Intrinsic Goodness

Why We Might Behave Better Than We Think

Antara Haldar

The plot of mainstream economics has been monopolized for generations now by the fictional character known as *Homo economicus*. *Homo economicus* has the cognitive capacities of a superhero: his ability to churn unlimited information and unflinching self-knowledge into instant and accurate decisions is infallible. But he differs from your everyday superhero in one key respect: he is in no way committed to using his powers to do good. His calculation of cost and benefit to relentlessly maximize utility is entirely clinical, and he is laser-focussed on his own material self-interest rather than fluffy considerations such as social moorings. Of course, we almost never encounter such people in the real world. Yet the dominant strain of what is known as "rational choice theory" – the model that lies at the heart of mainstream economics – has placed this unsympathetic antihero squarely at the centre of economic thinking, making rationality in effect synonymous with selfishness.

In order to avoid cognitive overload – a sort of bandwidth problem – we take mental shortcuts.

Happily, in recent years, we have started to glimpse what the arc of character development of the agent at the centre of economic analysis might look like – and how it might eventually evolve beyond *Homo economicus*. Beginning with the collaboration – now chronicled by Michael Lewis in The Undoing Project – of the psychologists Daniel Kahneman and Amos Tversky, the stick figure that had for so long populated economics has started to be filled out. Their work helped spawn a whole new field – behavioural economics – that identifies a plethora of instances where the standard assumptions of economic theory are shown not to operate. In particular, human beings appear to be both computationally inferior (lacking both complete information and self-control) and morally superior (motivated by things other than material self-interest) to the robotic *Homo economicus*.

Behavioural economics brings to the fore the fact the actions of real human beings are more flawed and faltering – replete with what mainstream economists would call "irrationality" – than the unerring mechanical logic of *Homo economicus*. In order to avoid cognitive overload – a sort of bandwidth problem – we take mental shortcuts. Psychologists call these "heuristics and biases" (terms popularized by Kahneman's *Thinking Fast and Slow*) and of-

ten they don't lead to the best decisions. We value what we have more than we value what we don't; we form attachments; we prioritize the present over the future; we get angry and aroused; we feel exhausted and frustrated; we forget to pay bills; we underinvest in insurance. These inconsistencies and foibles are ubiquitous features of human action.

Often called the "dismal science," even by its practitioners, economics has been relatively quick to accept the lapses of erring humans. The discipline is struggling, however, with the more revelatory findings of behavioural research – that most real people are also in many ways superior to the cardboard *Homo economicus*. We appear to be predisposed to feeling good about behaving well to others. We resent bad behaviour in others. We make and, perhaps surprisingly often, keep promises. We care (deeply) about fairness. Most of all, we are acutely sensitive to the approval and disapproval of others. In evolutionary terms, this all makes sense. The roots of these social, or co-operative, tendencies – even beyond kinship networks – arguably emerged as a matter of necessity in situations where individual and group survival were intimately connected. Scientists including E. O. Wilson have long studied this phenomenon, called "eusociality," in species ranging from ants and wolves to birds and bees. In humans, this is described as "prosocial" behaviour – and appears to be similarly deep-rooted. The research of the economist Ernst Fehr builds on this to show that "social preferences" or forces like altruism, reciprocity, pleasure in helping others, aversion to inequality, and ethical commitments seem to be hard-wired – and that cultural evolution has also tended to reinforce co-operative tendencies. To put it controversially, people appear to be inherently moral – or at least to possess a kernel of intrinsic goodness.

Homo economicus has been defended by those within the prevailing school of thought – neoclassical economics – as a device that, like so many of the discipline's other simplifying assumptions, has allowed economics to attempt to become a quantitative science: an abstraction tracing out the locus of the optimal path of an actor through the labyrinth of the economy – much as basic geometric shapes form the basis for understanding the much more complex physical realities of the world around us. But there is something different between the social and the physical worlds. The course of a falling ball is unlikely to be altered by our description of gravity, but our understanding of our own actions shapes those actions themselves.

The fact that we are the product of our passions is of fundamental economic importance. It justifies the need for interventions to save us from ourselves. This is the stuff of the economist Richard Thaler and the legal theorist Cass Sunstein's bestselling *Nudge*, which first appeared ten years ago: laws that will help us, for instance, to eat better (e.g., a tax on sugar) and save more (e.g., a compulsory pension contribution). But, equally, it implies that people often act in ways that conventional economic theory finds hard to predict. Examples range from people, paradoxically, donating less blood when they start being paid to do so, to parents leaving their children at a daycare centre longer after fines are imposed, to firefighters starting to take more leave when financial penalties are introduced. This is referred to as "crowding out" – when an interaction becomes transactional it ceases to engage people's finer feelings. These instances exemplify the fact that if we start out with the wrong assumptions about human nature, economic policies could radically misfire – or, worse still, that assuming the worst about people can make them behave badly. As the philosopher Michael Sandel has pointed out, monetization changes the character of the act and, hence, the nature of the psychological reward from engaging in it. The "helper's high" or guilt at transgressing a social rule, or the sense of identity that one derives from one's job, seem often to be more effective in driving action than a profit-and-loss calculation: morality appears, at least in certain key instances, to be a more powerful motive than money.

My own research illustrates this in the context of a Nobel Prize-winning social experiment: microfinance. The scheme involves providing small loans to people in underprivileged com-

munities traditionally denied finance by formal banks – originally women in rural Bangladesh, and then around the globe. The problem of how to persuade formal financial institutions to provide credit to those who need it most had long vexed conventional economic thinking – the size of the loans was too small relative to the costs of collection and, potentially, litigation to make it profitable. Microfinance – arguably the biggest, if unwitting, behavioural economics experiment ever conducted – found an innovative solution: by replacing economic incentives with emotional engagement, it managed to get a billion of the world's poorest women to repay loans almost 100 per cent of the time.

Defying the received economic wisdom, microfinance institutions neither drew up written contracts nor demanded collateral from borrowers so the incentives were not financial or legal in the traditional sense. Nor could the high repayment rates be entirely explained by more sophisticated accounts of strategic considerations like fear of losing access to future credit or even transferring the cost of monitoring loan use from the bank to other borrowers – the women kept repaying loans even when it ceased to be a matter of prudence (for instance, when the continued existence of lenders was threatened by calamities such as floods). Indeed, traditional economic scholarship is at a loss to explain the success of the system – and a mutant version of microfinance that attempted Starbucks-style scaling in India and treated the loans as purely commercial resulted in mass default. By assuming the best of human nature, microfinance succeeded in eliciting better behaviour. It can best be understood as a system that effectively tapped into good old-fashioned values: pride and shame and guilt and loyalty. Psychologists call this the moral machinery. Contrast this with traditional Wall Street-style finance that removes pride at behaving well and shame at misbehaving from the equation and eulogizes material aggrandizement above all else.

The contributions of behavioural economics were recognized last year with a Nobel Prize for Richard Thaler. This was the second Nobel award for the field: Kahneman won the prize for economics in 2002 without, as he jokes, having taken a single course in the subject. Popular books such as Dan Ariely's *Predictably Irrational* and Malcolm Gladwell's *The Tipping Point* have achieved almost cult status. Governments around the world – notably in the UK and US – are increasingly incorporating behavioural thinking into policy initiatives.

But has behavioural economics dealt *Homo economicus* the fatal blow? The evidence suggests not. While Thaler and Sunstein may have achieved a miracle by helping dislodge chocolate bars from some school lunches, they have – sadly – not yet managed to have the same impact on the core assumptions of economic theory. The contributions of behavioural economics have primarily led to the launch of what we might call the "new and improved" *Homo economicus* 2.0. In the instances that count, mainstream economics clings firmly to the belief that economic decision-making is both mechanistically algorithmic and, motivationally, driven by greed and selfishness. As the political scientist Margaret Levi, Director of Stanford's Center for Advanced Study in the Behavioral Sciences (where behavioural economics was first developed) puts it: "Instead of changing the paradigm of economics by challenging its fundamental methodology or questions embedded in prevailing conceptions of micro and macro economics, behavioural economics applies its insights to where nudges work". So rather than spur a revolution, it has settled for influence that is at best incremental.

While almost all major universities now consider it fashionable to offer a few courses on behavioural economics, it is treated as an addendum to the standard curriculum in mainstream microeconomics. Breakthroughs in the understanding of the cognitive contours of economic agents could have potentially profound implications for various domains of economic analysis – ranging from labour to development to inequality and finance. Yet, instead of stimulating a fundamental theoretical shift, it has been dealt with as an amalgam of quirky anecdotes about small-scale aberrations from gen-

eral models that largely cancel themselves out in the aggregate.

The assumption that standard economic theory is perhaps most reluctant to relinquish centres on material self-interest as the main fount of human action. This may be attributable to the long history of the concept in economics. Bernard Mandeville's *The Fable of the Bees* (1714) is a cautionary tale eulogizing individual self-interest as the most effective means of producing collective benefit. This is an idea encapsulated in Adam Smith's famous remark that "it is not from the benevolence of the butcher, the brewer, or the baker we expect our dinner, but from their regard to their own self-interest."

During the presidential debates of 2016, Donald Trump admitted that he had actively hoped for the crash of 2008, adding "that's called business, by the way." This remark sums up one dominant attitude towards economic activity as a parallel moral universe: a space untarnished by the basic constraints of human decency. The behavioural sciences show that this need not be so: drawing a sharp line between the economic and social is artificial. Not only is the assumption that greed and profit are the only spurs for human action descriptively incorrect, but as the economist Samuel Bowles has demonstrated graphically in *The Moral Economy*, assumptions about self-interest become self-fulfilling prophecies.

Donald Trump admitted that he had actively hoped for the crash of 2008, adding "that's called business, by the way." This remark sums up one dominant attitude towards economic activity as a parallel moral universe: a space untarnished by the basic constraints of human decency.

So why, then, does *Homo economicus* keep being resurrected? In large part, it is because mainstream economics derives its power from two things. First, the assurance of rigour and certainty provided by its mathematical models; and second, the alchemy of converting individual material self-interest into collective good through value-neutral institutions such as prices. The behavioural approach challenges both of these.

Of course, it is possible to fold behavioural insights into the dominant methodology: the respect that neoclassical economics accords individual choice is so great that preferences can include almost anything. Mainstream economics is content to shoehorn into its frame incremental alternative accounts of preferences, even ones it considers highly esoteric like caring about others or the climate.

But this misses the point. The notion of "preference" doesn't fully capture what the behavioural and cognitive sciences are telling us about human nature: that people possess values that are not mere whims, but deeply ingrained; constitutive of who they are. As the philosopher Richard Holton has put it,

> …economists sometimes talk as though it is just a question of extending the range of things that we have preferences for so that we don't just have preferences for our own well-being, but also for the well-being of others. But just as putting a price on something can change how we think of it, so seeing our values as mere preferences can change how we think of them. Values are fundamentally different: their authority doesn't come from the fact that we prefer them. We can't fit them into a scale of preferences without distorting what they are.

If people do, indeed, have deeper principles and not merely prudential preferences, that prem-

ise could be the basis for an entirely new paradigm in economics. But to incorporate more fully the insights from behavioural economics would entail the discipline going back to the drawing board: not only requiring it to rethink its seductively simple methodologies but also posing a threat to economics' very conception of itself. The epistemic hubris of neoclassical economics stems from its seeing itself as a value-neutral pure science, like physics. Behavioural analysis muddies the waters by introducing not only emotions, but also ethics into the picture.

The behavioural sciences may not have all the answers yet, but they have unquestionably discovered a chink in the epistemic armour of economics. They could well be what we need to save ourselves from the oversimplifying assumptions and obtuse abstractions of the current mode of economic analysis. By establishing that individual and collective interest need not be at odds, they show that what economists call the "tragedy of the commons" is not the inevitable fate of all social enterprises. A paradigm of economics centred on an economic agent capable of civic friendship – rather than on the pathological *Homo economicus* – may be exactly what we need to salvage the future of our broken societies and a planet on the precipice of disaster. This is no longer a mere philosophical trope, but is increasingly being established as an empirical reality. Until economics agrees to its script being rewritten, its plot is doomed to remain a tragedy.•

This article first appeared in *The Times Literary Supplement*, November 3, 2018, and is reprinted here with permission.

Books consulted in the writing of this piece include:

Elinor Ostrom, *Governing the Commons: The evolution of institutions for collective action* (Cambridge University Press, 2015).

Mark Granovetter, *Society and Economy: Framework and principles* (Harvard University Press, 2017).

Michael Lewis, *The Undoing Project: The friendship that changed our minds* (Penguin, 2017).

Michael Sandel, *What Money Can't Buy: The moral limits of markets* (Penguin, 2013).

Samuel Bowles, *The Moral Economy: Why good incentives are no substitute for good citizens* (Yale University Press, 2017).

***Antara Haldar**, PhD, teaches jurisprudence and law and economics at the University of Cambridge, where she holds the inaugural post in Empirical Legal Studies. She is currently Visiting Faculty at Harvard. She spends a lot of her time in cafes in London and New York, thinking about how cooperation works and plotting the epistemic revolution.*