

Title	Two philosophical blind spots in the pedagogy of undergraduate economics
Authors	Ross, Don
Publication date	2025-09-02
Original Citation	Ross, D. (2025) 'Two philosophical blind spots in the pedagogy of undergraduate economics', in Ianulardo, G., Davis, J. B. and Crespo, R. F. (eds.) Handbook of Teaching Philosophy to Economists. Cheltenham, United Kingdom: Edward Elgar Publishing, pp. 283-297. https://doi.org/10.4337/9781035336821.00028
Type of publication	Book chapter;book-chapter
Link to publisher's version	10.4337/9781035336821.00028
Rights	© 2025, Edward Elgar. This is a draft entry. The final version is available in Ianulardo, G., Davis, J. B. and Crespo, R. F. (eds.) Handbook of Teaching Philosophy to Economists, pp. 283-297. https://doi.org/10.4337/9781035336821.00028 . It is deposited under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/). - https://creativecommons.org/licenses/by-nc/4.0/
Download date	2026-10-24 21:25:40
Item downloaded from	https://hdl.handle.net/10468/18284

Two philosophical blind spots in the pedagogy of undergraduate economics

Don Ross

School of Society, Politics, and Ethics, University College Cork
School of Economics, University of Cape Town
Center for the Economic Analysis of Risk, Georgia State University

don.ross931@gmail.com

forthcoming in John Davis & Giancarlo Ianulardo, eds., *Elgar Handbook of the Teaching of Philosophy to Economists*, Edward Elgar 2025

Abstract

I consider two lessons from philosophy that, if economists were better aware of them, could improve undergraduate pedagogy in economics. First, economists should familiarise themselves with the understanding of the nature of preferences and beliefs that is becoming the majority view in the philosophy of cognitive science, that they are externally attributed from the intentional stance, rather than internal representational states. This would make the idea of revealed preference easier for students to grasp, and could be taught simply as part of the introduction of utility functions, thus requiring no extra-economic digression. Second, economists should recognise that the idea of general rationality adds nothing to the concept of agency, and should stop using the former idea in the classroom because it serves only to confirm counter-productive stereotypes of economists as bossy suppressors of warm values.

JEL codes: A22, D01

Introduction

Over the past few decades, an increasing number of economists have recognised that the standard undergraduate curriculum, which for decades varied little across the world's universities, was serving the discipline, the policy environment, and most students, very inefficiently. It had been inherited without significant change from a world in which far smaller shares of populations registered for degrees, and in which it was, consequently, not entirely unreasonable to teach whole classes as if everyone present might strive to become an economist.

In that context, teachers could play a long game: teach first-year students the basic logic and simple calculus of constrained maximisation under shifts in supply and demand; then spend most of the next two or three years (depending on national degree structures) enriching this basic comprehension by embedding problems in slightly more complex markets; and then, in the final year, use the technical concepts that had been built up like bricks in a wall to present efficient solutions to a few stylised, generic policy settings, such as international trade, market expansion, or invention of new technology. The idea was that the graduating student would exit with an understanding – meaning ability to work standard calculations and interpret standard graphs – of the core tools of economic reasoning. Those who went on to graduate school would have a solid platform on which to finally consider, at that stage, some actual problems arising in actual economies. Those who studied no further at least had acquired a sense of what economists did: use a small stock of carefully crafted concepts and

argument patterns to strip problems down to their theoretical cores, where clear general principles might be brought to bear. As citizens, when they read public media reports of economists' policy advice, they could understand in general terms what kind of work the relevant claims to authority were based on.

This approach makes little sense in an environment where most first- and second-year students will never study any further economics, and will depart with – unless they are unusually curious and active in their learning – no idea how the simple textbook relationships they examined scale up to any recognisable policy controversy they read about as citizens and savings investors. Furthermore, as ever more economists spend ever more of their time applying advanced game theory and econometrics to empirical data, and gathering those data themselves through surveys and experimentation, the traditional undergraduate base has ceased even to be a toy representation of the intellectual style of the real discipline.

I will return later to some thoughts on why this wasteful practice persisted for so many years past the point at which it had obviously ceased to serve any purpose commensurate to its expense in human hours. But its inefficiency *was*, eventually, widely recognised, and reform is well underway. A milestone was the launch of the CORE econ curricular resource (<https://www.core-econ.org>) in 2013. A comprehensive undergraduate programme that weaves economic theory and concepts with policy case studies from multiple countries, including poor and developing ones, the CORE curriculum is now taught in 400 global economics departments, including several top-ranked US ones, the LSE, Oxford, and Toulouse. This is sufficient head of steam to predict the future: the culture of economics is famously conformist, so the old pedagogy is truly doomed.

The sponsors and writers of CORE econ make no claim that the current suite of materials *complete* the modernisation of economics pedagogy, so it constitutes no disrespect to it to identify elements that might be regarded as missing. It includes little explicit philosophy of economics. The qualifier 'explicit' is important, since any curriculum that motivates methodologies and special concepts cannot avoid a good deal of *implicit* philosophising. There are also deft, but by no means incidental, reflections on social and moral values that interact with economic arguments and policy preferences. But to the many critical commentators who have routinely called for reanimation of the deeper philosophical foundations of the discipline, CORE is likely to seem unfinished in that respect.

The question of how much explicit philosophy *should* be integrated into undergraduate economics is obviously one on which a diverse spectrum of opinion is inevitable, as the range of views in the present volume attests. My own sympathies are with those on the minimalistic side of that spectrum, for practical reasons: curricular space is scarce, and in that setting, treating philosophical complexities brusquely is at least as likely to encourage dogmatic judgments as critical wisdom. The strongest case can be made for explicitly correcting relatively strong philosophical presuppositions that have constrained the most general features of economic reasoning, and that have undermined the consolidation of economics into the wider network of knowledge. In this chapter, I will focus on two of them, which have been closely related both historically and logically: atomistic assumptions about the causal basis of preferences and beliefs, and casual appeal to a notion of rationality that ignores its irreducibly polysemous semantics.

As this is an exercise in reasoned statement of opinion, I will identify my standpoint. As an economic methodologist, I have over the years defended numerous special commitments

concerning the strongest available foundations for economic theory and practice; Ross (2014a, 2023) reviews many of them as they pertain to descriptive economics, and they are in turn applied to welfare economics in Ross (2022) and Harrison and Ross (2023). But many of these commitments bear only indirectly or not at all on how economics should be taught to undergraduates. Equally important to my standpoint here is having taught economics, at all levels, for twenty years at the University of Cape Town (UCT) and for five years in the United States. The programme at UCT has large enrolments: about 1600 students register annually in the first-year course, and over 300 complete the major. UCT shifted to the CORE curriculum shortly after my teaching years there had ended, but I participated in the department-wide review that culminated in the decision to implement it. I continue to supervise Honours and postgraduate dissertations by UCT students who have emerged from that undergraduate programme, and in guiding their reading and modelling choices I must take account of what they have and haven't learned.

A general constraint that I find very pressing, and which should therefore also be mentioned upfront, is that, CORE notwithstanding, the space for economics itself in the economics curriculum, has increasingly had to yield acreage to time spent teaching students the software skills needed for the empirical data analysis that they must already be able to apply by the point of their Honours dissertations, and knowledge of which constitutes the primary human capital asset for the early career stages of most of them. Teaching economics at all levels is thus under continuous pressure to become more efficient, as the Stata and R manuals that students must know become ever longer. (Perhaps the capacity of AI packages to do econometrics will reverse this trend of many years, but we are not at that potential inflection point yet.)

The chapter is structured as follows. In the second section I argue that economists' inheritance of a tradition of social atomism has blinded them to the logical independence of normative from descriptive individualism. I explain how learning a bit of contemporary philosophy of cognitive science could allow economists to teach the idea of a utility function in a natural way that avoids implausible atomism about the nature and origins of preferences and beliefs. In the third section I argue that this pedagogical reform should be accompanied by banishing the word 'rationality' from the economics classroom, explaining why this could improve economists' social reputations while trading off nothing of pedagogical substance. The final section briefly concludes.

Atomism

Economics as it is taught throughout the world was born in intellectual cultures of European origin, and those cultures, especially during the formative years of the social sciences, were philosophically atomistic. By 'atomism' I refer to a metaphysical predisposition that encourages a range of more specific stances, including methodological individualism (Udehn 2001), and decompositional attitudes to structural and causal mechanisms. Atomism is compatible with recognition of macro-scale structure; 19th-century versions of emergentism (McLaughlin 1992) reflect an assumed atomistic base in their central metaphor. The historical depth of this metaphysic in Western culture is a matter of fundamental contestation, not least because, as argued by Quinn (2024) among others, the meaning and extension of the idea of 'Western culture' are themselves contested. Notwithstanding these nuances of deeper history, by the time modern economics arose there was unquestionably a predominant atomistic substratum in European thought, on which the discipline was bound to be constructed.

Atomism has complementary descriptive and normative aspects, the blending of which has been particularly significant in economics.

Normative atomism as applied by economists to human values is the view that individual persons are the basic sources of valuation, in several senses that can sometimes come apart: (1) individuals are the sources of those of their values which, in combination, make each person a unique contributor to the distribution and content of social welfare judgments; (2) individuals are the units that experience the welfare consequences of policies, and those are the experiences that justify choosing policies on the basis of welfare implications; and (3) individuals are the only units *entitled* to have their values factored into welfare analysis without requiring a consequentialist argument. Economists who refer without further analysis to consumer sovereignty typically argue in ways that imply all three of these commitments.

Normative atomism has frequently supplied crucial premises to defenders of libertarian principles for selecting economic policies. Such arguments often confront critics with a challenging kind of density. On the one hand, libertarianism is often taken to simply follow directly from the conjunction of the three normative atomist claims above. On the other hand, libertarian policies are often defended on grounds of welfare efficiency – but then at least one of the normative atomist claims is invoked to ground the welfare criteria on which efficiency judgments depend.

Critics of standard economics often emphasise connections between normative individualism and policies for minimising public regulation of markets and allocations of wealth. However, as a pervasive metaphysical framework, atomism has also influenced the thought of the left; Elster (1985) argues persuasively that Marx's economics and sociology incorporated methodological individualism, and thereby reflected descriptive, though not necessarily normative, atomism. Marx was aware of this tension, and his handling of it evolved over the history of his work.

The charge has been reiterated in every generation that the individualistic bias incorporated into economics encourages selfishness and undermines communal values and solidarity (Coleman 2002; Ross 2012). For a relatively recent iteration, see Marglin (2008). In undergraduate curricula, this kind of critique is strongly represented in sociology syllabi, and we may expect that most economists will prefer to leave it there. If economists did try to introduce it into their courses, they would likely do so shallowly, briefly, and unsympathetically, so it is not clear to me that this would generate net positive effects. It is therefore not the object of my focus here. My interest is, rather, in blind spots around atomism that undermine the clarity of economics pedagogy with respect to economists' own ambitions for it.

A first blind spot is that descriptive and normative individualism are presented as if they are a natural package. I hypothesise that unreflective atomism is the explanation for this: if society is, as a matter of fact, ultimately decomposable into individual people and their properties as a matter of fact, than it seems natural to suppose that whatever is measured as a proxy for valuation should similarly decompose. But if casual atomism is avoided, then it is not difficult to appreciate that descriptive and normative individualism are entirely independent theses. Economists and their students would be well served if curricula emphasised this point; and precisely because it is not difficult, it is a bit of philosophy that, if introduced, would not significantly crowd out either economic theory or econometrics.

I'll focus on the basis for normative individualism first. Most economists are normative individualists because there is very strong selection pressure attracting liberals into the discipline. I use 'liberals' in the widest possible sense here, to take in everyone from followers of Hayek to advocates of socialism mediated by market mechanisms. I refer, that is, to proponents of social order based on rule of law aimed at indiscriminately protecting freedom of expression and association from restriction by powerful institutions. Almost all economists are 'biased' in favour of such social orders because these are the arrangements in which the kinds of incentives that economists study are most important and most interesting. Modern markets arose when new legal mechanisms, beginning in maritime city states, brought to full complexity in the Dutch golden age, and then forming the basis for a massive expansion of wealth in England and her imitators, allowed the roles of investors and producers to be rigorously disaggregated by financial institutions (Hodgson 2015, 2023). The legal mechanisms in question rested on the idea that individuals can be incentivised to seek to materially flourish according to their own judgments about both what they want and how much risk they can tolerate.

The evolution of economic thought makes little sense without liberalism in this sense. Welfare analysis as economists theorise it, and as they practice when they work as consultants, presupposes it. Economists thus tend naturally to be people who are hostile to economies based on either state command or nepotistic networks (Rhoads 1985). In their professional context, this attitude is close to a primitive normative assumption. Economists differ considerably in the extent to which they value equality among individuals and groups, the special protection of vulnerable populations, stewardship of the natural environment, and other objectives that are recurrent subjects of political controversy. But a person who did not derive special satisfaction from efficiency would make a terrible career choice, guaranteeing a life of boredom and under-motivation, if they set out to be an economist. And, crucially, the relevant kind of efficiency is not the strictly natural concept on which engineers focus;¹ it is *welfare efficiency* that depends on individual agents responding to specific incentives, based on expectations stabilised by institutions and laws. Of course, 'individuals' here need not be literally confined in scope to single human organisms; they can be, and in economics often are, households or other small units built on thick personal ties. But they cannot be coincidental interests too disassociated for strategic coordination.

If appreciation of that kind of efficiency is indeed a foundational normative attitude, then it logically cannot depend on a contingent empirical fact about how individuals' preferences arise and stabilise. I am not here suggesting that most economists think that it does. I am, however, suggesting that they teach as if it does; and furthermore that this teaching has distorted economists' selection of research agendas.

By 'descriptive individualism,' in the context of economics, I refer to the idea that individual agents – typically meaning individual people – form and maintain their preferences (particularly their risk preferences, the most important part of the general structure of utility in economic models) separately and independently. As an empirical description of actual people this is clearly largely false. Children are intensely socialised, and one of the key social mechanisms they must learn to operate is the belief-preference-choice-action model of

¹ The point here is made, amusingly, by the policy technician Smil (2021) complaining that economists have mangled the *true* concept of efficiency as developed by engineers. Indeed economists starting in the 1870s borrowed the basic idea, and changed it. That is of course always what happens with important concepts, few of which are born *de novo*.

everyday psychology. That is, they must learn to take the *intentional stance* (Dennett 1987) toward observed behaviour. Unless use of the IS becomes automatic (in the way that driving or reading does for most people), they cannot successfully coordinate choices with others. That children are not born conditioning their choices on individual preferences and beliefs represented in their heads is indicated by the fact that children learn to apply the IS to other people, plus play agents such as dolls, before they can start applying it to themselves (McGeer 2001, 2002), and thereby start narrating their distinctive personal profiles into being – under strong normative guidance from parents and then from cohort peers.

Economists of course know that people acquire their adult preferences and beliefs as they grow up. But they often write as if this is a fundamentally *private* process. In this they are encouraged by philosophers and psychologists writing about social coordination through *mindreading* (Nichols and Stich 2003) – processes by which people allegedly infer ‘latent’, inner mental states from observed behaviour, plus ‘tells’ such as facial expressions. Zawidzki (2013) argues persuasively that, outside of contexts in which people share deep and intimate histories, or are constrained by very exact and rigid institutional constraints on choice and attention to information, such mindreading is impossible because far too many hypotheses are always consistent with all of the available evidence. The main device that allows for human coordination is *mindshaping*, processes by which people put normative pressure on one another to develop preferences and beliefs that align with social expectations.

This does not result in ant-like populations because people *also* reward idiosyncrasies of style, taste, and knowledge – but within bounds.² You can demonstrate enjoyment of baseball but not football, or of Bowie but not Celine Dion, but you cannot decide, by yourself, that it is immoral to sell purple objects without facing the stern sanction of having people decline to enter joint ventures with you. The grounds for this would be that you evidently didn’t understand the nature of norm setting.

Though socialisation processes in children and adolescents are the largest and most salient arenas of mindshaping in a typical person’s life, people engage in it continuously, including at micro-scales. To successfully navigate through social life is to engage in endless negotiations about the detailed contents of networks of reciprocal intentional-stance profiles.

Economists’ utility functions, structured summaries of an agent’s observed choices under incentives, thus track constantly moving targets in the case of individual people. Since the very point of constructing, identifying, and estimating a set of utility function in game-theoretic microeconomics is to fix the structures of games, the fluidity and pervasiveness of mindshaping would seem to render the exercise – which amounts to quantising the intentional stance – a quixotic pursuit in the case of people. But in fact, it works strikingly well as long as temporal and circumstantial contexts are carefully controlled, allowing microeconomists to design countless mechanisms for refining small-scale market outcomes. (Success is far more elusive at macro scales.) What makes this possible are the baseline institutional rules and the competitive pressures constraining the choices that both intentional-stance ascriptions and utility functions consolidate. For example, there is some wiggle room in how a person conducts themselves as a hair stylist – and indeed, such creative branding is typically the only available source of competitive advantage in such quotidian services – but most of the hair-care professionals’ choices are forced by prices, norms, laws, on the supply side, and

² This has an evolutionary economic explanation: differentiation of individual preferences supports division of labour, with its crucial efficiencies (Martens 2004).

regularities on the demand side. If someone is selling packets of sugar the scope for idiosyncrasy is even less.

The basic source of stability and order in the domains that economists can best predict and control are thus not intrinsic properties of human atoms, but social structures in markets. The basic ontological foundations of actual economic behaviour are thus the opposite of atomistic.

In this context, consider the traditional pedagogical sequence of the undergraduate curriculum before CORE. The professor or lecturer typically began – following a template exemplified in Robertson (1957) – by focussing on Robinson Crusoe, alone on his island, choosing how to optimise along his demand curve by allocating his time between gathering lower-valued food (e.g. coconuts) and investing in capital formation to access high-valued food (e.g., making a fishing rod). Later, he meets Friday, and students are then introduced to a model of a 2-agent market with specialisation and trade. Reliance on this story dates back to the classical period, and was criticised by Marx (see Karagöz 2014), but here I will focus on the marginalist version of the Crusoe trope that dominated textbooks at and following the time of mass expansion of universities. In every generation, critics pointed out the peculiarity of beginning to teach the study of markets with a fictional scenario in which there is no prospect of exchange (e.g. Spearman 1966).

Economists did not ignore this criticism. One widely retailed response depended on the rise of representative agent models of general equilibrium (Varian 1992; for an early anticipation of this logic, see Koopmans 1957). The very point of such models is to base the distinctions between investment, saving, consumption, and production on schedules of activity allocation rather than on distinctions between social roles assigned to different agents. Robinson Crusoe before he meets Friday is indeed a fully consistent simplification of this representative agent. Such macroeconomics famously sidesteps problems of aggregation, and provides technical microfoundations for macroeconomic theory without any pretence to decomposing the economy into social atoms. This defence of Robinson Crusoe pedagogy cannot be accused of involving a blind spot with respect to atomism – it renders the implicit philosophical issue moot.

The problem with this response is that it implicitly calls on most economists to ignore their own broader opinions about economics itself. The majority of economists are microeconomists, and happy not to be targets of critiques such as that of Hoover (1988), to the effect that new classical macroeconomic models are not idealisations of real economies because their representative agents have no counterparts in reality *at all*.³ The Varian defence of Robinson Crusoe pedagogy requires us to suppose that most writers of introductory textbooks (before CORE) thought it sensible to begin to teach curricula dominated by microeconomics as if the point of it was to prepare students to learn macroeconomic models that none would ever encounter unless they went to graduate school *and* that most economists – that is, most microeconomists – *always* regarded as speciously avoiding the kind of empirically reasonable microfoundations that they respected.⁴

³ This is particularly obvious when solutions require them to be infinitely lived.

⁴ This was the attitude toward new classical macro models that was instilled in graduate students in microeconomics at my rigorously mainstream alma mater in the 1980s.

Beginning in the 1980s, microeconomics became increasingly dominated by game theory. In this context, Robinson Crusoe pedagogy is particularly misaligned with basic modelling practice, which solves for equilibria unintended by any particular agents, based on vectors of heterogeneous utility functions and, often, heterogeneous information among them. Robinson Crusoe, before Friday, plays games only against Nature; and no one starts to teach game theory that way. In fact, as Karagöz (2014) points out, the misfit between Robinson Crusoe economics and game theory was anticipated at the moment of the birth of game theory by von Neumann and Morgenstern (1944, pp. 10-12), and economists were explicitly reminded of this by Schotter (1991, p. 98).

Why, in light of this, did Robinson Crusoe pedagogy persist for decades beyond the game theory revolution in microeconomics, and continues to be taught in many precincts that CORE has not yet reached? I suggest two complementary diagnoses, one purely practical and one more interesting for present purposes because it points to the philosophical blind spot around atomism.

First, a practical justification is that many students, particularly in less selective universities, come to their first economics course poorly prepared to apply calculus to handle optimisation problems. Crusoe's time allocation problem between two activities is as basic and simple a case of such a model as one can present. One then shows him facing a more complicated case of such a model when Friday's strategic responses become part of his environment. The virtues of this pedagogical path are obvious. But it should be pointed out that the narrative very directly encourages atomistic social ontology: economic life is a simple challenge for the lone agent, but then complications arise when other people are added. These others are thus *obstacles* to utility maximisation, potential sources of thwarted achievement; furthermore, to manage this interference one must try to read their concealed minds. Amadae (2016) draws attention to the way in which such an approach to teaching game theory can frame it as reflecting a paranoid attitude to society.

The more interesting motivation for sticking with Robinson Crusoe is that the economist might not merely encourage the framing of game theory that worries Amadae as an unintended consequence. Instead she might have read, or absorbed, a philosophy of psychology that emphasises internal representations of preferences and beliefs, which are privately computed by individuals and must be inferred through mindreading. This need not come her way through any acquaintance with explicit philosophy of cognitive science. She might instead – particularly if, in her research, she designs and performs behavioural experiments – derive this philosophy from the practice of laboratory psychologists. As argued by Ross (2023), the dominant experimental methodology in individual psychology is to try to *isolate* the contribution of the 'basic' or 'raw' mind by shielding subjects from 'confounds' introduced by social structure and environmental scaffolds to learning. Many behavioural economists who view their activity as integrating economics with psychology adopt this design principle in their labs.

The alternative, to bring as much as possible of varying social structure *into* the lab, by multiplying treatments for every reasonably hypothesised effect of social influence or scaffolding, is also common in the experimental economics literature. But, strikingly, there is no ongoing explicit *debate* between these two fundamentally different ways of handling independent variables in social models when studying individuals' choices. My sense is that the primary determinants of which approach a lab favours are their budgets for incentive payments – since multiplying treatment groups invariably multiplies subject payouts – and

the econometric estimation methods in which research teams happen to specialise (Harrison, Lau, and Ruström 2015). Roughly speaking, blocking the presence of ‘confounds’ fits naturally with using regression models to test null hypotheses about conjectured effects, while running multiple treatments complements efforts to estimate marginal effects using structural or Bayesian models.

The absence of self-conscious debate between these two approaches to dealing with social influences in economists’ empirical research is, I suggest, the most persuasive evidence available of a philosophical blind spot that is reproduced in pedagogy. Among philosophers, the debate between internalist and externalist models of belief and preference – that is to say, models of *mind* – has been a dominant controversy for over three decades. Throughout this period, the ranks of the externalists have steadily grown, to the point where I would wager that they are now the majority. But almost no trace of awareness of the issue has seeped into the economics literature.⁵ It is therefore not surprising that it is also absent from economics classrooms.

In identifying a blind spot around atomism in current economics, then, I am not merely reiterating the obvious point that economic theory assigns core importance to individuals’ responses to incentives. That is natural in light of the discipline’s policy focus on welfare efficiency, and it implicates them in no special ontological commitments. Utility function assignments follow the logic of the intentional stance, even if few economists are aware of that. Though it can be and is extended to all agents, use of the intentional stance across the behavioural sciences is anchored in interpretations of the behaviour of individual people, so economists are not outliers in this respect. Individualism in this sense is reflected in standard game theory, where outcomes are defined in terms of the separate utility functions of players. This requires utility functions in extensive-form games to be treated as exogenous, since equilibrium solutions are inferred from availability of outcomes. Insofar as this is a limitation, making it challenging to model the strategic aspects of preference formation (what Ross [2005] calls ‘game determination’), this is best treated as a technical problem for which mathematical innovations have been proposed (Kaneko and Matsui 1999; Kaneko and Kline 2008, 2015; Fershtman and Segal 2018; Ross and Sterling 2020, 2023). In general, the nuances of individualism remain an important topic in the philosophy of economics (Zahle and Kincaid 2019) but economists have never been blind to them, and they are usefully approached only when advanced microeconomic theory is well understood – so, not in undergraduate curricula.

The blind spot is, rather, casual internalism about the causal basis of agency, the assumption that this lies in a black box for psychologists and neuroscientists to potentially open. If this is thought to be a commitment of economic theory, then it must cast doubt on the foundations of that theory for anyone who follows current debates in cognitive science. These may include undergraduates who take courses in AI, philosophy, or anthropology, and even the educated public, to whom externalism about beliefs and preferences has been competently introduced (Paul 2021). It often leads economic models to incorporate assumptions about representation that appear naïve to students exposed to disciplines that *do* study the mechanics of information processing in brains (Chater 2018). All of this encourages economists to resort to Friedman-style ‘as if’ accounts of their methodology when pressed by students who see the

⁵ Exceptions include Ross (2005, 2011, 2014a, 2014b, , Lecouteux and Larrouy (2017), and Harrison and Ross (2023). Moscati (2024) shows in a dismissive footnote (#9, p. 17) that he does not distinguish externalism from behaviourism.

problems, and this directly, and quite unnecessarily, undermines the scientific seriousness of the discipline, as well as gratuitously raising the stakes on predictive success. This is particularly problematic at the introductory level, since toy teaching models, by their nature, do not deliver accurate predictions.

Philosophy of preference and belief based on the intentional stance and externalism does not require abstruse concepts or complex arguments to explain.⁶ It can be outlined in a few slides about the ways in which children are socialised and the ways in which adolescent personalities evolve toward adult ones – topics which are hardly exotic to undergraduates. Dennett (2013) provides a sourcebook from which a few charming and lucid explanatory thought experiments can be selected. Crucially, this need involve no awkward digression from the introduction of basic economic concepts. Preferences have always been an early, explicit curriculum topic, and, as I have stressed, the intentional stance simply *is* the everyday, qualitative version of the kind of revealed preference theory that economists want students to operate. One can go from Dennettian intuition pumps to classic money pumps in the same lecture.

Caution should be expressed, however, about a mis-step that could easily be made here. The intentional stance is often glossed as characterising beliefs and preferences that *rationalise* observed behaviour. This is generally harmless outside the context of economics. Within that context, however, it is likely to interact with another philosophical blind spot.

Rationality

The idea that economic theory models ‘rational’ choice and behaviour is, self-evidently, pervasive in the culture of the discipline and in its pedagogy. Students are sometimes told that economists assume rationality in their models in order to identify optimal policy options. Elsewhere they may be told that rationality is an idealisation that, in some settings and conditions, people approximate. Where behavioural elements are introduced into a curriculum, the story may be that we begin by modelling rational agents in order to identify ‘anomalies’ for that assumption, then systematically relax it to generate ‘bounds’ for application to actual people.

Students are to some extent expected to intuitively understand what rationality is, but then they are also usually given a specification of it as it applies in a particular kind of model they are studying: agents make the choices that would maximise consumer or producer surplus in a market, or they allocate budgets so as to minimise expected consumption variance over time, etc.; and they observe some general principles such as not trying to recover sunk costs and revising their preferences if they are shown to be cyclical. Students may be taught some axioms that are taken to define rationality *in general*, but more often smaller sets of axioms specify what counts as rational in a *role*, such as worker, manager, or investor, or person shopping for a spouse or admission to a medical school.

Defining *general* rationality in the undergraduate curriculum is unwise, and economists can glance over at the work of philosophers to see why. Economists are interested in rationality as

⁶ Of course it *has* required deep and dogged work by philosophers of mind and cognition to establish both the conceptual foundations and the persuasiveness of these ideas. Economics teachers, if they adopt my suggestion, will be in debt to decades of philosophical labour. But in the economics classroom, the destination can be easily shown without any attention to the maps that allowed it to be reached, because the ideas in question are both readily exemplified and not counter-intuitive.

a constraint for identifying solutions to problems of resource allocation, but many philosophers want to know, exactly, what rationality, in general, *is*. They have not settled this. They distinguish between epistemic rationality – how to make inferences so as to tend to believe true things and disbelieve false things – and practical rationality, how to choose so as to maximise the chances of living the best life. In the nature of philosophical projects, neither of these is ever likely to end in general consensus, or cease to throw up new technical challenges. Furthermore, most philosophers agree that the two projects are deeply entangled with one another: a person qualifying as ‘generally rational’ would not spend resources piling up true beliefs, most of which would be insignificant, but instead aim for true beliefs that serve their practical ends in the most efficient way.

Philosophical theories of rationality, especially rational decision theory, share axiomatic and historical foundations with the Bayesian theory of choice under risk, classically developed by Savage (1954), that has become most influential in economics. There is a literature that seeks to merge them to generate a theory of ideal public policy; see Adler (2012) for a magisterial consolidation. Such affinities can create the impression that development of a unified account of what it is to be rational, to which specific economists’ models of rational choice ultimately refer, is rolling along hopefully.

Such confidence is specious once details are taken into account. Economists often use ‘rationality’ in special ways that philosophers never take seriously, as for example in some models of rational expectations used in macroeconomics where rationality involves taking account of all information that can be extracted from the history of prices in a market. On the other side, many philosophers, Adler being an example, hold that it is irrational for a policy-maker to violate risk neutrality, a view to which few economists assent, since this would entail over-ruling the revealed preference structures of the majority of people. The depth of cross-disciplinary misalignment is indicated by adding a bit more detail on the second issue. A leading philosophical decision theorist, Buchak (2013), disagrees with her colleagues where risk attitudes are concerned, and argues that it is rational to generally weight probabilities in accordance with rank-dependent utility (RDU) theory rather than expected utility (EU) theory. Harrison and Ross (2023), taking an attitude typical of economists, agree with Buchak that people often commit no errors and should generally not be ‘corrected’ by an economic advisor, when they manifest RDU preferences, but they reject the very idea of trying to declare one risk preference structure more rational than another.

It is likely that the majority of philosophers believe that many ultimate ends people maintain, and not just means to them, are ‘rationally’ evaluable. As an economist, I could not in good conscience write that sentence without the scare quotes.

The standard pedagogy of economists, along with their public discourse, implicitly treats all of these disputes as if they were so much technical noise around a penumbra of general acknowledgment that rationality is an important good thing, at least outside of artistic expression, emotional entertainment, and the glow of love, and we all know essentially what it is. I see no basis for this attitude that can survive careful reflection. The concept of rationality is genuinely, irreducibly polysemous (see Ross 2022, Veit 2024). For economists (but rarely for philosophers) it is usually synonymous with consistency of behaviour. But that is not the step toward a general analysis (for economists), because *what* an economic agent is modelled as being consistent *about* varies limitlessly across models.

It is true that economic agency, like agency in general, requires *enough* consistency in behaviour for some goal states to be identified. Economists typically demand sufficient regularity to be able to derive a (stochastic) utility function from observed *choice* behaviour, but usually only over restricted domains of outcomes.⁷ Completeness of preferences over all states of the world is a special, strong assumption seldom used except sometimes as a harmless technical convenience in some applications (see Bourgeois-Gironde 2020, Chapter 2, for a rich discussion).⁸ Similarly, there is sometimes motivation to model strategic interactions under an assumption that all players in a game are ‘Bayes rational’ (and perhaps that all know that all are ‘Bayes rational’). Again, however, the reference to rationality does no semantic work here; one can simply say that all players update according to Bayes’s rule as they receive information.

I have stressed that utility-function derivation is simply the quantised version of the everyday qualitative intentional stance. Might that not point the way to a general, core concept of rationality after all? Could we not say that an entity is rational just to the extent that, unless we assume the intentional stance toward it, we will lose explanatory and predictive purchase on some of its behaviour? We *could* say that if we liked, but then this would amount to massive deflation of the rationality concept in the eyes of most who use it. Notice that I used the clumsy idea of an ‘entity’ in the suggested statement – this was to avoid circularly saying ‘agent’ instead. That would be circular because a need to apply the intentional stance is precisely how we distinguish agents from non-agents. Thus the phrase ‘rational agent’ involves semantic redundancy: putting ‘rational’ in front of ‘agent’ adds nothing.

It might be thought that casual use of ‘rational’, even if it amounts to nothing more than acknowledging agency, is harmless in economic pedagogy and in public conversations about economics. But I think it does a great deal of harm. It leads economists to frequently sound like scolds when they pronounce some behaviour or policy ‘rational’, implying that alternatives are not (as opposed to having more specific problems such as being unstable or unaffordable). It indeed encourages many economists to forget that scolding is almost always politically self-defeating and leads people to resent economists. It usually amounts to overstatement, implying that a critical target is egregiously, hopelessly, wrong or bad when it is merely being said to be less than the wisest option with respect to some efficiency metric. It incites anti-economic passions from romantic rebels against alleged technocratic hijacking of social values (Coleman 2002). Finally, it leads people to see economists as appointing themselves to the role of technical experts on a topic, ‘rationality’, which many non-economists rightly doubt is an actually possible domain of expertise. This obscures economists’ *genuine* expert knowledge on how to estimate marginal effects of incentive changes and new sources of productive efficiency, and how to model and measure strategic behaviour.

I therefore urge the radical policy of expunging references to rationality from economics.⁹ It is quixotic to try to legislate public discourse, but an obvious path to reform runs through the

⁷ Revealed preference turned from an idea in the foundations of pure economic theory to a tool for application only once Afriat (1967) showed how to restrict the scope of the choice sets over which it was applied; see Ross (2014), pp. 63-67.

⁸ Hausman (2012) argues that economic preferences are always at least implicitly supposed to be complete. See Angner (2018) for a rebuttal based persuasively on economists’ practice.

⁹ Consistency suggests that I should dispense the same advice to philosophers – after all, I have argued that the rationality concept is polysemous for everyone, and it is philosophers who actively resist that fact, as opposed to just not noticing it like economists. But whereas I think that economists could do as I suggest by nothing

economics classroom and the textbooks prescribed there. We should stop telling students that our models assume ‘rational agents’ or ‘rational choices’. Rather, they simply assume agents, with properties that vary across models, and they identify these agents’ choices by carefully observing them through theoretical lenses that quantise the intentional stance. Calling them ‘rational’ is empty, self-aggrandising decoration. Let us stop suggesting to students that we are providing them with a special sagacity to penetrate the depths of ‘irrationality’ among the hoi polloi. Keynes (1936) famously urged economists to adopt the mien of humble technicians rather than engineers of society. A good step toward following that advice from the master is to stop spreading the verbal holy water of pronouncements about rationality.

Conclusion

I have deliberately avoided suggesting that explicit philosophy be woven into undergraduate teaching of economics. Shallow philosophy is usually worse than none, which makes the curricular space constraint tightly binding. Economists are not formed in philosophical cultures and most are unlikely to be comfortable breaking their pedagogical style to engage with students Socratically, as effective philosophy teaching requires. At the lecture podium, we economists do not work our way dialectically toward nuanced interpretations of concepts; we stipulate their meanings for our practical purposes using technical definitions.

Leaving explicit philosophy to philosophy courses, however, does not imply that the best economics pedagogy should reflect obliviousness to what philosophers’ debates have revealed about some basic economic concepts. Most are convinced, by good arguments based on scientific evidence, that preferences and beliefs are not private representations that must be inferred by mindreading. Given the centrality of preferences and beliefs to economic modelling, we should avoid presenting an interpretation of what they are that is at odds with what students will learn in other courses. Furthermore, economists should be gratified that philosophers of cognitive science have arrived at a view of preference, in particular, that is elegantly aligned with how economists have long thought of them. Revealed preferences are quantised versions of the kinds of states attributed in everyday contexts from the intentional stance. Teaching the intentional stance, therefore, requires no special philosophical digression.

Economists should also be aware of what is revealed by the history of philosophers’ efforts to analyse rationality in general. Rationality is an irreducibly polysemous concept that, abstracted from any special context of application, adds nothing to the idea of agency under the intentional stance. Using the concept in introducing economics therefore achieves only the bad result of building a professional rhetorical style that causes economists to be stereotyped as arrogant scolds. Arguably, it leads some young economists to in fact *be* arrogant scolds. If we have taught the concept of agency framed through the intentional stance, then we will get across everything we’ve intended to communicate by talking about rationality. That word should therefore become extinct in the economics classroom. Here, then, is a lesson from philosophy that doesn’t ask the economist to add any content to her syllabus; she must merely disengage a rhetorical habit.

References

more strenuous than breaking a habit, philosophers would have to come to consensus on a major philosophical issue, which has absolutely no prospect of happening.

- Adler, M. (2012). *Well-Being and Fair Distribution*, Oxford University Press.
- Afriat, S. (1967). The construction of utility functions from expenditure data. *International Economic Review* 8: 67-77.
- Amadae, S. (2016). *Prisoners of Reason*, Cambridge University Press.
- Angner, E. (2018). What preferences really are. *Philosophy of Science* 85: 660-681.
- Bourgeois-Gironde, S. (2020). *The Mind Under the Axioms*, Academic Press.
- Buchak, L. (2013). *Risk and Rationality*, Oxford University Press.
- Chater, N. (2018). *The Mind is Flat*. Allen Lane.
- Coleman, W. (2002). *Economics and its Enemies*. Palgrave Macmillan.
- Dennett, D. (1987). *The Intentional Stance*. MIT Press.
- Dennett, D. (2013). *Intuition Pumps and Other Tools for Thinking*, Norton.
- Fershtman, C., & Segal, U. (2018). Preferences and social influences. *American Economic Journal: Microeconomics* 10:124-142.
- Hausman, D. (2012). *Preference, Value, Choice, and Welfare*, Cambridge University Press.
- Harrison, G., Lau, M., & Rutström, E.E. (2015). Theory, experimental design, and econometrics are complementary (and so are field experiments). In G. Frechette & A. Schotter, eds., *Handbook of Experimental Economic Methodology*, Oxford University Press, pp. 296-338.
- Harrison, G., & Ross, D. (2023). Behavioral welfare economics and the quantitative intentional stance. In G.W. Harrison & D. Ross, eds., *Models of Risk Preferences: Descriptive and Normative Challenges*, Emerald, pp. 7-67.
- Hodgson, G. (2015). *Conceptualizing Capitalism*, University of Chicago Press.
- Hodgson, G. (2023). *The Wealth of a Nation*, Princeton University Press.
- Hoover, K. (1988). *The New Classical Macroeconomics*, Blackwell.
- Kaneko, M., & Kline, J.J. (2008). Inductive game theory: A basic scenario. *Journal of Mathematical Economics* 44: 1332-1361.
- Kaneko, M., & Kline, J.J. (2015). Understanding the other through social roles. *International Game Theory Review* 17:
<https://www.worldscientific.com/doi/10.1142/S0219198915400058>
- Kaneko, M., & Matsui, M. (1999). Inductive game theory. *Journal of Public Economic Theory* 1: 101-137.

- Karagöz, U. (2014). The neoclassical Robinson: antecedents and implications. *History of Economic Ideas* 22: 75-100.
- Keynes, J.M. (1936). Economic possibilities for our grand-children. Re-printed in Keynes, *Essays in Persuasion*, Norton (1963), pp. 358-373.
- Koopmans, T. (1957). *Three Essays on the State of Economic Science*, McGraw-Hill.
- Lecouteux, G., & Larrouy, L. (2017). Mindreading and endogenous beliefs in games. *Journal of Economic Methodology* 24: 318-343.
- Marglin, S. (2008). *The Dismal Science*, Harvard University Press.
- Martens, B. (2004). *The Cognitive Mechanics of Economic Development and Institutional Change*, Routledge.
- McGeer, V. (2001). Psycho-practice, psycho-theory, and the contrastive case of autism: How processes of mind become second nature. *Journal of Consciousness Studies*, 8, 109–132.
- McGeer, V. (2002). Enculturating folk-psychologists. *Synthese*, 199, 1039–1063.
- McLaughlin, B. (1992). The rise and fall of British emergentism. In A. Beckermann, H. Flohr & J. Kim, eds. *Emergence or Reduction?: Essays on the Prospects of Nonreductive Physicalism*, De Gruyter, pp. 49-93.
- Moscatti, I. (2024). Psychological narratives in decision theory: what they are and what they are good for. Working paper: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4769060
- Nichols, S., & Stich, S. (2003). *Mindreading*, Oxford University Press.
- Paul, A. (2021). *The Extended Mind*, Mariner.
- Quinn, J. (2024). *How the World Made the West*, Random House.
- Rhoads, S. (1985). *The Economist's View of the World*, Cambridge University Press.
- Robertson, D. (1957). *Lectures on Economic Principles, Volume 1*, Staples.
- Ross, D. (2011). Estranged parents and a schizophrenic child: choice in economics, psychology and neuroeconomics. *Journal of Economic Methodology* 18: 215-229.
- Ross, D. (2012). Economic theory, anti-economics and political ideology. In U. Mäki, ed., *Handbook of the Philosophy of Science, v. 13: Economics*, Elsevier, pp. 241-285.
- Ross, D. (2014a). *Philosophy of Economics*, Palgrave Macmillan.
- Ross, D. (2014b). Psychological versus economic models of bounded rationality. *Journal of Economic Methodology* 21: 411-427.

Ross, D. (2022). Neo-Samuelsonian methodology, normative economics, and the quantitative intentional stance. In B. Caldwell, J. Davis, U. Mäki, & E.-M. Sent, eds., *Methodology and History of Economics: Reflections With and Without Rules*, Routledge, pp. 90-116.

Ross, D. (2023). Economics is converging with sociology but not with psychology. *Journal of Economic Methodology* 30: 135-156.

Ross, D., & Stirling, W. (2020). Economics, social neuroscience, and mindshaping. In J. Herbecke & C. Herrmann-Pillath, eds., *Social Neuroeconomics*, Routledge, pp. 174-201.

Ross, D., & Stirling, W. (2023). Mindshaping, conditional games, and the Harsanyi doctrine. CEAR Working Paper 2023-03: <https://cear.gsu.edu/wp-2023-03-mindshaping-conditional-games-and-the-harsanyi-doctrine/>

Savage, L. (1954). *The Foundations of Statistics*, Wiley.

Schotter, A. (1992). Oskar Morgenstern's contribution to the development of the theory of games. *History of Political Economy* 24: 95-112.

Smil, V. (2021). *Grand Transitions*, Oxford University Press.

Spearman, D. (1966). *The Novel and Society*, Routledge & Kegan Paul.

Udehn, L. (2001). *Methodological Individualism*, Routledge.

Varian, H. (1992). *Microeconomic Analysis*, 3rd Edition, Norton.

Veit, W. (2024). Rationality and intransitivity. *Croatian Journal of Philosophy* 24: 273-292.

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behaviour*, Princeton University Press.

Zahle, J., & Kincaid, H. (2019). Why be a methodological individualist? *Synthese* 196: 655-675.

Zawidzki, T. (2013). *Mindshaping*, MIT Press.