is sometimes applied include corporate agents and even non-human animals (in behavioural ecology). On corporate agency,

certain mental states, such as beliefs and desires, to the agents in question and (ii) showing that, under the assumption that those agents are rational, the ascribed mental states lead us to predict, and thereby to 'rationalize', the behaviour to be explained.[2] For example, we explain why Franz went to Starbucks in the afternoon and bought a cappuccino by saying that he had a desire to drink coffee and a belief that there was coffee available at Starbucks, so that it was rational for him to take the action. Classical economic theory formalizes this explanation, in the simplest form, by representing Franz's desires in terms of a preference ordering or utility function over various outcomes and his beliefs in terms of a subjective probability function over various states of the world, and by defining as rational any action that maximizes expected utility. Setting aside the technical terminology, the logic underlying this explanation is very similar to the logic underlying ordinary folk-psychological reasoning with its ascription of mental states to explain behaviour. Economic explanations can thus be seen as more sophisticated and scientifically elaborated reconstructions of folk psychology.[3]

But what is the status of the ascribed mental states and of the resulting explanation of Franz's behaviour? In particular, are the ascribed mental states (e.g., subjective probability and utility functions)

(1) mere *re-descriptions of behavioural patterns* and perhaps *instrumentally useful constructs* for organizing and making sense of empirical regularities,

or are they

(2) *representations of real mental/psychological phenomena*, no less existent in the world than the (also not directly observable) electrons, neutrinos, and electromagnetic fields postulated in the natural sciences?

Roughly, *behaviourism* is the first of these two views, whereas *mentalism* is the second; we will make this more precise later.

Behaviourism used to be the dominant view across the behavioural sciences, including not only economics, where it was pioneered by scholars such as Vilfredo Pareto (1848-1923), Paul Samuelson (1915-2009), and Milton Friedman (1912-2006), but also psychology and linguistics, where it was prominently expressed, for example, by Ivan

see List and Pettit (2011). On biological applications, see a special issue on group decision making in humans and animals, edited (with introduction) by Conradt and List (2009).

[2]For an overview of theories of choice and rationalization, see Bossert and Suzumura (2010).

[3]Economics thereby exemplies a familiar feature of science more generally, which Quine famously described as commonsense gone self-conscious (Quine 1960). On the relationship between economic decision theory and folk psychology, see also Pettit (1991) and Mongin (2011).

Pavlov (1849-1936), Leonard Bloomfield (1887-1949), and B. F. Skinner (1904-1990). Bloomfield wrote: 'The terminology in which at present we try to speak of human affairs – [...] "consciousness", "mind", "perception", "ideas", and so on – [...] will be discarded [...] Non-linguists [...] forget that a speaker is making noise, and credit him, instead, with the possession of impalpable "ideas". It remains for the linguist to show [...] that the speaker has no "ideas" and that the noise is sufficient' (quoted in Langendoen 1998).

In psychology and linguistics, especially since Noam Chomsky's influential critique (1959) of Skinner, behaviourism has long been replaced by some versions of mentalism (e.g., Katz 1964, Fodor 1975), though often under different names, such as 'cognitivism' or 'rationalism'. Many forms of behaviour, it is now widely accepted, are hard to explain unless we pay attention to the underlying cognitive mechanisms giving rise them. Chomsky argued that the way in which children learn languages (for example, they never make certain kinds of grammatical mistakes that, from a purely combinatorial perspective, we would expect) would be hard to explain if we thought of children as mere stimulus-response systems, without any innate language processing capacities (Pinker 1994, cf. Tomasello 1995).

In economics, by contrast, behaviourism continues to be very influential and, in some parts of the discipline, even the dominant orthodoxy.[4] The 'revealed preference' paradigm, in many of its forms, is behaviouristic, though there are more and less radical versions of it. Recently, Faruk Gul and Wolfgang Pesendorfer (2008) have offered a passionate defence of what they call a 'mindless economics', a particularly radical form of behaviourism.

In this paper, we aim to clear up some common conceptual confusions about behaviourism and mentalism in economics, situate the debate within the broader context of the philosophy of science, and defend a mentalist approach to economics, which we argue is in line with best scientific practice. We thereby reject Gul and Pesendorfer's case for behaviourism, though we do so from a different, more philosophy-of-science-oriented perspective than earlier, for instance normative-economic and neuroeconomic, responses to them (e.g., Kőszegi and Rabin 2007, Harrison 2008, and the contributions to Caplin and Schotter's 2008 collection; some of our criticisms are shared by Hausman 2008). Crucially, we show that a case for mentalism can be made even if economics is

[4]Behaviourism should not be conflated with behaviourial economics, which emphasizes the need for economic models to incorporate insights from psychology (see, e.g., Camerer, Loewenstein, and Rabin 2004). For this reason, the name 'behaviourial economics' may be somewhat misleading; arguably, a label such as 'psychological economics' would be more appropriate.

treated as a purely positive science of human socio-economic behaviour and not as any sort of normative enterprise. We briefly review some other responses to behaviourism at the end of this paper.

We agree with one methodological concern voiced by Gul and Pesendorfer: the concern about the appropriate *level of explanation* in economics. Here, we suggest, Gul and Pesendorfer are right in criticizing the attempts of the most radical economic psychologists to reduce decision theory to neuro-physiology. But Gul and Pesendorfer draw the wrong conclusions from this. Far from supporting a 'mindless economics', rejecting the attempt to reduce economics to neuroscience is entirely consistent with accepting a mentalist appproach to economic theory. The failure to recognize this point may stem from a failure to distinguish clearly between the notions of 'mind' and 'brain'. The former is a 'macro-level', psychological notion, the latter a 'micro-level', physiological one. The most compelling forms of mentalism entail precisely the view that the study of rationality and action cannot be reduced to the neuro-physiological study of the brain and body.

The paper is structured as follows. In Section 2, we review and contextualize Gul and Pesendorfer's central claims. In Section 3, we identify four misconceptions underlying them. In Section 4, we introduce some key concepts from the philosophy of science, which help us clarify the difference between behaviourism and mentalism. In Section 5, we distinguish between two kinds of 'revealed preference' approaches to economic theory – an 'epistemological' and an 'ontological' one – and show that only the more radical and less plausible approach commits us to behaviourism. In Section 6, we state our argument for mentalism more positively. In Section 7, we argue that the difference between mentalism and behaviourism is not just a metaphysical matter but relevant to the practice of economics. In Section 8, we distinguish mentalism from, and argue against, the radical neuroeconomic view that socio-economic behaviour should be explained in terms of the relevant agents' brain processes, as distinct from their mental states. In Section 9, we conclude.

2 The case for mindless economics

Gul and Pesendorfer's paper, 'The case for mindless economics' (2008), provides a useful starting point for our discussion. The paper makes at least three claims about economic science (the positive rather than normative part of economics):

- The only *evidence* that should be used to test economic theories is evidence about people's choice behaviour.

- The *content* of any economic theory consists solely in its choice-behavioural implications; two theories that are choice-behaviourally equivalent should be seen as equivalent simpliciter.

- Any economic theory should take the *form* of a representation of choice behaviour, and that representation should ideally take the form of attributing to the agents the maximization of some objective function.

The first of these claims concerns the *evidential base* of a theory in economics, the second its *semantic content* or *meaning*, and the third the *methodology of theory construction*. In addition to making these positive claims, Gul and Pesendorfer also express scepticism towards any form of normative economics that goes beyond a very thin kind of 'revealed-preference Paretianism', i.e., the assessment of socio-economic institutions or outcomes in terms of whether they are Pareto efficient relative to people's revealed preferences. For present purposes, however, we set the case of normative economics aside.

In essence, Gul and Pesendorfer hold that (positive) economics should be the science of choice behaviour, and that its evidence base, ontology of the world, and formal structure should focus solely on people's observed or observable choices. Although they do not situate their views in the context of earlier behaviouristic schools of thought in psychology and related disciplines, Gul and Pesendorfer's approach to economics mirrors Pavlov's and Skinner's approaches to psychology and the Vienna Circle's approach to the philosophy of science and language. In fact, each of their central claims corresponds to a different historical variant of behaviourism (using the taxonomy in Graham 2010).

The first claim – about the evidence base of economics – broadly corresponds to 'psychological behaviourism', the view that human (and animal) behaviour should be explained solely on the basis of behavioural evidence, such as evidence about 'external physical stimuli, responses, learning histories, and (for certain types of behavior) reinforcements' (Graham 2010). If anything, the evidence accepted by those earlier psychological behaviourists is *less* restricted than that accepted by Gul and Pesendorfer.

The second claim – about the semantic content or meaning of any theory in economics – corresponds to 'analytical or logical behaviourism', the view associated with the Vienna Circle, Gilbert Ryle (1900-1976), and some of Ludwig Wittgenstein's (1889-1951) work that 'the very idea of a mental state or condition is the idea of a behavioral disposition or family of behavioral tendencies' (Graham 2010). Accordingly, '[w]hen we attribute a belief ... to someone, we are not saying that he or she is in a particular internal state or condition. Instead, we are characterizing the person in terms of what he or she might do in particular situations or environmental interactions' (ibid.).

Figure 1: Gul and Pesendorfer's claims and their precursors

	Gul & Pesendorfer's claims	**Historical precursors**
of a theory in...	**economics**	**psychology**
Evidence base	agents' choice behaviour	external physical stimuli, responses, learning histories, reinforcements 'psychological behaviourism' (Pavlov, Skinner)
Semantic content	choice-behavioural implications	behavioural dispositions or behavioural tendencies described by the theory 'analytical / logical behaviourism' (Vienna Circle, Ryle, Wittgenstein)
Methodological form	representation of choice behaviour, in terms of the maximization of an objective function	representation of behaviour, no modelling of internal information processing mechanisms 'methodological behaviourism' (Watson)

The third claim – about the methodology of theory construction in economics – is analogous to 'methodological behaviourism' in psychology in that it prescribes a focus on the representation of behaviour rather than the modelling of mental states and mental processes in theory construction. Historically, methodological behaviourism, as defended for instance by John Watson (1878-1958), is the view that 'psychology should concern itself with the behavior of organisms' and not 'with mental states or events or with constructing internal information processing accounts of behavior' (Graham 2010). Accordingly, 'reference to mental states, such as an animal's beliefs or desires, adds nothing to what psychology can and should understand about the sources of behavior' (ibid.), and so a theory's goal should simply be to represent behavioural patterns. Gul and Pesendorfer strengthen that demand by requiring that this representation take the form of attributing to the agent the maximization of some objective function.

Figure 1 summarizes the parallels between Gul and Pesendorfer's claims and their historical precursors in psychology and related disciplines. Given the extent to which Gul and Pesendorfer's claims mirror (and perhaps reinvent) earlier behaviouristic claims, one might ask whether their views suffer from the same problems that those earlier behaviourisms suffered from and which ultimately led to their demise.[5] In what follows,

[5]The parallels between the mentalism-behaviourism debate in psychology and the one in economics have received very little attention in the literature. For a brief historical sketch, unrelated to Gul and

we draw on insights gained from some of those other cases to see what lessons can be learnt for the case of economics.

3 Four misconceptions

We begin our defence of mentalism by arguing that Gul and Pesendorfer's three positive claims, like their historical precursors, rest on at least four misconceptions, which we will call the 'misconception of a fixed evidence base', the 'evidence/content conflation', the '"unobservable, therefore non-existent" fallacy', and the 'maximization dogma'.

3.1 The misconception of a fixed evidence base

In line with psychological behaviourism, Gul and Pesendorfer argue that the only evidence that should be used to test economic theories is evidence about people's choice behaviour. But there is no systematic reason why the evidence base of economics should be restricted in this way. Across the sciences, it is a common phenomenon that our available evidence base occasionally grows. Various things or phenomena that people could not observe in the past, and which earlier generations might have regarded as speculative, have eventually turned out to be observable, through the use of more advanced instruments, more creative experimental designs, and so on.

In physics, entities and phenomena such as the Higgs boson and various elementary particles, forces, and fields seemed at some point to be purely theoretical constructs, but are being increasingly turned into observable entities and phenomena – albeit indirectly observable ones – through the advances in sophistication in our instruments and experimental techniques. The advances in microscopy over the centuries are a perfect illustration of this point (on the lack of a static distinction between what is observable and what is not, see, e.g., Maxwell 1962 and Shapere 1982).

In short, the idea that the evidence base of a particular scientific discipline should be fixed once and for all lacks any justification, given the history of science and the experience of other scientific disciplines. Rather, the evidence base of any science is changeable and dynamic, and there is no reason why economics should be an exception. Accordingly, even if there was a period in the history of economics when people's choice behaviour was the only evidence used to test theories, there is no principled reason why other kinds of evidence – from people's verbal reports and communicative behaviour to physiological and neuroscientific evidence – could not also be relevant.

Pesendorfer, see Edwards (2008).

3.2 The evidence/content conflation

In line with analytical or logical behaviourism, Gul and Pesendorfer argue that the content of any economic theory consists solely in its choice-behavioural implications; two theories that are choice-behaviourally equivalent should be seen as equivalent simpliciter. But even if the *evidence base* of economic theories were restricted to observable choice behaviour alone – and, as we have seen, there is no principled reason why it should be – it would not follow that the *content* of any economic theory should consist solely in its choice-behavioural implications. Rather, the content of a theory can, and often does, go well beyond its evidence base. To see that this is not just an esoteric possibility, but the norm across many scientific disciplines, consider a few simple examples:

Archaeology and ancient history: The evidence base for theories in these subjects consists of various archaeological objects and artefacts found, for instance, in excavations. But the content of those theories goes well beyond these objects and artefacts. The content, ultimately, is the life, social organization, and culture of the ancient societies in question. The reason why we are interested in old pots, pans, and other broken items is not just that these objects are interesting in their own right, but that they tell us something we cannot directly observe: namely how people lived in the societies under investigation.

Paleobiology: A natural- rather than social-scientific discipline that illustrates our point is paleobiology. Here the evidence base consists of geological findings and fossils, but the aim of the discipline is to answer biological questions about the evolution of life and its underlying molecular-biological mechanisms. Again, the content of the relevant theories goes well beyond the evidence base.

The point of much of science is precisely to make creative use of what is observable in order to get a better understanding of certain phenomena that are not by themselves observable. Making sense of, and organizing, empirical regularities is just one aim – but not the only aim – of science. By organizing empirical regularities, we often find evidential support for the existence of certain hitherto unobserved aspects of reality.

3.3 The 'unobservable, therefore non-existent' fallacy

The next misconception is also relevant to Gul and Pesendorfer's logical or analytic claim that the content of any theory in economics consists solely of its choice-behavioural implications, and that two choice-behaviourally equivalent theories should be seen as equivalent simpliciter. One route by which one might arrive at this claim is the following. Suppose one accepts, as Gul and Pesendorfer do, that observations about people's

choice behaviour are the only evidence that we are entitled to use to test our economic theories. And suppose, further, one somehow accepts the principle that *anything that is not observable does not exist*. It then follows that we are not entitled to treat as 'real' or 'existent' any properties or entities in economics that go beyond what we can directly observe. And this, by stipulation, is people's choice behaviour alone.

But even if we were to suspend our criticism of the assumption that only choice behaviour is observable in economics, it should be obvious, as a matter of logic, that, from the fact that a particular entity or phenomenon is not observable, it does not follow that this entity or phenomenon does not exist. And the conclusion that the entity or phenomenon does not exist follows even less from the fact that something is not *currently* observable. Sometimes we can have strong indirect evidence for something, even though it is not directly observable.

Electrons and other elementary particles are not, strictly speaking, directly observable; we can only see their traces, for example, when they travel through a cloud chamber (as water vapour condenses upon the impact of ionizing particles). But few people would seriously doubt their existence.

'Occam's razor' principle tells us not to postulate the existence of any unnecessary entities. So, before we can hypothesize that something exists despite being unobservable, we need to come up with at least some indirect evidence for its existence. But if the best confirmed and most parsimonious theory of a particular phenomenon commits us to postulating an entity, then it is fully consistent with Occam's razor principle to accept its existence. The key idea behind the principle is that we should not postulate too many entities, but neither should we postulate too few.[6]

[6]We here accept that mental states are not directly observable, and similar in status to the unobservable entities and properties in other sciences. Hausman (1998) denies that the mental states posited in economics (e.g., the utility and subjective probability functions) are unobservables of the same kind as electrons or neutrinos, and argues instead that they should be seen as part of 'commonsense reality', like tables and chairs. This is because the functional role played by utilities and probabilities in economics is 'virtually identical' to that played by desires and beliefs in folk psychology, and the latter are already among our everyday ontological commitments. We accept the analogy between the mental states in economics and those in folk psychology and agree with Hausman that those mental states should be considered real. Yet, we think a further argument is needed to convince the skeptic that mental states in *both* folk psychology *and* economics can be seen as real, *despite their prima-facie unobservability (or at most indirect observability)*. Our argument in this paper is intended to fill this gap. Several contributions to the 'realism-antirealism' debate in economics (as reviewed, e.g., in Hausman 1998) either deny or do not develop the analogy between the mental states posited in economics and the unobservables posited in the natural sciences, and hence that debate is somewhat orthogonal to our concerns here.

3.4 The maximization dogma

Implicitly relying on a particularly strong version of methodological behaviourism, Gul and Pesendorfer suggest that any economic theory should take the form of a representation of choice behaviour, and that this representation should ideally take the form of attributing to the agents in question the maximization of some objective function. However, while it may be a useful *starting point* for the explanation of behaviour to search for some objective function a given agent maximizes, there is no principled reason why our best theories of economic behaviour should *necessarily* be based on the notion of maximization.

Which *form* of a theory best explains human behaviour is a contingent, empirical question, which can be settled only by actual scientific research, not by methodological stipulation. Just as it has turned out to be wrong – given Einstein's general theory of relativity – that space and time must necessarily be Euclidean (as Immanuel Kant, for example, assumed), so there is no *a priori* reason to think that the explanation of social and economic behaviour must necessarily be based on the maximization of a single objective function. For example, an empirically adequate theory might model agents as being governed by a more complex system of constraints.

Of course, current attempts to explain economic behaviour in a non-maximization-based way, such as theories of satisficing as introduced by Herbert Simon (1956) or theories of fast and frugal heuristics as proposed by Gerd Gigerenzer and others (e.g., 2000), remain controversial. But the mere fact that these are well-defined and eligible contenders for economic theories illustrates that the maximization of a single objective function is not the only form an economic explanation can take. The reason economists are divided over Simon's and Gigerenzer's theories is *not* that these theories have the wrong form *per se*, but rather that it is unclear whether they offer the best explanations of the empirical phenomena they are intended to explain.

4 A primer in the philosophy of science

We have identified four misconceptions underlying Gul and Pesendorfer's (and no doubt others') arguments for behaviourism in economics. To clarify the distinction between behaviourism and mentalism further, we need to introduce some key concepts from the philosophy of science: the concepts of (i) a 'theory', (ii) 'empirical adequacy' of a theory, (iii) an 'ontological commitment', and (iv) 'underdetermination of theory by evidence'.

4.1 What is a theory?

On the standard approach (which goes back to Karl Popper and Carl Gustav Hempel; see, e.g., Woodward 2009), a *theory* is a set of sentences (in some definitions, a set of propositions), which is ideally:

(i) closed under implication (so that the theory can be identified with the body of its implications), and

(ii) expressible as the set of implications of a finite (ideally small) set of basic principles or axioms (called the *theory formulation*), perhaps together with some auxiliary assumptions.

Newtonian physics is a paradigm example of a theory in this sense. Here, the theory formulation consists of Newton's three basic laws of motion and his law of universal gravitation, and the theory itself consists of all the implications of those basic principles. To arrive at a Newtonian theory of a specific physical system, such as the solar system, we further need to add some auxiliary assumptions, especially about the initial configuration of the relevant bodies (their masses, positions, and forces acting on them). The theory's predictions about the system's behaviour over time will then be among the relevant body of implications. There are also some alternative definitions of a theory in the literature (e.g., van Fraassen 1980), but for present purposes, the standard definition will suffice.[7]

4.2 When is a theory (empirically) adequate?

A theory – call it T – is said to be *adequate* with respect to a body of sentences S if and only if those sentences are among the theory's implications, formally if and only if T logically entails S. Usually, we are interested in a theory's adequacy with respect to the set of those sentences whose truth we can empirically observe (the so-called *observation sentences*). We then also speak of *empirical adequacy*. (To make the definition applicable in practice, some relevant auxiliary assumptions may typically need to be included in T.)

For example, Newtonian physics, together with some auxiliary assumptions, is at least approximately adequate with respect to the observation sentences about the motion of the planets around the sun, or about the way an apple falls from a tree. It is not

[7]The main rival to the standard, *syntactic* definition of a theory given here is a *semantic* definition (as exemplified by van Fraassen 1980), according to which a theory is a set of models (with a certain structure), rather than a set of sentences (with a certain structure). Many subtly different variants of each definition can be given. The details are not the focus of this paper.

adequate, on the other hand, with respect to a body of sentences about the behaviour of objects whose velocity is close to the speed of light, as Einstein famously pointed out.

Empirical adequacy – or at least approximate empirical adequacy (a notion that could be analyzed further) – is typically considered a *minimal* desideratum on a good scientific theory. Importantly, *empirical adequacy* of a theory is not the same as *truth* of that theory. Truth is a more demanding, and more elusive, notion. According to the *correspondence theory of truth*, a necessary condition for the truth of a theory is the existence of a suitable homomorphism (structure-preserving mapping) between the relevant properties of the world and the theory's representation of those properties. As we will discuss below, logically, there can exist two or more rival theories that are each empirically adequate with respect to a particular body of observations, but only one of which, at most, may be true.

4.3 What are the ontological commitments of a theory?

To define the notion of an ontological commitment of a theory, we first need to introduce a basic notion from formal logic: the notion of a *semantic interpretation* of the language in which the theory is expressed. This is

- an *assignment of truth-values* to all sentences in that language,

which, in turn, is based on

- a definition of one or several *domains of objects* (depending on how many types of objects the theory refers to),
- an *interpretation of all predicates, relations, and functions* that the theory uses, as predicates, relations, and functions over the relevant objects, and
- an *assignment of objects to all constant symbols* used by the theory.

We call a semantic interpretation that renders a given theory true (i.e., which assigns the truth-value 'true' to all sentences of the theory) a *model* of that theory. Any consistent theory has at least one model, and typically many. Each such model corresponds to one possible way the world could be – one possible world – *consistently with the theory.* The domain of objects (or family of domains) of that model then represents the objects that exist in that particular world, and the predicates, relations, and functions correspond to the properties of, and relations between, those objects.

Obviously, some models of a given theory may be 'sparser' – i.e., have smaller domains of objects and/or fewer properties of, and relations between, these objects – than

others. However, by considering *all* possible models of the theory (at most excluding certain 'trivial' or 'non-standard' models), we can ask which kinds of objects, properties, and relations are present in *all* of them. These can be seen as the objects, properties, and relations the theory is *minimally* committed to. We call them the *ontological commitments* of the theory. (We set aside a number of subtleties here, which have been discussed in detail by model theorists and logicians.)

This notion of an ontological commitment is very natural. Consider, for example, the theory of arithmetic as defined by the Peano axioms, which are the fundamental axioms of arithmetic. Any standard model of these axioms, however it is defined, will have a domain of objects with the formal properties of the natural numbers. Therefore – and as we would intuitively expect – the natural numbers are among the ontological commitments of Peano arithmetic.

Similarly, consider the standard theory of particle physics, which offers a unified theory of electromagnetic, weak, and strong nuclear interactions, while still leaving out gravity. Just as the natural numbers are a common presence in any model of Peano arithmetic, so certain kinds of elementary particles can be found in any non-trivial model of the standard theory of particle physics, such as quarks, leptons (of which electrons are special cases), and different kinds of bosons. Most of these have also been experimentally identified, using instruments such as the Large Hadron Collider at CERN, Switzerland, but at least until recently the Higgs boson remained empirically undiscovered. The theory has always been committed to its existence, however, since the theory could not be true without it.

The present notion of an ontological commitment is central to the so-called *naturalistic* attitude towards ontological questions we find in normal scientific practice (Quine 1948, Fine 1984, Musgrave 1989).[8] To figure out what entities, properties, and relations there are in any given area, according to this attitude, we should not engage in armchair metaphysical reasoning, but consult our best scientific theories of that area. Unless we have independent reasons to doubt those theories, we should take their ontological commitments at face value. If our best physical theories tell us that there are quarks, leptons, and bosons, we have every reason to believe in these particles' existence, regardless of their unobservable status.

4.4 Underdetermination of theory by evidence

Let S (a set of sentences) be our body of evidence – perhaps even the maximal body of evidence we could hypothetically obtain – and let T be the theory that we have come up

[8] This attitude underlies Quine's famous dictum '[t]o be is to be the value of a variable' (1948).

with. Even if the theory is adequate with respect to the evidence, the logical relationship between theory and evidence is typically a one-way, rather than two-way, relationship. The theory, T, entails the evidence, S, but not the other way round; S is certainly a subset of T (assuming adequacy), but T goes beyond S. In particular, T also has some unobservable implications.

The key lesson of this point is that, in principle but often also in practice, there can be two or more distinct theories that coincide in their observable implications (and therefore in their adequacy with respect to our evidence), but which are in fact logically incompatible with respect to some unobservable implications. In such a case, we say that our theory is *underdetermined by the evidence*. This problem was famously discussed by Quine (e.g., 1975; see also List 1999).

The simplest illustration of this problem in economics is given by the assignment of a von-Neumann-Morgenstern utility function to an agent. As is well known, there is not just one utility function that fits a given agent's choice behaviour, but an infinite number. The function is unique only up to positive affine transformations. Of course, in the present example, nothing much hinges on the properties of the function that are left underdetermined, such as whether the agent's utility in one situation is twice as large as that in another. Indeed, most economists would not consider such statements meaningful; they would regard the question of which *specific* von-Neumann-Morgenstern utility function (as opposed to which equivalence class) is the right one as indeterminate. The underdetermination problem would come to trouble us only if we wanted to use von-Neumann-Morgenstern utilities as the basis for interpersonal comparisons, something many economists would reject (for discussion, see List 2003).

Generally, however, the problem of underdetermination of theory by evidence raises important questions for the status of the unobservable implications of any theory and its ontological commitments. When a theory is underdetermined by the evidence (so that there could be a rival theory with different unobservable implications and different ontological commitments), we face the question of whether there is a fact of the matter about the theory's unobservables:

- If there is a fact of the matter, we have an instance of *mere underdetermination*. One of the theories is correct in its unobservable implications, including its ontological commitments; we just do not know which one it is. This is an *epistemological* problem.

- If there is no fact of the matter, we have an instance of *indeterminacy*. The theory's unobservable implications do not correspond to anything in the world.

> The theory's unobservables are at best useful fictions, at worst meaningless. This is an *ontological* problem.

The main insight to be gained from this philosophy-of-science primer, for present purposes, is that the question of what our evidence for a particular theory is and, more broadly, what the largest body of observation sentences could possibly be, is fundamentally distinct from, and not to be confused with, the question of what the theory's ontological commitments are.

5 Two kinds of 'revealed preference' approaches

We are now in a position to distinguish more clearly between two kinds of 'revealed preference' approaches to economic theory, and to see whether they commit us to behaviourism (for classical works on revealed preferences, see Samuelson 1938, Richter 1966, and Sen 1971). One kind of approach is defined in terms of an epistemological thesis, the other in terms of an ontological one. As we will see, only one of the two theses – arguably the less plausible one – is genuinely behaviouristic, while the other is fully compatible with mentalism.[9]

An epistemological 'revealed preference' thesis: Our body of evidence for a theory in economics – the set of observation sentences – is restricted to agents' choice behaviour.

An ontological 'revealed preference' thesis: The ontological commitments of any theory in economics – or at least those ontological commitments that we are entitled to take seriously – are restricted to agents' choices and choice dispositions and therefore exclude mental states.

First consider the epistemological thesis. Although we have already disputed that the evidence base should be fixed as stated by that thesis, some economists might still accept it for stipulative reasons: they might stipulate that what demarcates economics from neighbouring disciplines such as psychology is its reliance on choice-behavioural evidence, rather than richer psychological evidence. This justification for the epistemological thesis may seem *ad hoc*, but it is not incoherent.

[9]The taxonomy of different kinds of psychological behaviourism in Moore (2001) also suggests that some more modest, 'methodological' (as opposed to 'radical') forms of behaviourism are compatible with mentalism.

The ontological thesis, by contrast, is harder to defend. At least in the technical sense explained above, the ontological commitments of standard economic theory simply include certain mental states such as preferences and/or beliefs. As soon as the theory refers to an agent's preference relation, utility function, or subjective probability function, any model of it will have such relations or functions among its structural elements, and so they will be among the theory's ontological commitments.

One might object that this shows only that certain mathematical functions or structures are among the theory's ontological commitments, not that these need to be interpreted as mental states; they could simply be viewed as abstract constructs, without any psychological interpretation. This objection, however, overlooks the insights of functionalism in the philosophy of mind (e.g., Block 1980). *Functionalism* is the widely accepted view that what makes a given property or relation a mental state (such as a belief or a preference) is precisely that it plays a particular functional role for the agent. Beliefs, for example, are those properties or relations that play the role of representing certain features of the world from the agent's perspective, and preferences are those properties or relations that play the role of directing the agent's actions (see, e.g., List and Pettit 2011, ch. 1). Standard economic models of individual decision making do include such properties or relations (for example, binary relations or real-valued functions playing the role of preferences and probability functions playing the role of beliefs), and so mental states, in functionalist terms, are present in them.[10]

Furthermore, standard economic theory has these ontological commitments even if we use only choice-behavioural evidence to establish its adequacy. This shows that the epistemological 'revealed preference' thesis does not imply the ontological one, and thus that the epistemological thesis is compatible with economic theory's commitment to underlying mental states. Indeed, the etymology of the term 'revealed preferences' suggests just this: an agent's behaviour 'reveals' – is evidence for – something other than behaviour, namely the agent's mental state – his or her preferences – which causes the behaviour in question.

[10]We here set aside the complicated and subtle philosophical debate about the relationship between the *functional role* of mental states (such as the representational or action-directing roles of beliefs and preferences) and their *phenomenal or conscious character* (roughly speaking, what it subjectively feels like to have such mental states). Some philosophers – especially dualists – hold that certain mental states have phenomenal (conscious) aspects *above and beyond* their functional aspects, while others consider these phenomenal aspects a *by-product* (in a sense that requires further analysis) of their functional aspects. (Yet others deny the existence of phenomenal aspects altogether.) For an overview of the debate, see Chalmers (2010). In this paper, we focus only on the functional aspects of mental states such as beliefs and preferences, which are most relevant for micro-economic theory.

Figure 2: Possible views about 'revealed preferences'

Evidence **Ontological status of ascribed mental states**	**Restricted to choice behaviour and dispositions**	**Not so restricted; other psychological evidence admissible**
Not real, but mere theoretical constructs	Radical behaviourism	Incoherent position
Real, to be taken seriously	Mentalism with narrow evidence base ('epistemic behaviourism')	Mentalism with broad evidence base

In sum, behaviourists and mentalists are divided on two questions: first, whether or not the evidence base of economics should be restricted to choice behaviour and choice dispositions, and secondly, whether the mental states ascribed by economic theories should be treated as mere theoretical constructs or as corresponding to real phenomena. Of course, taking the ascribed mental states to correspond to real phenomena is fully consistent with acknowleding that they depict these phenomena in very simplified or idealized ways, just as a physical theory's depiction of a planet or a volcano greatly simplifies or idealizes the details of the real planet or volcano it refers to. Recall that the relationship between the world – or the relevant target set of properties in the world – and their representation by a theory is at best a homomorphic one, which preserves certain key structural features but which still abstracts away from many substantive details. Figure 2 shows the different possible views.[11]

6 An argument for mentalism

Our objections to the radical behaviourist view should already be evident from our discussion. We now wish to state our argument for mentalism more positively. Recall that a radical behaviourist holds the view that even if certain mental states (or the relations or functions playing their role) are technically among economic theory's ontological commitments, they are still nothing more than theoretical constructs: they may be in-

[11]The distinction between radical behaviourism and mentalism with a narrow evidence base is similar to Cozic's (2012) distinction between a stronger and a weaker sense in which conventional models of choice can be 'cognitively mute'.

strumentally useful for making sense of behavioural regularities, but they should not be seen as corresponding to anything real. As our philosophy-of-science primer should indicate, however, this view misses the central idea underlying the naturalistic attitude towards ontological questions. When something – whether an eletron or a mental state – is an ontological commitment of a theory, then one's acceptance of that theory directly commits one to accepting the existence of the given entity or property. To ask whether that entity or property 'really' exists, after it has been established as one of the theory's ontological commitments, is to ask one question too many; or alternatively, it is to express doubts about the theory itself.

The naturalistic argument for mentalism in economics can be summarized as follows:[12]

Premise 1: Some mental states, such as beliefs and preferences, are technically among the ontological commitments of our current best theories of economic decision making.

Premise 2: In any normal science, the criterion for whether a theoretically postulated entity, property, or relation is to be treated as corresponding to a real entity, property, or relation in the world is whether it is among the ontological commitments of our current best theory or theories in the relevant area (assuming we have no special reasons to doubt those theories themselves).

Premise 3: Economics is a normal science.

Conclusion: The mental states that our best economic theories ascribe to economic agents are to be treated as corresponding to real phenomena (unless we have special reasons to doubt those theories themselves).

The argument is clearly valid (i.e., the premises logically entail the conclusion). Whether the argument is also sound depends on whether the premises are all true. Given the nature of practically all our current (micro-)economic theories, ranging from classical rational choice theory to more recent psychologically oriented theories (e.g., Camerer, Loewenstein, and Rabin 2004), Premise 1 is true in light of the technical definition of an ontological commitment and – for present purposes – the functionalist definition of a mental state. Premise 2 is also true, since it states a basic principle underlying standard scientific practice, namely the naturalistic ontological attitude. Premise 3 is a claim that critics of economics might wish to challenge, but scientifically minded economists are unlikely to object to it.

[12]Some of the philosophical ideas underlying this naturalistic argument are developed in List (2011).

Consequently, the only way to avoid the mentalistic conclusion would be to insist on having special doubts about our economic theories themselves, despite their status as our current best scientific theories in the relevant area. But those asserting such doubts would then have to explain what evidence underpins them. We suspect that few economists would wish to make their argument against mentalism dependent on a rejection of the adequacy of our best economic theories themselves. We conclude that just as we have strong *prima-facie* reasons to accept the reality of quarks, leptons, and bosons in particle physics, so we have strong *prima-facie* reasons to accept the reality of mental states in economics.

It is worth clarifying how this conclusion differs from the view held by radical behaviourists. We are not suggesting that radical behaviourists such as Gul and Pesendorfer will deny the reality of mental states when they take off their 'hats' as professional economists and adopt a commonsense view of the world, for instance while interacting with other people in their day-to-day lives. What they are committed to denying is that mental states should be part of the ontology of economics.

7 Does the difference between mentalism and behaviourism matter?

One might think that the difference between mentalism and behaviourism is a purely metaphysical matter, which is of little significance for the practice of economics itself. But this impression is misleading. That the difference matters also in practice can be seen by revisiting the empirical underdetermination problem, the problem that there can exist two or more distinct theories that are empirically equivalent but logically incompatible.

First consider the idealized limiting case of no underdetermination. Take a simple choice problem without risk or uncertainty, where an agent has perfectly well-behaved choice dispositions over some options, satisfying all the standard rationality conditions. The agent's choice dispositions – formally represented by a choice function – can then be uniquely rationalized by a preference ordering over the given options (e.g., Sen 1971, Bossert and Suzumura 2010) (note that the conditions for achieving such a unique rationalization are demanding). Although this rationalization involves a mental-state ascription – namely the ascription of a preference ordering, i.e., a binary relation that plays the role of a mental state – preference orderings and choice functions stand in a one-to-one correspondence in this case. As long as rationalization of choices is required to take the form of ascribing to the agent a weak ordering, there is no underdetermina-

tion of preferences by choice dispositions here: there exists one and only one preference ordering that entails the given choice dispositions.[13] Consequently, there are no logical implications of the mental state ascription that go beyond what is already encoded in the choice function itself, and no issues of indeterminacy arise: there are behaviourally observable facts about everything the theory says. Hence, one might think that the question of what the ontological status of the agent's preferences is, over and above his or her choice dispositions, is primarily metaphysical.

Now, however, consider a less idealized case. A much-discussed example is due to Amartya Sen (1993).

The polite dinner-party guest: Given a choice between a large, a medium-sized, and a small apple, a dinner-party guest (who at home would choose larger apples over smaller ones) chooses the medium-sized apple (for politeness). If the large apple is no longer available while the medium-sized and small ones are, the guest chooses the small apple (again for politeness).

The agent's choice function violates contraction consistency and cannot be rationalized by a preference ordering over apples. But it would be a bad explanation to suggest that the agent is irrational; this explanation would violate the *principle of charity* in interpretation (see, e.g., Davidson 1973). Rather, the agent is motivated by considerations over and above the sizes of the apples. However, *if* the agent's choice behaviour is the only evidence we can go by – for example, we cannot ask the agent any questions about the reasons for his or her choices – then we face an underdetermination problem. Several distinct hypotheses entail the same choice behaviour, ranging from the hypothesis that the agent has complicated (and perhaps 'non-consequentialist') preferences over 'extended alternatives' (object-context pairs) to the hypothesis that he or she is governed by various norms of politeness, approval- or esteem-seeking, or other social constraints (e.g., Bhattacharyya, Pattanaik, and Xu 2011; Bossert and Suzumura 2009; Suzumura and Xu 2001; Brennan and Pettit 2005). The agent's choice dispositions alone are insufficient to distinguish between these (and other) rival explanations.

Does this mean that there is no fact of the matter as to what the correct explanation is? Both our psychological understanding and the practices of other cognitive and behavioural sciences suggest that there *can* be a real difference between different rival explanations, despite their choice-behavioural equivalence. In addition to attributing

[13]Note that if we lift the requirement that rationalization take the form of the ascription of a weak ordering to the agent, and allow other forms of rationalization (e.g., in terms of other mathematical structures), then the underdetermination problem can arise even in the present case of choice without risk or uncertainty.

different internal cognitive mechanisms to the agent (as well as different first-person experiences, which would lead us to predict different introspective reports from him or her, if we could elicit a truthful response), they may also have different repercussions further down the line. Only some but not all explanations may cohere with our explanations of other related phenomena, so that good scientific practice would give us a coherence-based criterion for choosing some explanations over others.

Setting dogma aside, the natural view is that although choice-behavioural evidence often underdetermines our theoretical explanation of people's choices, a suitably broadened evidence base may allow us to distinguish between different rival hypotheses. Such a broadened evidence base might include evidence about other related social phenomena, different kinds of psychological data, verbal reports, and occasionally (for plausibility checks) even introspection. In short, the availability of different choice-behaviourally equivalent explanations does not imply that there is no fact of the matter as to what the real reasons for an agent's choices are.

Wakker (2010, p. 3, drawing on Harré 1970) distinguishes between *paramorphic* and *homeomorphic* models of decision making. A *paramorphic* model 'describes the empirical phenomena of interest correctly, but the processes underlying the empirical phenomena are not matched by processes in the model'. A *homeomorphic* model, by contrast, has the property that 'not only its empirical phenomena match reality, but also its underlying processes do so'. In outlining a research programme for decision theory, he suggests that we should aim to arrive at homeomorphic models and that this is what prospect theory seeks to do: 'Not only [should] the decisions predicted by the model match the decisions observed, but we also want the theoretical parameters in the model to have plausible psychological interpretations'.[14]

Sharing this goal, several recent works in decision theory emphasize the importance of 'reasons for choice' or 'psychological states' over and above the choice behaviour induced by them. Some of these works explicitly employ mentalist terminology, such as 'epistemic states', 'knowledge', and 'beliefs' in epistemic game theory (e.g., Aumann and Brandenburger 1995); 'belief-dependent emotions' in psychological games (Geanakoplos

[14]Wakker (2010, p. 3) also stresses that the evidence base and domain of economic explanations should not be considered fixed: '[Milton] Friedman's arguments in favor of paramorphic models are legitimate if all that is desired is to explain and predict a prespecified and limited domain of phenomena. It is, however, usually desirable if concepts are broadly applicable, also for future and as yet unforeseen developments in research. Homeomorphic models are best suited for this purpose. In recent years, economics has been opening up to introspective and neuro-imaging data. It is to be expected that the concepts of prospect theory, in view of their sound psychological basis, will be well suited for such future developments and for connections with such domains of research.'

and Pearce 1989); 'emotions' such as 'anger' or 'fear' (Elster 1998, Loewenstein 2000); 'thinking' and 'feeling' (Romer 2000); 'intrinsic' and 'extrinsic motivations', 'ego boosting' and 'ego bashing' (Bénabou and Tirole 2003); 'rationales' (Manzini and Mariotti 2007, Cherepanov, Feddersen, and Sandroni 2008); 'moods' and 'mindsets' (Manzini and Mariotti 2012); 'motivating reasons' and 'weighing of reasons' (Dietrich and List 2012a,b); 'experiences' (Dietrich 2012); and 'the minds of checklist users' (Mandler, Manzini, and Mariotti 2012).

In sum, since choice behaviour routinely underdetermines its theoretical explanation, good scientific practice requires us to consider all the different rival explanations and then creatively to identify an enriched evidence base, and more advanced empirical designs, to determine which explanation is most adequate – in particular, which is most homeomorphic and not merely paramorphic. Even if we fail to find a purely empirical criterion for picking out a unique correct theory, Occam's razor principle would tell us to choose a theory which is ontologically not too rich, but also not too sparse, to explain our observations parsimoniously.

8 Can economics be reduced to neuroscience?

Many neuroscientists hope to dispense with traditional psychological theories by explaining psychological phenomena in terms of neurophysiological processes in the brain (for a recent debate, see Bennett et al. 2007). Similarly, some of the most radical neuroeconomists hope to dispense with traditional economic theories by explaining economic behaviour in terms of the relevant agents' brain processes (for discussion, see Camerer, Loewenstein, and Prelec 2005). At first sight, one might think that scientific progress is inexorably headed in this direction, and many advances in science seem to confirm this picture. We are developing a better understanding of the 'micro-level' mechanisms underlying many 'macro-level' phenomena, for instance the biochemical mechanisms ('micro') underlying the functioning of cells ('macro'), the cellular mechanisms ('micro') underlying the life of organisms ('macro'), and the individual-level mechanisms ('micro') underlying larger social processes ('macro'). The search for micro-foundations of macroscopic phenomena, with a view to replacing less fundamental theories with more fundamental ones, seems *en vogue*.

Yet, there is a common misconception underlying many of these attempts at theory reduction. The misconception can be termed the 'supervenience implies explanatory reducibility' fallacy. To explain this fallacy, let us consider a familiar argument for theory reduction. Its (correct) premise is that the world is fundamentally made up of

elementary particles, atoms, and molecules, which stand in various physical and chemical relations to each other and whose interaction underlies all more complex phenomena, including the functioning and behaviour of organisms. More formally:

The supervenience thesis: The totality of 'micro-level', physical facts about the world determines all 'macro-level' facts, such as facts about organisms and their behaviour.

It is then argued that, because everything in the world 'supervenes' on the physical, the best explanation of any phenomenon must also be a physical one.

The explanatory-reducibility thesis: Any phenomenon in the world can and should ideally be explained in terms of underlying physical mechanisms. Any non-physical explanations – such as psychological or social explanations – are at best provisional and reflect a lack of understanding of underlying mechanisms.

The claim that psychology can be reduced to neuroscience is sometimes defended in just this way. Psychological phenomena are surely the result of underlying neurophysiological brain processes, and 'so', the reasoning goes, our most fundamental explanations of them should also be given at the neurophysiological level.

But does supervenience really imply explanatory reducibility? A large body of work in philosophy challenges this view, beginning with Jerry Fodor's (1974) and Hilary Putnam's (1975) classic arguments that the sheer combinatorial complexity of the relationship between the physical states of a person's brain and the psychological states of his or her mind rules out the effective reducibility of psychological 'natural kinds' (which are the relata of regularities that we are interested in) to purely neurophysiological ones.[15] What makes 'macro-level' mental states, such as beliefs and desires, more explanatorily useful than 'micro-level' patterns of neural activity is precisely that they abstract away from a large number of physical details that are irrelevant, and even detrimental, to the explanatory purposes at hand. Supervenience, in short, does not imply explanatory reducibility (for a recent defence of this anti-reductionistic view, see List and Menzies 2009; for a related discussion in the philosophy of social science, see List and Spiekermann 2012).

Consider, for example, how you would explain a cat's appearance in the kitchen when the owner is preparing some food. You could either try (and in reality fail) to

[15] More technically, the inverse image, with respect to the relevant supervenience function from physical brain states to psychological states, of any set of psychological states forming a 'natural kind' at the psychological level need not be a set of physical brain states forming a 'natural kind' at the physical level.

understand the cat's neurophysiological processes which begin with (i) some sensory stimuli, then (ii) trigger some complicated neural responses, and finally (iii) activate the cat's muscles so as to put it on a trajectory towards the kitchen. Or you could ascribe to the cat (i) the belief that there is food available in the kitchen, and (ii) the desire to eat, so that (iii) it is rational for the cat to go to the kitchen. It should be evident that the second explantion is both simpler and more illuminating, offering much greater predictive power. The belief-desire explanation can easily be adjusted, for example, if conditions change. If you give the cat some visible or smellable evidence that food will be available in the living room rather than the kitchen, you can predict that it will update its beliefs and go to the living room instead. By contrast, one cannot even begin to imagine the informational overload that would be involved in adjusting the neurophysiological explanation to accommodate this change.

Good explanations – ones that are parsimonious and predictively successful – should identify the most functionally relevant regularities, while leaving out extraneous details. Functionally relevant regularities, in turn, need not be found at the most fine-grained level of description. It is an empirical question at which level of description any given system exhibits the most tractable regularities. There is no reason, for example, why a good theory of forest ecology should refer to quantum-mechanical effects inside the individual atoms in each tree. Similarly, if you want to explain why Microsoft Windows crashes if you install a particular software package, you should first look at possible programme errors or incorrect system parameters before trying to give a detailed account of the flow of individual electrons in the computer's micro-processor and memory chips.

As Daniel Dennett (1987) has argued, we explain the behaviour of certain organisms in terms of their mental states and not in terms of complicated physical processes – thereby taking an 'intentional' rather than 'physical stance' – precisely because this is the level of explanation most suited for the explanatory purpose at hand. A doctor who wishes to treat a brain hemorrhage or a tumor may well take a physical stance towards the patient, at least during the medical intervention, but it is far from clear how much economists can gain from trying to explain socio-economic behaviour by looking at people's brains, rather than interpreting their minds.

All of this is consistent, of course, with the idea of enriching the evidence base of economics when this helps us to distinguish between different rival theories, and this could certainly include some neuroeconomic evidence. But it should be clear that neither the focus on behaviour alone, nor the focus on brain physiology alone, can deliver satisfactory economic theories.

9 Concluding remarks

We have offered an argument for mentalism, and against behaviourism, in economics. We have not only responded to the central epistemological and ontological claims made by behaviourists, but also distinguished mentalism from the more radical neuroeconomic view that economic behaviour should be explained in terms of the relevant agents' brain processes, as distinct from their mental states. Gul and Pesendorfer (2008) seem to miss this distinction, frequently equating the mental with the neural and treating what might charitably be understood as a case for a 'brainless economics' (i.e., for an economic science separate from, and not reducible to, neuroscience) as a case for a 'mindless economics' instead (i.e., for an economic science free from mental-state ascriptions).

Our present critique of behaviourism differs from other, more familiar critiques of behaviourism and 'revealed preference' approaches (see, among many others, Hausman 2000 and Kőszegi and Rabin 2007). The behaviouristic account of preferences (and other mental states such as beliefs) is often criticized for what it fails to deliver: (i) it fails to say anything about human psychology and motivation, from which it is explicitly disconnected; (ii) it fails to provide adequate foundations for normative economics, as it gives at most an impoverished account of human well-being, says nothing about fundamental desires and needs, and renders interpersonal comparisons of utility impossible (all of which may matter for policy-making); and (iii) it fails to 'explain' behaviour in a non-circular way, since behaviour is 'explained' by preferences (or other attributes) that are in turn defined in terms of behaviour.

While such arguments are important and can be (indeed have been) made, we have taken a different approach here. Those earlier arguments construe economics as a discipline that should deliver more than a theory of choice (providing an account of, e.g., some psychological features of agents, normatively relevant features beyond revealed preferences, or non-circular explanations of choice). This premise is not shared by those economists who, when pressed, are prepared to 'define' (micro-)economics as a science of choice behaviour. Such a science should be as free as possible from normative assumptions and play no 'therapeutic' role, in Gul and Pesendorfer's terms. Critics of behaviourism who presuppose a broader definition of the discipline have little hope of convincing those who endorse the narrow, choice-centered definition. By contrast, our critique should convince also those who view economics as a science of choice behaviour alone, devoid of any further psychological or normative goals. Our naturalistic argument shows that even if one is not interested in mental states *as such*, one's theory of choice may well have to take them on board. A theory *of choice* may have to be a theory *about more than choice*.

10 References

Aumann, R., and Brandenburger, A. (1995) 'Epistemic Conditions for Nash Equilibrium', *Econometrica* 63(5): 1161-1180.

Bénabou, R., and J. Tirole (2003) 'Intrinsic and Extrinsic Motivation', Review of Economic Studies 70, 489-520.

Bennett, M., D. Dennett, P. Hacker, and J. Searle (2007) *Neuroscience and Philosophy*, New York (Columbia University Press).

Bhattacharyya, A., P. K. Pattanaik, and Y. Xu (2011) 'Choice, Internal Consistency and Rationality', *Economics and Philosophy* 27(2): 123-149.

Block, N. (1980) 'What is Functionalism?' in *Readings in Philosophy of Psychology*, vol. 1, Cambridge/MA (Harvard University Press), pp. 171-184.

Bossert, W., and K. Suzumura (2009) 'External Norms and Rationality of Choice', *Economics and Philosophy* 25: 139-152.

Bossert, W., and K. Suzumura (2010) *Consistency, Choice, and Rationality*, Cambridge/MA (Harvard University Press).

Brennan, G., and P. Pettit (2005) *The Economy of Esteem*, Oxford (Oxford University Press).

Camerer, C. F., G. Loewenstein, and D. Prelec (2005) 'Neuroeconomics: How Neuroscience Can Inform Economics', *Journal of Economic Literature* 43(1): 9-64.

Camerer, C. F., G. Loewenstein, and M. Rabin (2004) *Advances in Behavioral Economics*, Princeton (Princeton University Press).

Caplin, A., and A. Schotter (eds.) (2008), *The Foundations of Positive and Normative Economics*, Oxford/New York (Oxford University Press).

Chalmers, D. (2010), *The Character of Consciousness*, Oxford (Oxford University Press).

Cherepanov, V., T. Feddersen, and A. Sandroni (2008) 'Rationalization', working paper, University of Pennsylvania.

Chomsky, N. (1959) 'A Review of B. F. Skinner's *Verbal Behavior*', *Language* 35(1): 26-58.

Conradt, L., and C. List (eds.) (2009) 'Group decision making in humans and animals', theme issue of *Philosophical Tansactions of the Royal Society* B 364: 717-852.

Cozic, M. (2012) 'Economie "sans esprit" et données cognitives', working paper, Institut d'Histoire et de Philosophie des Sciences et des Techniques, Paris.

Davidson, D. (1973) 'Radical Interpretation', *Dialectica* 27(3-4): 313-328.

Dennett, D. (1987) *The Intentional Stance*, Cambridge/MA (MIT Press).

Dietrich, F. (2012) 'Modelling change in individual characteristics: an axiomatic framework', *Games and Economic Behavior* (in press)

Dietrich, F., and C. List (2012a) 'A reason-based theory of rational choice', *Nous* (in press).

Dietrich, F., and C. List (2012b) 'Where do preferences come from?', *International Journal of Game Theory* (in press).

Edwards, J. M. (2008) 'On Behaviorism, Introspection, Psychology and Economics', working paper, University of Paris 1, Panthéon-Sorbonne.

Elster, J. (1998) 'Emotions and Economic Theory', *Journal of Economic Literature* 36(1): 47-74.

Fine, A. (1984) 'The Natural Ontological Attitude', in J. Leplin (ed.), *Philosophy of Science*, Berkeley (University of California Press).

Fodor, J. A. (1974) 'Special sciences (or: The disunity of science as a working hypothesis)', *Synthese* 28(2): 97-115.

Fodor, J. A. (1975) *The Language of Thought*, Cambridge/MA (Harvard University Press).

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

Gigerenzer, G., P. M. Todd, and the ABC Research Group (2000) *Simple Heuristics That Make Us Smart*, New York (Oxford University Press).

Graham, G. (2010) 'Behaviorism', in E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Fall 2010 Edition), available at: http://plato.stanford.edu/archives/fall2010/entries/behaviorism/

Gul, F., and W. Pesendorfer (2008) 'The Case for Mindless Economics', in A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics*, Oxford/New York (Oxford University Press), pp. 3-39.

Harré, R. (1970) *The principles of scientific thinking*, London (Macmillan).

Harrison, G. W. (2008) Neuroeconomics: A Critical Reconsideration, *Economics and Philosophy* 24(3): 303-344.

Hausman, D. (1998) 'Problems with Realism in Economics', *Economics and Philosophy* 14: 185-213.

Hausman, D. (2000) 'Revealed Preference, Belief, and Game Theory', *Economics and Philosophy* 16: 99-115.

Hausman, D. (2008) 'Mindless or Mindful Economics: A Methodological Evaluation', in A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics*, Oxford/New York (Oxford University Press), pp. 125-151.

Katz, J. J. (1964) 'Mentalism in Linguistics', *Language* 40(2): 124-137.

Kőszegi, B., and M. Rabin (2007) 'Mistakes in Choice-Based Welfare Analysis', *American Economic Review* 97(2): 477-481.

Langendoen, D. T. (1998) 'Bloomfield', in R. A. Wilson and F. C. Keil (eds.), *The MIT Encyclopedia of Cognitive Science*, Cambridge/MA (MIT Press), pp. 90-91.

List, C. (1999) 'Craig's Theorem and the Empirical Underdetermination Thesis Reassessed', *Disputatio* 7: 28-39.

List, C. (2003) 'Are Interpersonal Comparisons of Utility Indeterminate', *Erkenntnis* 58: 229-260.

List, C., and P. Menzies (2009) 'Non-reductive physicalism and the limits of the exclusion principle', *Journal of Philosophy* CVI (9): 475-502.

List, C. (2011) 'Free will, determinism, and the possibility of doing otherwise', *Nous* (forthcoming).

List, C., and P. Pettit (2011) *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford (Oxford University Press).

List, C., and K. Spiekermann (2012) 'Methodological Individualism and Holism in Political Science: A Reconciliation', working paper, London School of Economics, available at: http://personal.lse.ac.uk/list/PDF-files/IndividualismHolism.pdf

Loewenstein, G. (2000) 'Emotions in Economic Theory and Economic Behavior', *American Economic Review* 90(2): 426-432.

Mandler, M., P. Manzini, and M. Mariotti (2012) 'A million answers to twenty questions: Choosing by checklist', *Journal of Economic Theory* 147: 71-92.

Manzini, P., and M. Mariotti (2007) 'Sequentially Rationalizable Choice', *American Economic Review* 97(5): 1824-1839.

Manzini, P., and M. Mariotti (2012) 'Moody choice', working paper, University of St Andrews.

Maxwell, G. (1962) 'On the Ontological Status of Theoretical Entities', in H. Feigl and G. Maxwell (eds.), *Scientific Explanation, Space, and Time; Minnesota Studies in the Philosophy of Science*, Volume III, Minneapolis (University of Minnesota Press).

Mongin, P. (2011) 'La théorie de la décision et la psychologie du sens commun', *Social Science Information* 50(3-4): 351-374.

Moore, J. (2001) 'On Distinguishing Methodological from Radical Behaviorism', *European Journal of Behavior Analysis* 2: 221-244.

Musgrave, A. (1989) 'Noa's Ark – Fine for Realism', *The Philosophical Quarterly* 39(157): 383-398.

Pettit, P. (1991) 'Decision Theory and Folk Psychology', in M. Bacharach and S. Hur-

ley (eds.), *Foundations of Decision Theory: Issues and Advances*, Oxford (Blackwell), pp. 147-175.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*, New York (William Morrow).

Putnam, H. (1975) 'Philosophy and our mental life', in *Mind, Language and Reality*, Cambridge (Cambridge University Press).

Quine, W. V. (1948) 'On What There Is', *Review of Metaphysics* 2: 21-38.

Quine, W. V. (1960) *Word and Object*, Cambridge/MA (MIT Press).

Quine, W. V. (1975) 'Empirically Equivalent Systems of the World', *Erkenntnis* 9: 313-328.

van Fraassen, B. C. (1980) *The Scientific Image*, Oxford (Oxford University Press).

Richter, M. K. (1966) 'Revealed Preference Theory', *Econometrica* 34(3): 635-645.

Romer, P. M. (2000) 'Thinking and Feeling' *American Economic Review* 90(2): 439-443.

Samuelson, P. (1938) 'A Note on the Pure Theory of Consumer's Behaviour', *Economica* (New Series) 5(17): 61-71.

Sen, A. K. (1971) 'Choice Functions and Revealed Preference', *Review of Economic Studies* 38(3): 307-317.

Sen, A. K. (1993) 'Internal Consistency of Choice', *Econometrica* 61(3): 495-521.

Shapere, D. (1982) 'The Concept of Observation in Science and Philosophy', *Philosophy of Science* 49(4): 485-525.

Simon, H. A. (1956) 'Rational choice and the structure of the environment', *Psychological Review* 63(2): 129-138.

Suzumura, K., and Y. Xu (2001) 'Characterizations of Consequentialism and Non-consequentialism', *Journal of Economic Theory* 101(2): 423-436.

Tomasello, M. (1995) 'Language is Not an Instinct', *Cognitive Development* 10: 131-156.

Wakker, P. (2010) *Prospect Theory: For Risk and Ambiguity*, Cambridge (Cambridge University Press).

Woodward, J. (2011) 'Scientific Explanation', in E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), available at: http://plato.stanford.edu/archives/win2011/entries/scientific-explanation/