

The Self According to Others: Explaining Social Preferences with Social Approbation

Oswin Krüger Ruiz, University of Oxford, UK
oswin.krg@gmail.com

Abstract

In past decades, significant work in behavioural economics has decisively revealed the limitations of the human agency model known as *Homo Economicus*, whereby humans are purely driven by material self-interest.¹ These behavioural findings are, however, far from integrated in mainstream economic theory, which builds heavily on the neoclassical tradition. Unbeknown to modern economics, Bernard Mandeville and Adam Smith already proposed a richer model of human agency in which choices also depend on the desire for social approbation. The social approbation mechanism complements material self-interest and provides a more diverse toolset, which is able to explain social preferences.² Mainstream economic agency confines the study of human action to an artificially-limited spectrum because it reduces society to atomistic individuals who maximise one all-purpose measure of value: utility, which is often instrumented by consumption. Collective action is therefore only sustainable where material incentives are in place, as the economic agent rides for free unless financially penalised. To explain pro-social behaviour from the standpoint of self-interest, Mandeville and Smith proposed that agents also maximise social approbation, which conveys incentives to act pro-socially because the desire for others' approval encourages compliance with social norms. The upshot for collective action is that, assuming social norms represent common interests, approval from others provides an extrinsic motive for pro-social behaviour. I formalise the mechanism by proposing a simple utility function in which agents maximise social approbation as well as material self-interest.

JEL codes: B41, D16, B12, D91, B50

1. Failing to Account for Social Preferences

The field of economics has long been subject to the criticism that it models human agency in a completely unrepresentative manner by not allowing for the possibility of social preferences. Its simplistic assumptions, often called the *Homo Economicus*, presuppose that humans only seek to maximise their own material payoffs driven by pure material self-interest. This view of human agency has a long-lasting tradition in economic thought; previously Bernard Mandeville and Adam Smith constructed economic propositions of how selfish individuals could potentially generate social benefits in a market context.³ Through Smith's famous invisible hand, selfish agents would generate social benefits by following their 'propensity to

¹ Material self-interest as regularly modelled in microeconomics: a human agent only cares about maximising utility, which is generally represented by individual material payoffs (i.e. consumption of goods and services).

² Preferences to act pro-socially beyond individual material payoff, in the sense of Ernst Fehr and others behavioural economists.

³ *Ibid.*, i.28.

truck, barter, and exchange one thing for another'.⁴ The supposedly selfish individual depicted in *The Wealth of Nations* (WN) by Adam Smith, which ignored the more comprehensive account of human behaviour in *The Theory of Moral Sentiments* (TMS), quickly became – and thanks to its methodological convenience still is – the primary model of human agency in economic analysis.⁵

Despite its prevalence in modern economic textbooks, *Homo Economicus* was already seen as a limited model of agency from its conception. Even Mandeville, who greatly praised the economic gains of self-interest, failed to abstract these benefits from a social and political setting.

'So Vice (self-interest) is beneficial found; When it's by Justice lopt and bound' (Mandeville, 1988[1732], i. 36-7).

Mandeville is hinting at the fact that a market of selfish agents cannot be decontextualised from the rest of society – for example, politics. Adam Smith, on the other hand, was more concerned with the known difficulty of modelling human agency, as 'the source of human motivation... cannot be easily reduced to a few simple principles'.⁶ Beyond the classics, economists of all ages have agreed that the *Homo Economicus* painted a highly incomplete picture of human agency. However, these assumptions greatly facilitated the mathematical modelling that economists sought to achieve, often to emulate the natural sciences.⁷ These economic models create an abstract world that might be consistent and render useful insights without corresponding to reality.⁸ Yet while internal validity is warranted by mathematical logic, external validity relies on the appropriateness of the underlining propositions to the context a model is being applied to.

Economists have hence long been aware of the limiting assumptions surrounding *Homo Economicus*, even if those doubts did not always extend to the insights derived from it. Yet even ignoring these moments of self-awareness, the last two decades of behavioural field work and randomised controlled experiments have been able to fully falsify the *Homo Economicus*. This work has undermined the generalisation that human behaviour can be studied predominantly by reference to selfish material incentives.⁹ Beyond material self-interest, humans can be 'conditional co-operators and altruistic punishers', clearly revealing social preferences, as evidenced by multiple empirical experiments conducted within the realm of behavioural economics.¹⁰ According to Ernst Fehr, prominent examples include instances in which 'people vote, pay their taxes honestly, participate in unions and protest movements', which often lack a discernible material payoff to the individual.¹¹ Offering a potential explanation, survey data collected by Kahneman and Deaton indicates that extra income beyond the threshold of \$75,000 per annum (based on an average in the US) quickly decreases in added value to the individual.¹² Social preferences, caring for more than just oneself or being incentivised by more than material gains, have been thoroughly documented by the work of Daniel Kahneman, Ernst Fehr, Samuel Bowles, Dan Ariely, and other behavioural economists.

⁴ Smith, *WN*, I.ii. p.1.

⁵ Laffont and Marmimort, 2001, p.18.

⁶ Mehta, 2006, p. 249.

⁷ Bodenhon, 1956, pp.25-32.

⁸ Morgan, 2012, p.405.

⁹ Sassower, 2010

¹⁰ Fehr, 2005, xii.

¹¹ Fehr and Schmidt, 1999, p.818.

¹² Kahneman and Deaton, 2010.

The documented existence of social preferences poses a significant challenge to conventional economic modelling and, more importantly, the benefits of market behaviour carried out by purely selfish agents who are only maximising material payoffs. Smith and Mandeville had some answers to this challenge that might have been previously overlooked.

2. Instrumenting Social Preferences with Social Approbation

Smith's account of human agency was much richer than is often depicted. While Mary Morgan considers it too multifaceted for modelling, a first step is to incorporate the motive of social approbation. Both Smith and Mandeville departed from the tensions between self-interest and altruism to reach conclusions that looked beyond the dichotomy. Their thoughts indicate that human behaviour is driven by the desire for social approbation just as much, if not more, than by material self-interest. For this, they presumed that people sympathise and hence approve of each other. Society approves of actions that conform to social norms, turning the desire for social approbation into a motivation for social preferences. In other words, Smith's and Mandeville's agents gain positive utility by being approved of socially.

2.i Breaking the Dichotomy

The 18th century intellectual discussions of society revolved around depictions of people as either egoists or pro-social beings, cementing a dichotomy that humans are either one or the other. Economics found materialistic egoists easier to model in a market context, clearly delighted by the apparent paradox of deriving common good from selfish behaviour. Embedded in the foundations of economics is the belief that ethical and social considerations are beyond its scope. It can be questioned, however, whether the selfish-selfless dichotomy is even useful. In its barest form, self-interest simply indicates the interest of an individual agent, her reason-to-act so to speak, which could be directed towards herself (egoism) as well as towards others (social preferences).

This notion of neutral self-interest can, admittedly, be regarded as a driver of any form of behaviour. A contemporary critic of Mandeville, Joseph Butler, argued that people could place self-interested value on morality, rendering morality and self-interest compatible.¹³ This approach reinforced one of Mandeville's crucial contributions. It is pointless to categorise human activity by distinguishing between selfish and selfless choices, if both can always be traced back to some form or other of (neutral) self-interest. I am choosing to help another selflessly because I want to, meaning by definition that it is in my (self) interest. According to Mandeville, 'to judge a man's performance' we need to know 'the principle and motive from which he acts'.¹⁴ Yet the fact that generosity can be interpreted as a genuine impulse to help others, as with Rousseau's pity-mechanism, it can also be considered 'ambition in disguise', as expressed by La Rochefoucauld,¹⁵ shows the futility of trying to judge decisions on a normative categorisation of intentions.

Discoveries in modern neuroscience corroborate this view. The human brain displays activity in the prefrontal cortex when confronted with material incentives, an area associated with cost-benefit and consequentialist thinking and mainly concerned with 'deliberative processes'.¹⁶ In contrast, the limbic system deals with behaviour that is considered impulsive. Speaking against the egoism-altruism dichotomy is the fact that this part of the brain is

¹³ Internet Encyclopedia of Philosophy: A peer-reviewed academic source. Accessed 15/08/2016, <http://www.iep.utm.edu/mandevil/>

¹⁴ Mandeville, 1988[1732], i.43.

¹⁵ Force, 2003, p.137.

¹⁶ Bowles, 2016, pp.103-110.

activated when engaged in collaborative activity, such as ‘conformity of an action with a set of rules’, yet also with activity regarded to be driven by selfish motives.¹⁷ The distinction between selfishness and selflessness breaks, as it rests, on ambiguous intentionality.

Revealing a false dichotomy between egoism and collaboration is a relevant challenge to the theoretical apparatus of economics. Even Edgeworth distinguished between economics and ethics on the basis of self-interest and selflessness respectively, suggesting these could be separated.¹⁸ It is as if economics simply excluded the existence of any form of collaboration that did not fit within the realm of marked exchange between selfish individuals. As a result of its limited conception of agency, neoclassical economics has been shaped as a science that studies only the self-interested behaviour of humans, leaving out important aspects of social interaction. When studying human economic activity it would be unwise to not consider all the motives that influence human choice. There is thus a need to look beyond the false dichotomy that justifies a model of human agency driven solely by material self-interest.

2.ii Social Approbation Maximisation

Numerous social scientists argue that the borders between the self and others are not that clear-cut. Humans are methodologically indivisible from one another, partly a distinctive entity and partly a component of a greater social structure, their identity and choices influenced by their surroundings. It is therefore necessary to incorporate influences arising from a social context in any model of individual utility-maximisation. Since people reveal social preferences, there must be a behavioural mechanism to explain it. Smith and Mandeville already identified this issue and proposed a solution that has received little attention. They argued that individual agents seek to be approved of by society, which creates a mechanism that allows for social influences. When acting pro-socially, driven by expectations of receiving social approbation, agents would be maximising their individual utility.

In Smith’s complex conception of agency, humans are driven by the ‘great purpose of human life which we call bettering our condition’.¹⁹ In their critique of expected utility theory, Kahneman and Tversky advanced the idea that ‘value is assigned to gains and losses rather than to final assets’ because ‘our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes’.²⁰ In Kahneman’s prospect theory, which fits remarkably well with Smith’s idea of dynamic betterment, value is a function of two things: the ‘reference point’ and the ‘magnitude of the change’ from that point.²¹ Humans care about improvements (change) beyond their current state (reference point).

Smith also argued that humans are driven by an ‘innate desire to receive the approbation of others’.²² His views on social regard bear resemblance to Rousseau’s conception of *amour-propre*, recognising the idea that agents judge improvement on the basis of a positive change relative to others. Assuming that the reference point is some measure of the average recognition of the people around us, our valuation would be directly correlated to the difference between that point and our own position. Human agency therefore seeks improvement relative to two different reference points: 1) the agent’s current state in time and 2) what the agent perceives as the average reference point of her social context.

¹⁷ *Ibid.*

¹⁸ Force, 2003, p.171.

¹⁹ Smith, *The Theory of Moral Sentiments* [henceforth *TMS*], 1790, I.iii.2.1.

²⁰ Kahneman and Tversky, 1979, p. 263.

²¹ *Ibid.*, p. 277.

²² Paganelli, 2008, pp.80-91.

While Rousseau argued that ‘all equality vanished’ as a result of this constant search for improvement relative to others,²³ Mandeville saw the resulting desire for social esteem as a motive for collaborative activity. Positioned in between both views, Smith seemed to consider that social approval can be both the object of ‘congratulations’ and ‘envy’.²⁴

Regardless, in their quest for improvement, Smith’s and Mandeville’s agents pursue social approbation as well as wealth. People could be incentivised to act pro-socially through the desire for social approbation, which signals improvement in the eyes of others.

Similar to the research of Kahneman and Deaton, Smith acknowledges diminishing returns to consumption after a satiation point of wealth, when ‘the necessities of nature... (which) the meanest labourer can supply’ are covered.²⁵ As consumption is bounded by satiation, he reasoned, wealth acquisition shifted from having an intrinsic value to an instrumental one.

‘the rich man glories in his riches, because he feels that they naturally draw upon him the attention of the world’ (Smith, 1790, *TMS*, I.III.ii.1).

Smith argued that wealth is accumulated beyond the point of which it provides material benefits. Instead, wealth accumulation is a way of drawing the admiration of others or, in other words, seeking social approbation. Nevertheless, even if wealth accumulation ‘constantly obtain(s)’ social recognition, it is not the only behaviour conducive of social distinction.²⁶ An agency model that incorporated social approbation maximisation would, according to Smith’s propositions, be more comprehensive than instrumenting the human quest for betterment solely by wealth accumulation.

Smith’s behavioural model therefore isolates the desire for social approval as a central motivation, alongside the long-run benefits it might render. In other words, social approbation has intrinsic value.

‘[Humans seek to] be observed, to be attended to, to be taken notice of with sympathy [as well as] all advantages which we can propose to derive from it’ (Smith, 1790, *TMS*, I.III.ii.1).

According to Smith, receiving sympathy – positive approbation – constitutes a direct way of measuring our own social worth. The desire for sympathy combines with our desire for bettering our condition, leading to what Smith calls vanity.

‘[Vanity] involves the constant comparison between ourselves and others, and constant computation of the ways in which we could improve our position in the eyes of others’ (Force, 2003, p.132).

While wealth constitutes one of these ways, social approbation provides a direct measure of the value of our behaviour in the eyes of others. Improvements relative to others can be directly inferred from the approval and admiration received from others.

Smith further explains how, through social approbation, society influences the actions of individuals, providing a way to reconcile individual self-interest with pro-social behaviour.

²³ Rousseau, 1754.

²⁴ Smith, *TMS*, I.II.v.1.

²⁵ *Ibid.*, I.III.ii.1.

²⁶ *Ibid.*, I.III.iii.3.

'Bring him into society, and all his own passions will immediately become the causes of new passions. He will observe that mankind approves of some of them, and are disgusted by others. He will be elevated in the one case, and cast down in the other' (Smith, 1790, *TMS*, III.I.3).

As a result, Smith believed that the self is formed within society, raising doubts about whether individuals can be isolated. The individual desire for approbation means that agents will seek to comply with social norms. It is through this mechanism that our desire for social approval can interact with, or even oppose, material self-interest. As long as social norms endorse common goals, this mechanism can incite pro-social behaviour. While it lies beyond the scope and aim of this paper, it should be noted that social convention could also sanction socially-harmful behaviour.

Thinking about the supply-side, a neoclassical economist might at this point question why, from a standpoint of individualistic maximisation, anyone would praise or applaud the activity of others, a necessary condition for the social approbation mechanism. Smith himself appealed to human nature, proposing that individuals have a propensity to sympathise with fellow humans,²⁷ an idea that closely resembles empathy. Sympathy makes us relate to another by 'conceiving what we ourselves would feel in... (their) situation'.²⁸ What Charles Griswold calls 'sympathetic imagination' refers to our capacity to imagine ourselves in the shoes of another. Our degree of sympathy towards another person is determined by the degree to which we can identify with their situation and react similarly.

'Sympathy, therefore, does not arise so much from the view of the passion, as from that of the situation which excites it' (Smith, 2007[1776], *WN*, I.I.i.10).

This argument challenges methodological individualism because it intertwines the individual well-being, and hence preferences, to that of others. Emma Rothschild considered Smith's self-interest a 'warm sentiment' due to the capacity to sympathise with others.²⁹ The idea is that our senses cannot 'carry us beyond our own persons' but envisioning what we would feel in another person's shoes would make us 'become in some measure (weaker in degree) the same person'.³⁰

Furthermore, Smith proposed that the capacity to imagine oneself in another's situation drives the individual to approve or disapprove of another's behaviour, since it provides a criterion against which to judge the propriety of their response to a given situation. It thus follows that the degree of approval depends on the extent to which we can relate with another's situation and agree with their behavioural response. This mechanism does not contradict the view that we primarily care about our own situation and well-being. Instead, this mechanism is relevant for social collaboration because the feelings and judgements of supposedly isolated individuals can be connected and aligned.³¹

The sympathy mechanism also implies that we would approve when we can relate – imagine ourselves acting similarly – and disapprove when we cannot. The giving of social approbation is therefore driven by a mutual understanding of what constitutes socially-acceptable conduct, otherwise known as social convention. Crucially, sympathy is not a result of categorical moral truths but relies on conventional agreement over what should be approved of.

²⁷ Force, 2003, p.261.

²⁸ Smith, *WN*, I.I.i.1.

²⁹ Rothschild, 2001, p.27.

³⁰ Smith, *TMS*, I.I.i.3.

³¹ *Ibid.*, I.I.iii.1.

'To us, surely, that action must appear to deserve reward, which everybody who knows of it would wish to reward, and therefore delights to see rewarded: and that action must as surely appear to deserve punishment, which everybody who hears of it is angry with, and upon that account rejoices to see punished' (Smith, *TMS*, II.I.ii.3).

Smith's crucial insight is that an individual's criteria for approving is determined by the expectation of what others would approve of; in other words, the consensus. This constitutes an incentive for pro-social behaviour because conforming to social conventions is rewarded by others with social approbation, making its pursuit individually rational. From the perspective of a social consensus, it may follow that pro-social behaviour (others acting in our interest) should always be approved of (while anti-social behaviour frowned upon) by agents who seek to maximise their own individual utility. In the context of economic externalities, social approbation hence provides a mechanism to attenuate negative ones and encourage positive ones. This mechanism does, however, require the assumption that humans have a natural inclination to sympathise with each other. Consequently, they approve of what they deem deserving of applause, which is a function of what they think others would applaud.

Mandeville developed similar ideas about human behaviour despite reaching different conclusions. His pride-mechanism aimed to explain the apparent morality and sociability amongst humans in terms of self-interest by making use of a straight-forward version of the 'countervailing passions principle'.³² Mandeville reduced behaviour to the individual maximisation of pleasurable passions without consideration for others,³³ combining methodological individualism and hedonism. Mandeville then proposed that humans take joy in being applauded, considering 'pride' intrinsically valuable. He argued that social approbation compensates selfish agents for restraining passions which are deemed socially improper.

'Flattery must be the most powerful Argument that could be used to Human Creatures... with unbounded Praises ... bestow'd a thousand Encomiums on the (selfish) Rationality of our Souls, by the Help of which we were capable of performing the most noble (other-regarding) Achievements' (Mandeville, 1988[1732], i.30).

Because humans 'are solicitous about gratifying... (their) appetites',³⁴ the desire for social approbation trumps the passions which the agent knows are not considered praiseworthy by others.

'virtue, who from Politicks Had learn'd a Thousand Cunning Tricks, Was, by their happy influence, Made Friends with Vice' (Mandeville, 1988[1732], i.7).

This is a specific scenario of the pride-mechanism in which social approbation can explain social preferences. In practice, the actions driven by social approbation could go different ways depending on the social conventions. Heterogeneity of behaviour, which is driven by various interpretations of social norms, is hence accounted for.

Whilst the language of Mandeville is heavy with idiosyncratic views about vice and virtue, the claim that individuals seek to maximise social approbation can be made

³² Force, 2003, p.152.

³³ Mandeville, 1988[1732], i.42.

³⁴ *Ibid.*, i.30.

independently of categorical ethical propositions. In fact, social and moral preferences, which he called *virtue*, are not universal truths in Mandeville's account of society, but rather the result of a desire for social approbation. According to Mandeville, moral rules are variant over time and space because they are arbitrary tools invented by rulers and moralists to render people 'useful to one another'.³⁵ He argued that by defining virtue as 'regard for the publick', and making virtue the object of social praise, wise philosophers and skilful politicians were facilitating social collaboration.

The only ethically-neutral inference that can be made from the social approbation mechanism is that people will inevitably approve of behaviour that also advances their own private interests. It could be argued that social norms are the manifestation of common interests over time and thus, the desire to conform to social rules in order to receive the approbation of others provides an individual incentive for pro-social behaviour. Note, however, that this conclusion does not suggest humans are endowed with intrinsic morality. Instead, it proposes that humans can learn over time to build social-reward mechanisms that promote collectively beneficial activity. Through the individual desire for social approbation, conforming to social conventions becomes individualistically rational.

The mechanism hence provides a rationalisation of collective action which might play a role in the challenge of efficient allocation. In a discussion about fruit distribution amongst various guests, Mandeville exemplifies how social approbation can result in a voluntary, and therefore individually optimal, allocation that reveals social preferences.

'If there are Seven or Eight Apples or Peaches among Six People... that are pretty near equal, he who is prevail'd upon to choose first, will take that, which, if there be any considerable difference, a Child would know to be the worst' (Mandeville, 1988[1732], i.72).

By doing so, the agent chooses to accept a material loss (opportunity cost) in the interest of others, revealing social preferences. The reasoning is that in the agent might reach higher utility by maximising social approbation instead of material gain.

'By this Civility the Best remains for others, which being a Compliment to all that are present, every Body is pleas'd with it: The more they love themselves, the more they are forc'd to approve of his Behaviour' (Mandeville, 1988[1732], i.72).³⁶

By bringing 'self-love' into the equation, Mandeville is reiterating how this pro-social behaviour is self-interested in nature. The simple fruit case outlines how approbation, prompted by the propensity to sympathise, constitutes an extrinsic reward for the selfish agent. The First Axiom of Revealed Preferences specifies that, if agents act in this manner voluntarily, the post-distribution scenario would be considered superior to the pre-distribution one, revealing an individual preference for social approbation. If empirical testing were to show this, one could argue that the social approbation mechanism provides a self-interested incentive for social preferences.

Mandeville recognised the challenge of validating his cognitive hypothesis. Through simple observation he found a puzzle that could potentially support his behavioural mechanism. By inquiring into the self-interested reasons for entering military action, he

³⁵ *Ibid.*, i.34.

³⁶ *Ibid.*

identified public recognition as one of the few convincing explanations without relaxing the axiom of self-interest, 'soldiers, that were forc'd to fight, If they surviv'd, got Honour by't'.³⁷

Public honour might seem an unreasonable justification to participate in warfare for *Homo Economicus*. Yet Mandeville thought that a self-interested agent might value the resulting social approbation more than the increased risk of death.

2.iii Evidence

Using methodologies unavailable in the 18th century, modern behavioural researchers have found experimental evidence for the social approbation motive. For instance, Andreoni and Petrie show that removing anonymity in public good's games significantly increases voluntary individual contributions.³⁸ Beyond the external validity of a public goods experiment, the fact that only varying anonymity within the same sample increases pro-social behaviour provides a controlled account of the potential behavioural effect of overt social approbation. Andreoni and Bernheim further propose that the apparent fairness – a conventionally praiseworthy trait – revealed repeatedly in field and lab experiments, can be driven by an inclination for being 'perceived as fair' by others. In fact, their research suggests that the signalling effect accounts for previously unexplained pro-social conduct more fully than a simple appeal to exogenous social preferences.³⁹ The large public exposure and praise following substantial charitable donations provides an indicative real-world example.

List's investigation into the second-hand trading of baseball cards also shows that, contrary to consumption-maximisation, sales involve fairer prices (higher correlation between quality and price) when the trading parties live in closer geographical locations. List argued that traders with a higher degree of social relatability or higher likelihood of crossing paths again experienced an erosion of material maximisation. List finally concluded that half of the effect originally attributed to social preferences actually arose from reputational considerations, which are not internalised by the price mechanism.⁴⁰

Further research indicates that social approbation maximisation might be based on social norms. For instance, certain criminal groups have a form of social convention in which the act of confessing is highly disapproved of,⁴¹ making non-confession (cooperation) the preferred strategy. This challenges the conventional prisoner's dilemma. Furthermore, a cross-cultural experiment conducted by Bowles identified cultural differences when voluntarily contributing in a public good's game. Introducing the ability to identify and punish low contributors increased individual contributions but also revealed that players were willing to undergo a personal financial cost to punish others. Being a cross-cultural experiment, the researchers were also able to identify that punishment varied depending on social norms. In Western societies, low-contributors were generally punished, revealing a social distaste for low-contributors. Yet in other societies, high contributors were punished as retribution by low contributors who had been previously punished.⁴² While the experiment identified the importance of punishment, social conventions also seem to have a distinctive influence on behavioural patterns. Furthermore, the experiment could indicate that the maximisation of social approbation works as a risk-averse inclination to prevent punishment for behaving antisocially. This characterises compliance with social norms as a pre-emptive instinct to avoid potential punishments from society – being ostracised would be a famous example. The

³⁷ Mandeville, 1988[1732], i.7.

³⁸ Andreoni and Petrie, 2004, pp.1605-1623.

³⁹ Andreoni Bernheim, 2009.

⁴⁰ List, 2004.

⁴¹ Anderson, 1996, p.35.

⁴² Bowles, 2016, pp.139-141.

notion that there might be a need for such a mechanism suggests that *Homo Economicus*' selfish materialistic behaviour does not cover all externalities of an agent's activity.

2.iv The Challenge of Anonymity

Behavioural research indicates that anonymity poses an important challenge to the social approbation motive, as it removes the agent from the public's eye. The motive seems to require identifiability, implying a failure to explain pro-social actions completed in anonymity, such as tax payments or small donations. If applause is not forthcoming, the incentive is absent.

Recognising this issue, Smith argued for a cognitive reward mechanism in which the "pleasure of seeing" the results of pro-social behaviour, also thought of as a warm glow, engenders the self-interested incentive behind the act of unconditional collaboration.⁴³ This complies with a hedonistic model of behaviour; the act is self-interested because there is a 'reward' for the agent and unconditional because it does not depend on reciprocity. Instead, the 'reward' comes from triggering an internal reward mechanism which, as is the argument here, is based on expectations of the approbation of others: 'With the eyes of other people, scrutinize the propriety of our own conduct'.⁴⁴

By internalising this mechanism, the desire for hypothetical social approval makes agents judge their own behaviour in the same way others would, using pre-acquired norms of social conduct as the criterion. Similarly to Smith, Akerlof and Kranton argue that 'people have identity-based payoffs' which they derive 'from their own actions' and 'from others' actions'.⁴⁵ Furthermore, they sketch how some identity characteristics might be chosen individually whilst others are exogenous, ultimately defined by socially-constructed conventions such as gender, class or aesthetics. Bowles summarises their argument when explaining that agents facing a consumer or production choice are not only trying to *acquire* something but also building towards *being* someone, 'both in their own eyes and in the eyes of others'.⁴⁶

The resulting view is that judgement about oneself is contingent on others, a clear challenge to methodological individualism.

'Every faculty in one man is the measure by which he judges of the like faculty in another. I judge of your sight by my sight, of your ear by my ear, of your reason by my reason' (Smith, *TMS*, I.iii.10).

Smith further argues that we judge our own actions and behaviours the same way we would judge that of others, which he embodies with the idea of an impartial spectator. The impartial spectator, according to István Hont, is the mechanism behind turning 'a balanced normalised appraisal of society... into the psychological armoury of individuals'.⁴⁷ As such, the impartial spectator (an imaginary social judge of our actions) drives the pro-social without overt social approval from our surroundings. In a way, it internalises social norms and makes agents act in a way in which others would approve of. With the impartial spectator, agents reward themselves with social approbation when their choices conform to social conventions.

While the social approbation motive can, in theory, be made fully intrinsic through the impartial spectator, it is difficult to believe that social preferences could parallel the intensity of

⁴³ Smith, *TMS*, I.i.1.

⁴⁴ *Ibid.*, III.i.5.

⁴⁵ Akerlof and Kranton, 2000, p.717.

⁴⁶ Bowles, 2016, p. 192.

⁴⁷ Hont, 2015), p. 37.

the regard for oneself. Mandeville's and Smith's ideas show instead how the desire to maximise social approbation constitutes a mechanism that can drive pro-social behaviour amongst self-interested agents. Nonetheless, the social approbation incentive is not all-inclusive and merely complements material self-interest.

3. Proposing a Simple Utility Function

Since neoclassical economic agency has been sanctioned through its integration into mathematical language, this section formalises the notion of social approbation maximisation derived from the 18th-century classical economists. The aim is to incorporate their insights about social approbation as a framework for understanding heterogeneous behaviour into a basic utility model of economic agency.

3.i Existing Approaches

The economic literature that acknowledges the existence of social preferences generally invokes two types of utility functions. First, utility models in which agents care for the utility of others, especially amongst families and small groups, as shown by the utility function:⁴⁸

$$u_i(c_i, u_j(c_j, u_i))$$

Agent i cares about her own consumption and also the utility that j derives from consumption, c . This model only acknowledges consumption gains yet includes other-regarding preferences, without providing an endogenous cognitive mechanism for them.

An alternative utility theory to explain social preferences invokes models of inequality aversion, as formalised by Fehr and Schmidt (1999) or Bolton and Ockenfels (2002). In this formalisation, agents prefer outcomes that minimise inequality, implicitly caring for the material well-being of others.

$$u_i(c) = c_i - \gamma_i \max(c_j - c_i, 0) - \beta_i \max(c_i - c_j, 0)$$

Where γ_i captures the agent's disutility of having less than j and β_i the disutility of having more than j . In the case of $c_i > c_j$, agent i cares for the consumption of j .

While both of these approaches provide a framework for understanding social preferences by interconnecting individual utilities, they still adhere to the neoclassical assumption that consumption is the sole source of utility.

3.ii A Simple Utility Function with Overt Social Approbation

The approach taken here is different, as it includes an alternative source of utility: social approbation. Social approbation is an individual desire consistent with the maximisation of abstract utility units. The aim is to provide a model for understanding social preferences in social behaviour beyond the family, where simply evoking other-regarding preferences is harder to justify. By including social approbation as a utility variable, the strictness of methodological individualism and material self-interest is relaxed; the model acknowledges that humans act within a social context and care for others.

A Cobb-Douglas utility function is chosen to represent the relationship between material self-interest and social approbation, in order to represent diminishing marginal rates

⁴⁸ Bergstrom, 1988.

of substitution with convexity. Amongst low-consumption agents, extra consumption increases utility sharply while social approbation might be less relevant. However, social approbation becomes more valuable relative to consumption beyond the point of material satiation, where extra consumption provides little intrinsic utility.

The proposed utility function is therefore:

$$u_i(q) = \alpha \log S_i(q) + (1 - \alpha) \log A_i(q) \quad (1)$$

Where utility is a function of the latently separable material self-interest, S , and social approbation, A , of an individual, I , and which depend on continuous actions, q . Material self-interest encompasses commercial activity in which acquisition of wealth, consumption or investment opportunities are the aim. These should account for the utility-enhancing variables traditionally studied in economics. The effect of social approbation, on the other hand, depends on changes in utility that come from A . Social approbation represents the amount of approval an agent receives from society as a result of his activity – overt approbation. The agent chooses a feasible combination of choice variables, q , that bring him onto the highest indifference curve.

The potential of this model to explain social preferences is outlined in the following scenario, in which the positive material utility derived from a choice is offset by the disutility of negative social approbation. This case represents and formalises Smith's and Mandeville's principle of counterbalancing passions, as identified by Pierre Force.

Imagine, for instance, a consumer decides on the level of loud music to play, q , which also affects others through negative externalities. If only driven by material self-interest, the agent would proceed choosing a high q .

$$\frac{\partial S}{\partial q} > 0 \quad (1.1)$$

$$\frac{\partial A}{\partial q} < 0 \quad (1.2)$$

$$\left| \frac{\partial A}{\partial q} \right| > \left| \frac{\partial S}{\partial q} \right| \quad (1.3)$$

According to condition (1.1), playing loud music would increase her utility. However, if society overtly disapproves of her activity, she would derive negative utility from proceeding, condition (1.2). As a result, the agent faces a trade-off between material self-interest and social disapprobation. Furthermore, if the absolute magnitude of the social approbation effect is larger, condition (1.3), the agent would derive more utility by conforming with social norms, thus voluntarily choosing a lower level of q .

Turning the direction of the effects (1.1) and (1.2) around outlines a different scenario in which an agent might choose to assume a material loss if the extra utility derived from increased social approbation is larger in absolute terms, condition (1.3). This scenario could explain altruistic punishment – when agents punish low contributors at a personal loss – observed in some public goods games. Thus, even when clearly distinguishing between material self-interest and the desire for social approbation, condition (1.3) enables comparability between the two driving forces. Reversing the inequality in (1.3) would indicate that the agent puts more weight on material self-interest.

3.iii Modelling Social Approbation

For a more detailed analysis, it is necessary to consider the supply side of social approbation. Following the ideas of Smith and Mandeville, social approbation depends on overt and covert approbation. The previous case provides a simple illustration of the mechanism, ignoring some of the difficulties surrounding social approbation maximisation. Equation (1) outlines a scenario that does not account for the fact that overt social approbation is limited to the people that can ensure their approval reaches the agent; family, friends, public figures. Social approbation would be purely a function of the number of people overtly approving and the intensity of their approbation, which might be larger than expected due to social media. Nevertheless, as with the challenge of anonymity, overt approbation cannot account for actions that reveal social preferences in which the agent's identity is unknown or indistinguishable from others.

Building on expected utility theory, this is where the notion of covert approbation comes into play. Social approbation is not only a function of post-action overt approbation but also depends on the agent's own expectation of the approbation she would receive if everyone knew her choice, but which is not overtly expressed and might have a probabilistic distribution. The agent is taken to maximise the expected social approbation of some function defined by social norms. To represent this, the following equation could be thought of as the production function of social approbation; it is inspired by the identity-based approach pioneered by Akerof and Kranton (2002).

$$A_i = A_i(o_i(q), c_i^e(q)) \quad (2)$$

Social approbation depends on changes in the total level of overt approbation, o_i , and the agent's expectation of changes in total social approbation, c_i^e . The covert self-approval based on social expectations, c_i^e , characterises Smith's impartial spectator. It reveals the drive to comply with the social norms acquired by the agent, learned through experience in society, since the agent expects approbation would be bestowed on that basis. These might be ambiguous, as both vectors are ultimately dependent on possibly different interpretations of the extent to which the agent's choice conforms to social norms.

Furthermore, overt approbation can be broken down into:

$$c_i = c_i(n, \bar{k}_i(q)) \quad (3)$$

Where approbation by others (3) depends on the number of agents close enough to express their approval, n , and its intensity, defined by the average level of relatability or agreement with the agent's choice, $\bar{k}_i(q)$. One could think of this as a modified version of the gravity trade model, where 'mass' is degree of relatability. Overt approbation of others is therefore dependent, as Smith's sympathy mechanism outlines, on the degree to which others relate and empathise with the agent's situation and choice.

$$w_i^e = w_i^e(k_i^e(q)) \quad (4)$$

On the other hand, covert approbation in (4) varies with the agent's expectation of the degree to which others would agree and relate to her choice, $k_i^e(q)$, which Smith called the impartial spectator. This can be instrumented through the degree to which an agent's choice conforms to the social norms of the agent.

4. Conclusions

The aim of this paper has been to pinpoint the social approbation mechanism as a potential way of accounting for social preferences. Reassessments of the most fundamental behavioural axioms embedded in economic thought have gained traction in past decades, driven by recent behavioural research. This constitutes a response to the neoclassical tradition which has long championed consumption-based utility maximisation. This paper has sought to revive and modernise a more comprehensive utility model of human behaviour derived from the thought of two classical economists – Mandeville and Smith. The capacity to have pro-social preferences has revealed that material self-interest is a limited model to explain individualistic behaviour, partly because it ignores social dynamics. The central idea obtained from these thinkers is that a behavioural model should account for the desire to *be* someone rather than just to *acquire* something. For this purpose, agents seek the approval of others.

Due to this craving to be approved of by others, social preferences are made individualistically rational within a utility-maximisation framework. For this driver to become an incentive for collective action, it must be assumed that time-variant social norms reward behaviour which furthers common interests. Contrary to how the private desire for scarce consumption bundles leads to a confrontation of interests, the individual desire for social approbation can prompt the alignment of private interests. The direction of its effect ultimately depends on social norms, which could incite conflict as well as cohesion. Further research into the role of social norms could identify whether it provides an incentive for collective action, undermining the free-rider problem.

The formalisation of the ideas discussed in this paper provides a model of economic behaviour that permits agents to have social preferences whilst maintaining material self-interest. It loosens methodological individualism by making an agent's utility function dependent on the interests of others, taken to be represented by social norms. Self-interested agents find, in their desire for social approbation, a motivation for their social preferences.

Acknowledgements

With thanks to Avner Offer and John King for their comments via the Economic Thought Open Peer Discussion forum.

Bibliography

Akerof, G., & Kranton, R. (2000) 'Economics and Identity.' *Quarterly Journal of Economics*, Vol.115, No.3.

Anderson, A. (1996) 'Organised Crime, Mafia and Governments.' In *The Economics of Organised Crime*. Cambridge University Press.

Andreoni J., & Miller J. (2002) 'Giving According to Garp: An Experimental Test of the Consistency of Preferences for Altruism.' *Econometrica*, Vol.70, No.2.

Andreoni, J., & Petrie, R. (2004) 'Public Goods Experiments Without Confidentiality.' *Journal of Public Economics*, Vol.88, No.7-8, pp. 1605-1623.

Andreoni, J., & Bernheim B. (2009) 'Social Image and the 50-50 Norm.' *Econometrica*, Vol.77 No.5.

Arrow, K., and Debreu, G. (1945) 'Existence of Equilibrium for a Competitive Economy.' *Econometrica*, Vol. 22.

- Black, A. (2009) *A World History of Ancient Political Thought*. Oxford University Press.
- Bellinger, K. (2007) *The Economic Analysis of Public Policy*. Oxford & New York: Routledge, pp. 50-63.
- Bergstrom, T. (1988) 'System of Benevolent Utility Interdependence.' University of Michigan working paper.
- Bhaskar, R. (1998) *The Possibility of Naturalism*. London: Routledge.
- Bodenhon, D. (1956) 'The Problem of Economic Assumptions in Mathematical Economics.' *Journal of Political Economy*. Vol.64. No.1.
- Bowles S. (2016) *The Moral Economy: Why Good Incentives Are No Substitute for Moral Citizens*. Yale University Press.
- Boadway, R., & Bruce, N. (1984) *Welfare Economics*. Oxford: Basil Blackwell.
- Carlin, W. & Soskice, D. (2015) *Macroeconomics*. Oxford University Press, pp.136-137.
- Chang, H.J. (2015) 'Too Much Maths, Too Little History' in a lecture by the LSE Economic History Department. <https://www.youtube.com/watch?v=6rXBBqMmIP8>. Accessed 09/16
- Cramp, T. (1991) 'Pleasures, prices and principles.' In Meeks, G. (ed.), *Thoughtful Economic Man*. Cambridge University Press.
- Colandar D., H. Föllmer, A. Haas, M. Goldberg, K. Juselius, A. Kirman, T. Lux, and B. Sloth (2009) 'The Financial Crisis and the Systemic Failure of Academic Economics.' Kiel Working Paper.
- Davies, E., & Fefchamps, M. (2017) 'Pledging, Parising and Shaming: Experimental Labour Markets in Ghana.' IZA Institute of Labour Economics.
- Edgeworth, F. (1881) *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. London, pp.1-40.
- Elster, J. (1989) 'Social Norms and Economic Theory.' *Journal of Economic Perspectives*, Vol.3.
- Fehr, E. et al, (2005) *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. London: The MIT Press, pp. 1-35.
- Fehr, E., & Fisher, U. (2002) 'Why Social Preferences Matter.' *The Economic Journal*, Vol.112.
- Fehr, E., & Schmidt, K. (1999) 'A Theory of Fairness, Competition and Cooperation.' *Quarterly Journal of Economics*, Vol.114, No.3, p. 818.
- Friedman D. & Sákovics J. (2011) 'The Marginal Utility of Money.' Sire Discussion Paper.
- Friedman, M. (1953) *Essays in Positive Economics*. Chicago: University of Chicago Press.
- Force P. (2003) *Self-Interest Before Adam Smith*. Cambridge University Press.
- Gravelle, H., Rees R. (2004) *Microeconomics*, 3rd edition. Prentice Hall, ch.1.
- Harfor, T., (2012) 'Black-Scholes: the maths formula linked to the financial crash' in *BBC Magazine* (28-04-12). Accessed online 23/11/20, <https://www.bbc.co.uk/news/magazine-17866646>.
- Hayek, F. (1952) *The Counter-Revolution of Science*. Liberty Fund Inc., pp.51-53
- Hume, D. (1738) *A Treatise of Human Nature, Part 3*.
- Hundert, E. (1994) *The Enlightenment's Fable*. Cambridge: Cambridge University Press.
- Heilbroner, R. (1953) *The Worldly Philosophers*. New York: Simon & Schuster.

- Henrich et al. (2005) 'Economic Man in Cross-Cultural Perspective' *Behavioral and Brain Sciences* Vol. 28, pp. 795-885.
- Hont, I. (2015) *Politics in Commercial Society*. Harvard University Press.
- Kahneman, D. & Deaton, A. (2010) 'High income improves evaluation of life but not emotional well-being' in *PNAS* Vol.107 No.38.
- Kahneman, D., & Tversky, A. (1979) 'Prospect Theory: An Analysis of Decisions under Risk.' *Econometrica* Vol.47 No.2.
- Keller, G. (2015) 'Information Economics.' *Oxford Faculty of Economics Lectures*.
- Koford, K., & Miller, J. (1991) *Social Norms and Economic Institutions*. Ann Arbor, Michigan: University of Michigan Press, pp.39-59.
- Krugman P. (2009) 'How did Economics Get it so Wrong.' *The New York Times Magazine* (02/09).
- Kuhn, T. S. (1973) 'Objectivity, Value Judgement, and Theory Choice.' *The Essential Tension*, pp.357-359
- Laffont, J. and Marmimort, D. (2001) *The Theory of Incentives I: The Principal-Agent Model* Princeton University Press, pp. 1-30.
- Lazear, E.P. (2000) 'Economic Imperialism.' *The Quarterly Journal of Economics*, Vol.115, No.1.
- List, J. (2004) 'The Behavioralist Meets the Market.' *Journal of Political Economy*, Vol.114 No.1.
- Skidelsky, R. (2015) 'Too Much Maths, Too Little History.' Lecture by the LSE Economic History Department, <https://www.youtube.com/watch?v=6rXBBqMmIP88>, Accessed 09/16.
- Mandeville, B. (1988[1732]) *The Fable of the Bees or Private Vices, Publick Benefits, 2 vols.* With a Commentary Critical, Historical, and Explanatory by F.B. Kaye. Indianapolis: Liberty Fund.
- Manslow, A.H. (1943) 'A Theory of Human Motivation.' *Psychological Review*, Vol.50.
- Manokha I. (2015) 'The Conundrum of Economics: Uncompromising Empiricism Alongside Blind Faith in the "Magic" of the "Invisible Hand".' *Raisons Politiques*, Vol, pp.1-12.
- Morgan, M. (2012) *The World in the Model*. Cambridge University Press, p. 8.
- McCloskey, D. (1994) *Knowledge and Persuasion in Economics*. Cambridge University Press, pp. 42-60.
- Mehta, P. (2006) 'Self-Interest and other Interest.' In Haakonssen, K. (ed.), *The Cambridge Companion to Adam Smith*. Cambridge University Press.
- Offer, A., & Söderberg, G. (2016) *The Noble Factor*. Princeton University Press.
- Offer, A. (1997) 'Between the Gift and the Market: The Economy of Regard.' *Economic History Review*, Vol. 50, No. 3, pp.450-476.
- Olson, M. (1965) *The Logic of Collective Action*. Harvard University Press.
- Ostrom, E. (1990) *Governing the Commons: The evolution of institutions for Collective Action* Cambridge University Press.
- Paganelli, M., 'Approbation and the Desire to Better One's Condition in Adam Smith' *Journal of History of Economic Thought*, pp.80-91.
- Robbins, L. (2008) 'The Wealth of Nations: Analytical II.' *LSE Lecture 14 in A History of Economic Thought* (2000), p.125-132.

- Robinson, J. (1962) *Economic Philosophy*. Chicago: Aldine Transaction.
- Rothbard, M. (2010) *An Austrian Perspective on the History of Economic Thought*. Alabama, pp. 261-263.
- Rothschild, E. (2001) *Economic Sentiments*. Harvard University Press.
- Rousseau, J. (1754) *Discourse on the Origins of Inequality*. Holland: Marc-Michel Rey.
- Sassower, R. (2010) 'Is the Homo Economicus Extinct?' *Philosophy of the Social Sciences*, 40 (4), pp. 603-615.
- Schneider, L. (1987) *Paradox and Society: The Work of Bernard Mandeville*. Oxford University Press, pp. 50-135.
- Sen, A. (1977) 'Rational Fools' A Critique of the Behavioral Foundations of Economic Theory.' *Philosophy & Public Affairs*, Vol.6 No.4.
- Shumpeter, J. (1954) *History of Economic Analysis*. New York: Oxford University Press.
- Smith, A. (2007[1776]) *An Inquiry Into the Nature and Causes of The Wealth of Nations*, edited by S.M. Soares, MetaLibri Digital Library.
- Smith, A. (1790) *The Theory of Moral Sentiments*, 6th Edition.
- Stevenson, B., and Wolfers, J. (2008) 'Economic Growth and Subjective Well-Being.' *Brookings Papers on Economic Activity*.
- Synder, C. and Nicholson W. (1980) *Microeconomic theory: basic principles and extensions*. Cengage Learning, pp.441-442
- Tolonen, M. (2009) *Self-love and Self-liking* Helsinki University Press.
- Von Neumann, J. (1944) *Theory of Games and Economic Behaviour*. Princeton University Press.
- Weber, M. (1922) *Economy and Society*. University of California Press.
- Weintraub, R.E. (2002) *How Economics Became a Mathematical Science*. Duke University Press, p. 36-37.
- Youyou W., Kosinski M., Stillwell D. (2015) 'Computer-Based personality judgements are more accurate than those made by humans.' *Proceedings of the National Academy of Sciences*, Jan, 112 (4) pp. 1036-1040.

SUGGESTED CITATION:

Ruiz, Oswin Krüger (2020) 'The Self According to Others: Explaining Social Preferences with Social Approbation.' *Economic Thought*, 9.2, pp. 38-54. <http://www.worldeconomicsassociation.org/files/journals/economicthought/WEA-ET-9-2-Ruiz.pdf>