Disengaging liberalism from rationality

Robert Sugden

r.sugden@uea.ac.uuk

School of Economics
University of East Anglia
Norwich NR4 7TJ
United Kingdom

Brief biography

Robert Sugden is Professor of Economics at the University of East Anglia in the UK. From the 1970s, he has used theoretical, experimental and philosophical methods to investigate issues in normative economics, behavioural 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.

Abstract

From Adam Smith onwards, a liberal tradition of economics has described the market system as both wealth-creating and liberty-enhancing. Modern economics has formalized this description in 'neoclassical' models of markets populated by rational agents. However, there is growing evidence that individuals' decisions often reveal inconsistent preferences. I review three influential books that present this evidence as a challenge to liberal justifications for the market. In response, I argue that individuals, whether neoclassically rational or not, can value the market as an institution that allows them to get whatever they want and are willing to pay for. This justificatory strategy illustrates how liberalism might benefit by disengaging itself from rationality.

From the very beginnings of economics as a discipline, a central liberal tradition has presented the market system as a mechanism that not only creates wealth but also promotes individual liberty. The first of these claims is implicit in the title of Adam Smith's *Wealth of Nations*. Smith shows how the market allows the division of labour, and how this creates the wealth that is visible even in the woollen coat of the Scottish day-labourer and in the glass window of his cottage. But he also describes the market system as 'the obvious and simple system of natural liberty' in which 'every man, as long as he does not violate the laws of justice, is left perfectly free to pursue his own interest in his own way'. For Smith, there is a fundamental connection between these two valuable properties: each individual's pursuit of his own interest in his own way is the motive power for the wealth-creating tendencies of the market system. In Smith's famous example, a merchant pursues his own interest in choosing to invest his capital where it will create most value, but, as if 'led by an invisible hand', his actions nevertheless promote the interests of society (Smith [1776] 1976, 22–23, 687, 456).

Over the course of the first half of the twentieth century, these ideas were formalized in terms of equilibrium properties of 'neoclassical' models of economies populated by maximizing agents. That firms were assumed to maximize profit was a natural updating of Smith's depiction of merchants. However, attributing maximization to the economic behaviour of private individuals or households was more questionable. Maximization of what, and why? In the first years of neoclassical economics, it was common to assume that individuals maximized 'utility', interpreted as a measure of hedonic satisfaction. But, in a move initiated by Vilfredo Pareto and grounded in scepticism about the cardinal measurability and interpersonal comparability of utility, twentieth-century economists converged on the assumption that each private agent acts on a stable and internally consistent preference ordering over the set of relevant economic outcomes; the agent's 'utility' was reinterpreted as a numerical representation of that ordering. Fundamental assumptions about preferences were justified on the grounds that they were requirements of individual rationality.

Within this theoretical framework, the wealth-creating properties of markets are represented in their most abstract form by the two Fundamental Theorems of welfare economics. An outcome of an economic system is *Pareto efficient* if, relative to that outcome, there is no way of making one individual better off (in terms of his preferences) without making someone else worse off (in terms of hers). The First Fundamental Theorem establishes that, in an idealized model of a market economy, every competitive equilibrium (roughly, every state of the market in which each commodity and each type of labour is traded at a single market-clearing price) is Pareto efficient. The Second Fundamental Theorem establishes that every Pareto-efficient outcome is a competitive equilibrium in

relation to *some* initial distribution of property rights. As Kenneth Arrow, a co-author of a celebrated proof of those theorems, explained in his 1972 Nobel Lecture, they can be interpreted as a highly stylized representation of what is good about the market system, if combined with an appropriate scheme of income redistribution. At the same time, the defining properties of the idealized market economy – in particular, the absence of public goods, externalities, market power or asymmetries of information – can be read as identifying sources of 'market failure' that require policy interventions if Pareto efficiency is to be achieved (Arrow 1972, 109–110).

In neoclassical models, preferences have a dual role: individuals choose the actions that best satisfy their preferences, and preference-satisfaction is the normative criterion against which economic systems and policy proposals are assessed. Because of this duality, the neoclassical approach has an underlying presumption in favour of freedom of choice. A market economy with a reasonably equal distribution of income gives each individual a wide range of choice, and incentivizes firms to supply products that consumers want to buy. If policy interventions to counter market failures are assessed in terms of the preference-satisfaction criterion, favoured interventions will tend to be ones that allow scope for individual choice (for example, carbon taxes rather than selective prohibitions) or that simulate voluntary transactions (for example, using willingness-to-pay measures to determine the supply of public goods). Thus, neoclassical economists have been able to defend regulated market economies as upholding the value of 'consumer sovereignty' or of being 'free to choose'.

This understanding of markets has been undermined by developments in behavioural economics that began in the late 1970s. A recurring finding of behavioural economics, and of the cognitive psychology from which it developed, is that individuals' choices are *context-dependent*: they are sensitive to features of the context or 'framing' of decision problems that seem to have no relevance to individuals' interests or well-being. The implication is that individuals often do not act on, and perhaps do not even have, the consistent preferences that neoclassical economists assumed. If such preferences do not exist, the whole concept of Pareto efficiency – indeed, the whole idea of using preference-satisfaction as a normative criterion – is fatally flawed. If they exist in some latent form but are not acted on, it might be possible to retain preference-satisfaction as a normative criterion, but economists would need a theory of what latent preferences are, and a method for discovering them. Supposing that these problems could be overcome, the solution would necessarily break the connection between preference-satisfaction and freedom of choice, as the latter has normally been understood by liberal economists.

This essay is about one of the main ways in which economists have dealt with these issues. In broad terms, behaviour that contravenes the rationality axioms of standard decision

theory is explained as the result of individuals' systematic errors, biases, and failures of self-control – effects that can be explained by psychological mechanisms. Normative conclusions are derived by inferring the preferences that individuals *would have* acted on, had they not been subject to these effects, and by seeking to satisfy these purified preferences. Such conclusions are often unashamedly paternalistic, and are sometimes explicitly presented as evidence of the untenability of liberal justifications for the market. In this domain of economic debate, liberalism *is* under siege. I will propose a strategy for breaking the siege – a strategy that disengages liberalism from rationality.

1 Three perspectives on behavioural findings

In discussing how economists have responded to the findings of behavioural research, I will focus on three books that have been written with serious intent for general readerships: Daniel Kahneman's *Thinking, Fast and Slow* (2011), Richard Thaler and Cass Sunstein's *Nudge* (2008), and George Akerlof and Robert Shiller's *Phishing for Phools* (2015). On separate occasions, Kahneman, Thaler, Akerlof and Shiller have been winners of the Nobel Memorial Prize in Economic Sciences. These books can safely be taken as expressing views that are now widely held in the economics profession and as making significant contributions to public debate.

The order in which I have listed these three books is not chronological, but represents an underlying progression of ideas. Kahneman's book surveys a programme of research that he (with Amos Tversky, who died before Kahneman's Nobel Prize was awarded) had been pursuing since the 1970s. This was a programme of cognitive psychology, based on experimental investigations of how individuals make decisions. It was explicitly presented as a challenge to the assumption, fundamental to neoclassical economics, that individuals' economic behaviour is consistent with maximizing theories of rational choice. Kahneman and Tversky's 1979 paper 'Prospect Theory', published in the most prestigious journal of economic theory, threw down this challenge by presenting experimental evidence of *anomalies* in decision-making behaviour – systematic violations of the axioms of rational choice theory – and proposing a theory that could explain them. It initiated the development of behavioural economics as a recognized branch of the discipline.

At first, many economists responded by arguing that the findings of laboratory experiments should not be expected to apply to decisions made when individuals have opportunities and incentives to discover and correct anomalies in their preferences, and that these were the kinds of decision most relevant for economics (e.g., Smith 1998; Plott 1999). But this defence became less and less tenable in the face of an accumulation of evidence (both from controlled experiments and from observations of economic behaviour 'in the field') of systematic violations of rational choice theory in situations in which decision makers faced significant incentives and had ample opportunity to learn from experience.

A second phase of behavioural economics began in the early 2000s, when economists started to consider the implications of behavioural findings for *normative* analysis. *Nudge* expands on and popularizes a seminal paper published by Sunstein and Thaler in 2003, 'Libertarian Paternalism Is Not an Oxymoron'. The central idea of their paper and of the later book is that the findings of behavioural economics invalidate the non-paternalistic stance of neoclassical welfare economics, and that this gives governments and firms a legitimate role of helping individuals to overcome their own psychologically-induced biases, errors, and temptations. However (it is claimed), these interventions can be made by *nudges* which, although paternalistic in intent, do not compromise individuals' freedom of choice – hence the slogan of 'libertarian paternalism'.

The idea that nudges can be used as a policy tool has been highly influential in academic and public debate, and has been taken up by policy makers around the world. But it has also provoked a reaction from commentators such as Nick Chater and George Loewenstein (2023), who argue that, by characterizing social problems as resulting from the biases and errors of private individuals, it has diverted attention from more fundamental 'structural' causal mechanisms. One strand of this argument starts from the same hypotheses about bias, error and temptation as are used in the nudge literature, but proposes harder forms of intervention. Significantly, the emphasis is on the role of firms in *creating* situations that induce biases, errors and temptations that are profitable for those firms but harmful to their customers.

This is a running theme in Akerlof and Shiller's *Phishing for Phools*. The topic of the book is encapsulated in its subtitle, *The Economics of Manipulation and Deception*. Much of the book is about deception in financial markets, and its role in setting off the world financial crisis of 2008. But deception in financial markets and the 'manipulation' of consumers in retail markets are treated as two manifestations of a single mechanism that is intrinsic to the market system. In the words of the blurb on the book jacket, 'markets are inherently filled with tricks and traps and will "phish" us as "phools"'. My concern is with Akerlof and Shiller's analysis of retail markets.

2 Kahneman on bias and error

The organising principle of *Thinking, Fast and Slow* is a distinction between two systems of human mental processing, first proposed by Peter Wason and Jonathan Evans (1975). System 1 is fast and automatic, and generates impressions, intuitions, feelings and impulses. System 2 is slow and under conscious control; using it is effortful. Operating on the outputs of System 1, System 2 constructs explicit thoughts in an orderly way. According to Kahneman, a person's sense of identity is located in her System 2. He describes the reader's System 2 as 'the conscious being you call "I"', and claims that '[w]hen we think of ourselves, we identify with System 2, the conscious, reasoning self that has beliefs, makes choices, and decides

what to think about and what to do' (pp. 21, 27). Nevertheless, 'the thoughts and actions that System 2 believes it has chosen are often guided by the figure at the center of the story [i.e., the story told in Kahneman's book], System 1' (p. 32). The implication is that we do not perceive our immediate intuitions, feelings and impulses as truly ours until we have converted them into judgements that we consciously endorse.

Kahneman introduces the distinction between the two systems by reminding readers of the familiar optical illusion of two lines of equal length but with arrows at their ends that point either inward or outward. System 1 produces the perception that the line with the inward pointing arrows is longer. The realisation that the perception is illusory is a product of System 2. Although this example is about visual perception, Kahneman's main concern is with 'cognitive' illusions which, unless corrected by System 2, affect judgements and decisions. Since System 1 is fast and effortless and usually creates impressions that System 2 would endorse, we are wise to rely on it in most of the situations we face. But Kahneman advises us to 'learn to recognize situations in which mistakes are likely and try harder to avoid significant mistakes when the stakes are high'. His book is about how to do this (pp. 26–28).

However, Kahneman skirts round a significant disanalogy between the optical illusion of the two lines and many of the cognitive illusions that he presents as influencing decisions. The context of the optical illusion is a judgement about a question (the relative length of two lines) for which there is an objectively correct answer. There is a correspondingly objective sense in which the System 1 perception is an *error* and the effect of the arrows is a *bias*. Further, with the help of a ruler, or just a sheet of paper and a pencil, the person making the judgement can access the correct answer by System 2 reasoning. This provides a straightforward way of acting on Kahneman's advice to try to avoid making mistakes.

In contrast, consider the effect of *loss aversion*, theorised by Kahneman and Tversky (1979) and Tversky and Kahneman (1991) and discussed at length in Kahneman's book (pp. 269–321). A famous example of this effect was observed by Jack Knetsch (1991) in a beautifully simple classroom experiment using coffee mugs and chocolate bars. In one session, each participant received one of the mugs as a free gift, and shortly afterwards was invited to exchange it for one of the chocolate bars; 89 per cent of participants chose not to exchange, implicitly reporting a preference for the mug. In a different session with similar participants, each participant received a chocolate bar and was invited to exchange it for a mug; 90 percent chose not to exchange, implicitly reporting a preference for the chocolate. One can reasonably infer that, for many of Knetsch's participants, which of the two goods they would perceive as preferable would depend on which they had been given a few minutes before. Kahneman explains this *endowment effect* as a result of loss aversion. (In reaching a decision, the loss of giving up something you already have is given more psychological

weight than the gain of getting something else in its place.) Loss aversion, like many of the other anomalous choice patterns described by Kahneman, is a form of context dependence. But is context dependence a *bias*? Is it an *error*?

As I have argued elsewhere, the concepts of bias and error are intelligible only in relation to a concept of truth (Infante et al. 2016). In a decision problem, the analogue of true length in the optical illusion must be *true preference*. The problem is that preferences (at least between consumption goods such as mugs and chocolates) are fundamentally subjective. At most, the axioms of rational choice theory require that an individual has *some* context-independent preference (or indifference) between each pair of potential options; the axioms are silent about what that preference should be. Psychological theories that explain context dependence have a corresponding limitation. For example, they can explain how, perceived in relation to any given reference point, losses are psychologically more salient than gains; but they cannot tell us which reference point induces *true* preferences over any particular pair of options. Truth in preferences is not an empirical concept.

The disanalogy between objective and subjective entities may not be particularly troubling if (as in Kahneman's advice to his readers) the person who is to correct 'biases' and 'errors' is also the person whose System 1 supposedly created them. Suppose I realize that my unreflective preference between two given options depends on the context in which I have to choose between them. In some particular context, I have to make this choice. What I take to be Kahneman's advice – that if the stakes are high, I should try to form a judgement about which I *really* prefer – might not be much help, but it is sensible enough. But the disanalogy is more worrying if, as proposed by Thaler and Sunstein, the concepts of bias and error are to be used to guide paternalistic interventions, and the designer of the interventions needs to make judgements about what people in general really prefer.

3. Thaler and Sunstein on libertarian paternalism

In the introduction to *Nudge*, Thaler and Sunstein undertake to show that

> in many cases, individuals make pretty bad decisions – decisions that they would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control. (p. 5)

This proposition is counterposed against the supposed commitment of 'textbook' economics to the assumption that 'each of us thinks and chooses unfailingly well' and thereby to the rejection of paternalism (p. 6). Thaler and Sunstein will argue for paternalistic interventions (by governments and firms) that are designed to counter individuals' tendencies to make bad decisions. The aim of these interventions will be to 'make choosers better off, *as judged by themselves*' (p. 5, italics in original). The paternalist's job requires her to reconstruct those judgements.

The evidence of bad decision making is presented in the first two chapters of *Nudge*. The first chapter, 'Biases and blunders', has essentially the same structure as Kahneman's argument. It begins with an optical illusion – a drawing of two table tops with the same dimensions, one of which looks much longer and narrower than the other. The lesson to be drawn is that 'our understanding of human behavior can be improved by appreciating how people systematically go wrong'. Like Kahneman, Thaler and Sunstein use the System 1/System 2 distinction to interpret the evidence about how decision making can go wrong (pp. 17–22). Their catalogue of decision-making blunders includes loss aversion, endowment effects, and framing effects (i.e., tendencies for choices between given options to vary according to how the choice problem is described). Since these effects are different forms of context-dependence, they can be countered by nudges which affect only the 'choice architecture' within which decisions are made.

The second chapter of *Nudge*, 'Resisting temptation', is about self-control problems. Thaler and Sunstein represent individuals as having two selves – a rational self (the 'Planner') whose intentions can be subverted by an impulsive self (the 'Doer'). When psychological arousal is low ('cold' states), the Planner is in control, but 'hot' states of high arousal activate the Doer. Thus, an individual's choices between given options can vary according to properties of the choice context which affect arousal. As an illustration, Thaler and Sunstein describe a Cinnabon outlet at Chicago O'Hare Airport, whose oven aromas subvert the health-oriented intentions of Planners heading for the nearby fruit and yoghurt stand (pp. 40–42, 49). Again, context-dependence allows room for nudges to influence individual behaviour.

But how are policy makers (or 'choice architects' as Thaler and Sunstein prefer to call them) to decide the directions in which individuals are to be nudged? Thaler and Sunstein's official answer is that individuals should be nudged towards the choices they would have made in the absence of biases, errors, and failures of self-control. There is an implicit assumption that after correcting for these reasoning failures, each individual will be found to have determinate, context-independent latent preferences over the outcomes that feature in the situations in which nudges are to be used. When Thaler and Sunstein use the 'as judged by themselves' clause, they are presumably referring to these assumed latent preferences. But they do not explain why, as a matter of empirical psychology, we should expect such preferences to exist. Presumably (given their endorsement of the System 1/System 2 distinction), Thaler and Sunstein are assuming that each person's System 2 has access to some mode of reasoning which, if used, would generate context-independent preferences – in something like the way that a person can use a ruler to arrive at a judgement about the relative length of two lines. But what is that mode of reasoning? We are never told. Given the disanalogy between the objectivity of length and the subjectivity of preference, we are entitled to be sceptical about whether such a mode of reasoning exists at all.

Nudge is full of concrete proposals for nudges, but when Thaler and Sunstein need to claim that individuals are being helped to overcome what *the nudgees themselves* would judge to be errors, their arguments become rather casual. For example, when arguing for nudges against smoking, drinking and obesity-inducing diets, they simply report familiar statistics about the associated health risks and then conclude: 'With respect to diet, smoking, and drinking, people's current choices cannot reasonably be claimed to be the best means of promoting their well-being' (pp. 7, 44). But when some particular person, say Jane, reflects on her choice of a cream doughnut rather than a kiwi fruit in a particular café on a particular day, is she committed to the claim that a high-fat, high-sugar diet over her lifetime is the best means of promoting her lifetime well-being? Or just to the claim that, on the day and all things considered, she prefers the doughnut? The latter claim need not involve any bias, error, or failure of self-control. It is a self-endorsement of a context-dependent preference.

The argumentative moves that I have called 'casual' mask significant shifts in the normative standard by which nudges are being justified. The shift is from the subjective standard of each individual's latent (i.e., error-free, context-independent) preferences to a more objective standard of well-being. The claim that Jane's long-term well-being would be better served by a diet that included fewer doughnuts and more kiwi fruit is entirely plausible, but it does not imply that, when she chooses the doughnut on the particular day in the particular café, she is making what she would acknowledge as an error. These shifts suggest to me that Thaler and Sunstein are finding it difficult to create convincing arguments about the content of nudgees' latent preference – a difficulty which would be unsurprising if those preferences were ill-defined or did not exist.

4. Akerlof and Shiller on phishing

Even if Thaler and Sunstein's proposals leave open crucial questions about how policy makers should direct their interventions, there is some reassurance for liberal-minded readers in the fact that what are being proposed are only nudges. Nevertheless, Thaler and Sunstein sometimes present behavioural findings as grounds for questioning the value of the freedom of choice that competitive markets provide. For example, in their discussion of the Cinnabon outlet, they treat it as an example of an imperfection of the market system. With the implicit suggestion that the outlet may have been deliberately designed to allow its aromas to reach passers-by, Thaler and Sunstein draw the general lesson that 'even when we're on our way to making good choices, competitive markets find ways to get us to overcome our last shred of resistance to bad ones'. And as an explanation of this effect, they refer to an undeniable fact about the market mechanism: 'Markets provide strong incentives for firms to cater to the demands of consumers, and firms will compete to meet these demands, whether or not those demands represent the wisest choices' (p. 49).

This fact is the starting point for Akerlof and Shiller's *Phishing for Phools*. Akerlof and Shiller describe themselves as 'admirers of the free market system', but they express a very general sense of malaise about the workings of that system:

> But, inevitably, the competitive pressures for businessmen to practice deception and manipulation in free markets lead us to buy, and pay too much for, products that we do not need; to work at jobs that give us little sense of purpose; and to wonder why our lives have gone amiss. (p. vii)

Deception and manipulation are represented as *phishing*, defined as 'getting people to do things that are in the interest of the phisherman, but not in the interests of the target'. A *phool* is 'someone who, for whatever reason, is successfully phished'. My concern in the present paper is with retail business practices that Akerlof and Shiller would classify as manipulative bur not deceptive. Such practices involve what they call *psychological phishing*. This occurs in two forms, depending on the psychological effects that are exploited. In one case, those effects are 'cognitive biases, which are like optical illusions, [leading the phool] to misrepresent reality [and to act] on the basis of that misrepresentation'. These effects correspond with 'biases' and 'errors', as understood by Kahneman, Thaler and Sunstein. In the other case, which corresponds with Thaler and Sunstein's 'temptation', 'the emotions of a psychological phool override the dictates of his common sense' (p. xi).

Psychological phishing is possible because of what Akerlof and Shiller present as an established finding of psychology – that 'people frequently make decisions that are not in their best interest. Put bluntly, they do not do what is really good for them; they do not choose what they really want' (p. 1, repeated in slightly different words on p. 5). It is difficult to know what to make of this sweeping but ambiguous claim. On the most natural interpretation, goodness is a matter of normative judgement, but wanting is a matter of psychological fact. The concept of 'really good' seems to imply an objective normative standard; the concept of 'really wanting' suggests an inference about the deeper psychological mechanisms than underlie actual wants. On one reading of Akerlof and Shiller's equation between the two concepts, 'really wanting' corresponds with Thaler and Sunstein's latent preferences, and 'really good' expresses a judgement about the moral value of satisfying those preferences. That interpretation leads to the problems I described in Section 3. On another reading, what is 'really good' for a person is determined by some theory of welfare that may have only a contingent relationship with her actual wants or judgements. It is a matter of definition that what each person 'really wants' is whatever is really good for her: if in fact she wants something different, that is false consciousness. On that reading, there would be a fundamental break with the liberal tradition.

Akerlof and Shiller recognise that they will be criticized for presuming to know that individuals are 'making decisions that, applying just a bit of their own common sense, they

would know are not to their benefit'. How could Akerlof and Shiller be so sure about this? Their wake-up-and-smell-the-coffee answer expresses impatience with philosophical niceties:

> We do not have to be presumptuous to see that people are making such [phoolish] decisions. We know because we see people making decisions that NO ONE COULD POSSIBLY WANT. (p. xii, capitalization in the original)

As examples of such decisions in the domain of consumer choice, Akerlof and Shiller refer to the health effects of smoking, alcohol consumption, and high-fat, high-sugar, high-salt diets, and pronounce that 'no one now thinks it is smart to smoke', 'no one wants to be an alcoholic', and 'no one wants to be obese' (pp. xv–xvi). But that is not what is at issue. Take the case of alcoholism. No one *wants* to be an alcoholic, but neither does anyone *decide* to be one. People develop alcoholism as a result of repeated decisions to have another drink. It is all too obvious that people *can* want to have another drink and that thoughts about the dangers of alcohol dependency can be countered by more comfortable thoughts along the line of 'It won't happen to me'. It may well be true of a whole population over a period of decades that the conjunction of 'It won't happen to me' beliefs understates the overall incidence of alcoholism. Viewed at the population level and over the long run, that would indeed be an objectively definable bias. But that does not imply that the steps by which an individual becomes an alcoholic are errors *as judged by himself*. Despite the capital letters, the 'NO ONE COULD POSSIBLY WANT' criterion cannot justify Akerlof and Shiller's treatment of 'really want' and 'really good' as self-explanatory concepts. This leaves a huge gap in their arguments.

To organize observations of phoolish decisions, Akerlof and Shiller use a dual-self model. After a brief reference to an experiment in which capuchin monkeys who had been taught to trade revealed tastes for Fruit Roll-Ups:

> We can imagine us humans, like the capuchins, as having two different types of tastes. The first concept of 'tastes' describes what is really good for us. But, as in the case of the capuchins, that is not always the basis for all of our decisions. The second concept of 'tastes' determine how we really, actually make our choices. And those choices may not, in fact, be 'good for us'. (p. 4).

Then, mixing metaphors:

> [W]e can think about our economy as if we all have monkeys on our shoulders when we go shopping or when we make economic decisions. Those monkeys on our shoulders are in the form of the weaknesses that have been exploited by marketeers for ages. Because of those weaknesses, many of our choices differ from what we 'really want', or, alternatively stated, they differ from what is good for us. (p. 5).

This monkey-on-the-shoulder metaphor has clear parallels both with Kahneman's System 1/ System 2 model and with Thaler and Sunstein's Planner/Doer model. But Akerlof and Shiller take the idea a step further by representing the psychological mechanisms that supposedly induce error and bias in a person's decisions as entirely alien to her – as external manipulators of her real wants and as deadweights to be carried around. It is as if the features of the world that cue those mechanisms do not engage with the real person at all, but only with the imaginary monkey on her shoulder.

Given Akerlof and Shiller's professed admiration for free markets, their accounts of what happens in these markets are strikingly dystopian. For example, when picturing an ordinary supermarket, they begin with the candy displayed at the checkout counter and the cigarettes that used to be displayed there before this practice was prohibited. They continue:

> There are thousands more phishes in the supermarket, embodied in all the different products on the shelves, each with its own team of marketing experts and advertising campaigns, each the product of experimentation with many other possible marketing forms. (p. 21)

On purchases of potato chips:

> [Consumers] know what they are buying. The chips come in bags that are correctly labeled, even with the number of calories; but the companies are phishing their consumers in another way. The potato chips are now scientifically designed, with the optimal amounts of fat and salt, to maximize their sales. (p. 94)

And drawing a general lesson from all this:

> The ability of free markets to engender phishing for phools … is inherent in the workings of competitive markets. And the same motives for profit that give us a healthy benign economy if everyone is fully rational are the same motives that give us the economic pathologies of phishing for phools. (p. 166)

Think what is being said here. Firms try to discover products and ways of displaying them such that, when consumers see these products, they will choose to buy them at prices that give firms a margin of profit. If consumers are fully rational – if what they choose is what they really want – this discovery process is benign: free markets are to be admired. But if consumers are *not* rational, the discovery process is a pathology: firms are catering to the demands of the monkeys on the consumers' shoulders. In other words, the freedom of choice that markets offer is unambiguously valuable *only for rational individuals*.

When the Fundamental Theorems of welfare economics were first proved, the assumption of maximizing behavior by economic agents was a modelling simplification. Because this assumption was treated as an acceptable approximation to normal human

behaviour, liberal economists were able to interpret the theorems as a stylized demonstration of the value that *everyone* derives from competitive markets. It would be a sad retreat for the liberal tradition if the most that could be said was that *rational* individuals have reason to value competitive markets, and that psychologically normal responses to the cues of everyday life are sufficient to demonstrate a person's irrationality. Is economic liberalism addressed only to people whose preferences satisfy the rarefied standards of rational choice theory?

5. Opportunity

Thaler and Sunstein's example of the Cinnabon outlet in the airport reappears in *Phishing for Phools* as Akerlof and Shiller's 'favorite example' of phishing in retail markets:

> Most of us probably take it for granted that there just happens to be [a Cinnabon outlet] right where we are waiting for our delayed flight. We fail to appreciate how much effort and expertise went into understanding our weak moments and developing a strategy to take advantage of them. (pp. 2, 171).

But there is another way of thinking about the example. Suppose I have arrived at an airport on a long-haul flight and have just learned that my connecting flight has been delayed. I don't normally buy snacks while travelling, but the combined effects of tiredness, frustration and boredom induce an unanticipated desire to eat something sweet and cake-like. No sooner has this desire appeared than I notice the Cinnabon outlet (perhaps aided by its appetizing aromas). Taking for granted the coincidence of my desire and the means of satisfying it, I fail to appreciate that this is an instance of a general property of competitive markets.

In a competitive market, firms have incentives to predict regularities in what people in general desire and are willing to pay for, and to look for profitable ways of satisfying those desires. If I hadn't taken the Cinnabon outlet for granted, I might have reasoned my way to the expectation that something like it would exist. Having experienced the desire, I can recognize the features of my situation (the delay, my tiredness, frustration and boredom) that have caused it. I can recognize that air travellers frequently find themselves in similar situations, and that the airport is full of air travellers. So, it is a reasonable expectation that other people in the airport often have desires similar to mine, and that satisfying them might be profitable for firms. And if it *is* profitable, it is a reasonable expectation that some firm will have found a business model that creates profit by satisfying these desires.

In more general terms, competitive markets tend to supply whatever people *in fact* want and are willing to pay for, whenever they want it and are willing to pay for it. This tendency exists irrespective of whether individuals are rational or irrational. In the example, my want for a cinnamon bun exists at the particular time and place at which I express my willingness to pay for it. I may have had no previous expectation of having this want, and I might regret it later. It might not have existed in the absence of the cinnamon aromas.

Nevertheless and as a matter of psychological fact, it is what I – the human being, not some normative conception of my 'true self' – want at the time.

This tendency of competitive markets can be expressed in a stylized form as the Strong Market Opportunity Theorem (McQuillin and Sugden 2012; Sugden 2018, 107–139). Mathematically, this theorem is closely related to the First Fundamental Theorem of welfare economics. (Formally, the Strong Market Opportunity Theorem can be interpreted as a lemma in a proof both of the First Fundamental Theorem and of the stronger result that every competitive equilibrium is in the 'core' of the relevant economy. What needs to be added to this lemma to complete those proofs is the assumption that individuals act on consistent preferences.) Conceptually, however, the theorems apply to fundamentally different domains: the Strong Market Opportunity Theorem is about opportunity rather than welfare.

For simplicity, consider an economy in which the only economic activity is the exchange of commodities between individual *consumers*, and in which exchange takes place in only one period. A *competitive equilibrium* can be defined as a state of the economy in which, for each tradable commodity, the following two conditions are satisfied. First, there is a single *market price* at which all consumers are free to buy and sell (the *Law of One Price*). Prices are expressed in units of money, which can be treated as one of the commodities. Second, consumers' total desired purchases of the commodity, given these prices, are equal to their total desired sales (the *Law of Supply and Demand*). To explain why these two laws can be expected to hold in a competitive market, it is sufficient to assume that exchanges between consumers are intermediated by a class of profit-seeking and non-colluding *traders*. The only assumption that needs to be made about each consumer's desired transactions is that *dominated* transactions – transactions that involve unambiguous monetary loss relative to other available opportunities – are never desired (Sugden, 2004). Since there has been no other reference to consumers' preferences, the concept of Pareto efficiency has no application. Nevertheless, it is still possible to talk about consumers' opportunities to buy and sell.

The Strong Market Opportunity Theorem states that, in every competitive equilibrium, every set of consumers has the collective opportunity to make any combination of non-dominated exchanges among themselves that they might conceivably desire to make. To see the intuition behind the theorem, consider any competitive equilibrium, any potential exchange transaction, and any consumer (say Joe) who would be a party to it. Necessarily, Joe's component of that transaction must either increase, hold constant or decrease the value of his holdings of commodities, measured at market prices. If the effect is a decrease in value, Joe's component is dominated by a trade that he could make by trading at market prices. Since the transaction is an exchange, the total market value of the holdings of the consumers who are parties to the transaction must remain constant. Thus, if Joe's holdings

increase in value, the holdings of some other party (say Jane) must decrease in value, implying that the transaction is dominated for her. The only remaining possibility is that value remains constant for *all* parties. But then each party's component is one that he or she can achieve by trading at market prices. In other words, the collective opportunity to make this transaction is available in the competitive equilibrium.

The Strong Opportunity Theorem can be extended to economies in which exchanges can be made in more than one period. The extended theorem does not use any assumptions about the consistency of consumers' desires over time: a combination of transactions over two or more periods counts as non-dominated if, in each period considered separately, each consumer's transactions *in that period* are not dominated by other transactions available in the same period. It is in this sense that a competitive market allows individuals to get whatever they want and are willing to pay for, whenever they want it and are willing to pay for it. I maintain that this statement can be read as a reason for an individual to value the opportunities that she would be given in a competitive market. If that is right, it is a reason that can be offered to individuals irrespective of whether or not their preferences satisfy economists' principles of rationality. The market is blind to consumers' reasons – or even their lack of reasons – for wanting to make the transactions they choose to make.

But, one might ask, if a person's wants have been induced by a firm in its pursuit of profit, does that compromise her reason to value the opportunity to get what she wants? Given Akerlof and Shiller's remarks about the Cinnabon outlet and about the scientific design of potato chips, and given Thaler and Sunstein's hint that the cinnamon aroma might be a deliberate nudge, I take it that both pairs of authors would say 'Yes'. I disagree.

Let me return to the example in which I am the delayed air traveller. Before I learned that my connecting flight was delayed and before I noticed the Cinnabon outlet, I may have had no preference about what to eat or not eat conditional on a delay that I was not expecting to occur. Suppose that was the case. In the psychological sense of the word, my desire for a particular brand of cinnamon bun is *constructed* in response to my learning about the delay and by my awareness of an offer that Cinnabon and its airport franchisee are making to me. Considered in relation to my decision about eating, the delay is a random event, but the offer is an intentional act by Cinnabon and the franchisee. Does this kind of intentionality compromise my reason to value an offer?

The answer to this question depends on one's conception of the role of entrepreneurship in a market economy. As viewed in economics, an entrepreneur's objective is to discover and implement combinations of transactions that the parties to those transactions are willing to make and that also generate a surplus that the entrepreneur can appropriate. These two facets of entrepreneurship are inseparable from one another: it is

because entrepreneurs take a share of gains from trade when they discover them that they have an incentive to seek out those gains.

Much of the creativity of entrepreneurship consists in discovering opportunities for gains from trade that would-be transactors are unaware of. Successful entrepreneurship involves directing the attention of potential trading partners to the opportunities the entrepreneur is offering and to the properties of those offers that she expects they will find attractive. If preferences are context-dependent, so too are perceptions of attractiveness. Attention is a psychological phenomenon, responsive to cues that theories of rational choice assume away. Inescapably, entrepreneurship involves forms of persuasion that do not work through the narrow channels that those theories recognize. But this is equally true of other domains of voluntary interaction: think of the rhetorical content of discourse in politics, journalism, religion or science. Non-deceptive persuasion that activates psychological cues is not a pathology; it a fundamental facet of human interaction in a free society. That Cinnabon's outlets are located in places where desires for its products are most likely to arise, and that the visibility and aromas of these outlets are designed to activate those desires, are examples of the workings of an opportunity-generating mechanism that individuals can recognize as valuable for them.

6. Disengaging liberalism from rationality

I have tried to show that there are coherent and cogent reasons for valuing competitive markets, and that an individual can endorse those reasons while also recognizing that his own economic behaviour is that of a psychologically normal human being whose preferences and judgements are often context-dependent. In short, liberal economics is more than a belief system for rational individuals.

Of course, the whole idea of offering reasons for valuing some institution presupposes that one's addressees are rational in the sense of being willing and able to engage in reasoning. In this broad sense of 'rationality', any justification for competitive markets must be addressed to individuals as rational agents. But it is fundamental to many forms of liberal argument that what are being justified are general rules that are to be applied in many different circumstances, the details of which are unknown when the justification is being offered. If the rules are to govern interactions between people, the justification typically needs some assumptions about individuals' behaviour and mental states in the situations to which the rules apply, but those assumptions are necessarily general rather than specific. The neoclassical justification for competitive markets follows this schema. A competitive market is a system of general rules. The neoclassical justification assumes that each individual has context-independent preferences over relevant outcomes, but makes no substantive assumptions about the content of those preferences. I have tried to show that a liberal justification for the market does not need even those formal assumptions about preferences.

Just as the addressees of the neoclassical justification do not need to know what their actual preferences will be, my addressees do not need to know whether their preferences will be context-independent. That does not prevent them from reasoning coherently about the opportunities that competitive markets provide.

How much does all this matter? Economists' responses to the findings of behavioural science may seem peripheral to a discussion of how political liberalism may be threatened by the forces of populism, nationalism and authoritarianism. But those responses may be symptoms of a general property of modern liberal thought that connects with its vulnerability to these threats. Many political commentators have noticed how populism builds on a widespread sense of resentment of what is perceived as the condescension of a 'metropolitan elite' of highly-educated professionals who want to impose their self-certified senses of expertise and moral rightness on the rest of the population. Think about what is signalled when academic writers claim that markets offers freedoms that are valuable only for rational agents, and that other people are 'phools' who need help to steer them away from choosing things that 'NO ONE COULD POSSIBLY WANT'. Think about the confidence with which academic writers claim insights into individuals' true preferences, despite the casualness of the arguments they present in support of these claims. All this is a far cry from the robust liberalism of John Stuart Mill's ([1859] 1972 :75) principle of 'liberty of tastes and pursuits; of framing the plan of our life to suit our own character; of doing as we like, subject to such consequences as may follow'. Now may be the time to disengage liberalism from rationality.

References

Akerlof, George, and Robert Shiller. 2015. *Phishing for Phools: the Economics of Manipulation and Deception*. Princeton, NJ: Princeton University Press.

Arrow, Kenneth. 1972. "General Economic Equilibrium: Purpose, Analytic Techniques, Collective Choice." Nobel Memorial Lecture. https://www.nobelprize.org/uploads/2018/06/arrow-lecture.pdf

Chater, Nick, and George Loewenstein. 2023. "The i-frame and the s-frame: How Focusing on Individual-Level Solutions has Led Behavioral Public Policy Astray." *Behavioral and Brain Sciences* 46: e147.

Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016. "Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioural Welfare Economics. *Journal of Economic Methodology* 23 (1): 1–25.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–291.

Knetsch, Jack. 1989. "The Endowment Effect and Evidence of Non-Reversible Indifference Curves." *American Economic Review* 79 (5): 1277–1284

McQuillin, Ben, and Robert Sugden. 2012. "How the Market Responds to Dynamically Inconsistent Preferences." *Social Choice and Welfare* 38 (4): 617–634.

Mill, John Stuart. [1859] 1972. *On Liberty*. London: Dent.

Plott, Charles. 1996. "Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis." In *The Rational Foundations of Economic Behaviour* ed. Kenneth Arrow, Enrico Colombatto, Mark Perlman and Christian Schmidt, 225–250. Basingstoke, UK: Macmillan and International Economic Association.

Smith, Vernon. 1989. "Theory, Experiment and Economics." *Journal of Economic Perspectives* 3 (1): 151–169.

Sugden, Robert. 2004. "The Opportunity Criterion: Consumer Sovereignty without the Assumption of Coherent Preferences." *American Economic Review* 94 (4): 1014–1035.

Sugden, Robert. 2018. *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford: Oxford University Press.

Sunstein, Cass, and Richard Thaler. 2003. "Libertarian Paternalism Is Not an Oxymoron." *University of Chicago Law Review* 70 (4): 1159–1202.

Thaler, Richard, and Cass Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Tversky, Amos, and Daniel Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics* 106 (4): 1039–1061.

Wason, Peter, and Jonathan Evans. 1975. "Dual Processes in Reasoning?" *Cognition* 3 (2): 141–154.