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

Oswin Krüger Ruiz
University of Oxford

Abstract
In past decades, significant behavioural work has been done in economics contesting the human agency model known as *Homo Economicus*. These findings are, however, far from integrated in mainstream economic theory, which builds heavily on the neoclassical tradition. Beyond tracing the axiomatic foundations of mainstream economic agency, I reconstruct the argument made by Bernard Mandeville and Adam Smith that human choice also depends on the desire for social approbation. The social approbation mechanism complements material self-interest and provides a more diverse toolset when explaining social preferences. I begin by proposing that mainstream economic agency confines the study of human action, in particular large-scale collaboration, 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 with consumption. Collective action is therefore only sustainable when material incentives are in place, as the economic agent rides for free unless financially penalized. To explain pro-social behaviour from the standpoint of self-interest, Mandeville and Smith propositioned 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 goals, 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.

1 An economic model of human behaviour.
2 Preferences to act pro-socially, in the sense of Ernst Fehr and others behavioural economists.
I. Neoclassical Economic Agency

I.i The Origins

Challenging the conception that Adam Smith is the father of economics, the inquiry into the beneficial role of self-interest in a commercial society has forerunners in even earlier thinkers. According to Friedrich Hayek, it was Bernard Mandeville who, in the 18th century, first identified the paradoxical idea that the free pursuit of selfish desires could lead to unintended public benefits, pinpointing “the object for theory in the social and economic realms”.³ By arguing that “every individual is a little world by itself” that exclusively seeks its own pleasure, Mandeville endorsed the two axioms of methodological individualism and hedonism.⁴

Individuals are “only solicitous of pleasing themselves… without considering the good or harm that from their being pleased will accrue to others”.⁵

Mandeville had a cynical view of people, deriving inspiration from Pierre Bayle and François de La Rochefoucauld, both of whom fundamentally judged humans as “fallen” and “vastly more impelled to evil than to good”.⁶ Rooted in this tradition, which later influenced Smith, is the Epicurean proposition that human beings are isolated entities that maximise their own interest, utility, or pleasure. The historical origin of this assumption is not empirical, as it builds on metaphysical opinions about human nature.

⁵ Ibid., i.28.
⁶ Schneider, Paradox and Society, pp.68-70.
It includes the corollary that individuals lack inclinations to care about the well-being of others.

In 1776, Adam Smith embarked on the challenge of responding to Mandeville’s cynicism about human nature, departing from similar propositions. The supposedly selfish individual depicted in *The Wealth of Nations (WN)*, which ignored the more comprehensive account of human behaviour in *The Theory of Moral Sentiments (TMS)*, is even today thought of as the subject-matter of economic analysis.\(^7\) The justification for this view comes from his discussions about the market mechanism in the *WN*. Adam Smith recognised that, by construction, the activity of voluntary trade necessarily involves mutually-beneficial outcomes and that at no instance does it require an agent’s active regard towards the well-being of the other party, instead building on the “propensity to truck, barter, and exchange one thing for another”.\(^8\) Despite his depiction of the invisible hand as somewhat “ironical”;\(^9\) Adam Smith provided one of the earliest, if somewhat simplistic, accounts of a system in which the choices of self-interested individuals have unintended consequences that promote social interests. He constructed the mutually-beneficial trade relationship between an agent and the “butcher, the brewer, or the baker”, not based on intrinsic “benevolence” or sympathy, but on our propensity to persuade others to help us by appealing to their self-interest. The prime engine of Smith’s commercial society seems to be our material self-interest and that of others.

Agent tries to “show them (other agents) that it is for their own advantage to do… what he requires of them”.\(^10\)

---

\(^10\) Smith, *WN*, i.ii.2.
With our “faculties of reason and speech” we endeavour to persuade others to collaborate through exchange, using their self-interest as the reason-to-trade. When we propose to trade by the act of offering a “shilling”, we provide an “argument” to align another agent’s interests with our own, generating a form of social collaboration. This interpretation of the market mechanism portrays exchange as communication and money as a persuasion device. This reading is supported by Pierre Force, who also cautions that Smith’s material self-interest is only one aspect of human behaviour rather than the sole motor of individual and social choice. Smith’s oft-cited account of the market system in Book I is rudimentary when contrasted to the totality of his writings because it presents a historical argument about how trade might have originated rather than a comprehensive account of the behavioural components of markets in a modern commercial society, which his other texts reveal to be much more elaborate. A simplistic reading of Smith might mistakenly suggest an endorsement of material self-interest and a disregard for shared authority or social interaction. The proposition in Book I, underpinning neoclassical economics, merely outlines the hypothesis that agents cannot be expected to trade on the basis of social preferences alone but must rely on the material self-interest of other individuals.

I.ii The Rise of the Homo Economicus

Building on Smith’s Book I, neoclassical economists generalised the selfish motive as the driver of human behaviour and sought to compute unintentional cooperation, driven by an invisible hand, as “the result of the interaction of selfish agents maximising their

\textsuperscript{11} \textit{Ibid.}
long-term individual material interests”.

Determining “to what extent egoistic behaviour (would) achieve general good” is, according to Amartya Sen, the “relatively abstract dispute” which attracted the attention of prominent economic thinkers like Jevons, Marshal, Edgeworth and Pareto.

When talking about utility-maximisation in economics, the most appropriate utilitarian understanding is Benthamite hedonism, which sees the maximisation of pleasure and minimisation of pain as the proper aim of human life. Mill’s utilitarianism, in contrast, diverges from economic maxims by allowing utility to be derived from ethical goals. To eliminate complications, neoclassical economists adhered to the analytical technique of considering for a fact that humans are calculating machines which could be studied through the “calculus of pleasure and pain”. By embracing this agency model, Francis Edgeworth made a substantial contribution to validating economic agency in formal analysis, its progress driven by the emulation of Newtonian physics. His *Mathematical Psychics (1881)* is an important cornerstone of textbook microeconomics. Therein, he advocated the quantitative study of the human mind, which he deemed feasible by calculating discernible changes in the intensity of *pleasure-units* over time. In other words, individual utility, or pleasure, is considered the standardised measure of the objective of human action. The easiest step thereafter involved assuming that humans would maximise this single form of abstract pleasure-accounting for themselves. The paradox of unintended consequences simply meant that

---

15 Ibid.
agents would exchange until their marginal rate of substitution equals the price ratio, the famous condition for utility-maximisation.

\[
\frac{MU_1}{MU_2} = \frac{P_1}{P_2}
\]

Edgeworth thus proposed that economic activity tends to the “equilibrium of hedonic forces each tending to maximum individual utility”, creating parallels to physical phenomena.\(^\text{19}\) Justifying his approach meant entering the realm of philosophy, in which he regarded the quest for a “principle of arbitration” as the challenge of allocating value and resources amongst individuals. The ambiguity and unquantifiability of common moral principles implied these should not be studied; categorical moral truths were too imprecise for economics. Instead, he sanctioned methodological individualism by arguing that selfish utility-maximisation drives human agents, who interact towards exhausting all gains from trade. Despite building his utilitarian validation primarily on Sidgwick’s consequentialism, justification for this system is derived by appealing to contracts. These make self-interest socially permissible because they grant an individual “the consent of others affected by his actions”,\(^\text{20}\) as long as actions are accounted for – Edgeworth overlooked the difficulties that arise from externalities. As a result, an agency model based on individualistic egoism received its moral sanctioning on the basis of contractualism as well as consequentialism.

Simplifying the subject-matter of economics became a crucial step in the development of analytical tools based on formal logic and mathematics. Edgeworth and Pareto saw the mathematical turn in their field as the representation of “basic

\(^{19}\) Ibid., p.15.
\(^{20}\) Ibid., pp.14-17.
propositions in mathematical logic” and computation of the “implication of those derivations” in ways that eliminated ambiguity.\textsuperscript{21} Finding a formal solution to Mandeville’s paradox spurred on the quest of proving the existence of a competitive equilibrium, considered the 20\textsuperscript{th}-century holy grail of economics. In 1954, Debreu and Arrow defined modern microeconomics by showing that “under suitable assumptions…the allocation of resources in a competitive equilibrium is optimal in the sense of Pareto”.\textsuperscript{22} Pareto efficiency ensures that a state is reached in which no further gains are possible without producing a loss for somebody. The resulting First Fundamental Welfare Theorem is still considered one of the foundations of textbook economic teaching.

\textit{I.iii In Modern Economics}

Despite facing increasing challenges from behavioural economists, the self-interest axiom still underpins much economic thinking through textbook economics. It is considered the benchmark from which deviations happen, making it more of a paradigm than a theory.\textsuperscript{23} Modern microeconomics deals with agents – consumers and firms – who own an initial ‘endowment’ and embark on economic exchange by ranking all the options in order to choose the most preferred.\textsuperscript{24} Firms try to maximise profits for the purpose of maximising the individual consumption opportunities of its stakeholders. All economic activity is therefore reduced to individual economic agents who try to optimise their utility functions, generally computed through consumption-leisure

\textsuperscript{23} Koford, K., & Miller, J., \textit{Social Norms and Economic Institutions} (Michigan, 1991), p.44.
choices and limited by their budget constraint. The agent decides on the consumption bundles that give maximum possible utility. This is a cyclical concept in which utility is the subjective reward people derive from the chosen bundles. The fact that they are chosen reveals that those bundles are preferred and therefore have maximum utility.

Beyond microeconomic theory, the 2016 Nobel Prize in Economic Sciences further cemented the prevalence of the self-interest axiom; the prize was awarded to Hart and Holmström for their contribution to contract theory. Within this field, the principal-agent model builds on individualistic self-interest to study the minimum expected payoff required for an agent – participation constraint – to complete an action desired by a principal. The model makes material payoff and the inverse of effort the sole explanatory variables of human behaviour, giving rise to moral hazard, a hidden action due to imperfect information and a sequential incentive structure. The agent has an incentive not to complete the task if the payment is made beforehand because the model infers that effort can only be overcome by the material gain which has already been attained. Game theory, pioneered by von Neuman and others, also characterises behaviour as self-interested, despite allowing for scenarios of cooperation in certain games.

The supposed lack of social preferences in neoclassical economic agency means that collective action is eroded by the free-rider problem. The combination of self-interest and methodological individualism is what made Thomas Hobbes conclude in 1651 that, without coercion, social cooperation could not be sustained. In 1965, Mancur Olson published the Logic of Collective Action, which outlined how, based on propositions of economic agency, common interests are weakened by individual ones

---

due to the free-rider problem, especially as the group’s size increases. The analysis is that, in the absence of coercion, “rational self-interested individuals will not act to achieve their common or group interest”.\textsuperscript{28} If an individual can benefit from collective action without having to contribute towards it or face a cost for their inaction, there is no incentive to contribute. From a stance of individual self-interest, collectively rational actions become unsustainable because they are individually irrational, even if the action is in the interest of the collective.

Collective goals are, however, built on the idea that “groups act in their self-interest because individuals do so”.\textsuperscript{29} There are incentives for collaborating, therefore, if the average individual benefit of collective action is preferable. Here there is a difference to be drawn between the preferred consequence and action. The consequences of collective action are individually preferred. Yet an individual might prefer the consequence of a collective action but not prefer the individual action towards it – generally depicted as a contribution. As soon as a critical number of people contribute and an individual can benefit without contributing, there is no incentive to do so. This problem is generally tackled by punishment, coercion, and contracts. The point is that agents require a personal incentive to contribute towards common goals. However, to argue for the inexistence of such an incentive in human nature based on the model of individualistic and material utility-maximisation is a different matter.

By tracing its development, mainstream economic agency seems to be characterised by two main notions. Firstly, individuals are distinctive agents that make decisions with regard to their interest, known as methodological individualism. Secondly, agents maximise utility, which is generally instrumented through

\textsuperscript{29} \textit{Ibid.}, p.3.
consumption. In the context of collective action in society, this model does not allow for pro-social contributions that are not based on material self-interest.

II. A Methodological Convenience

The generalisation of material and individualistic self-interest that characterises neoclassical economics may be valuable as a simple model of human behaviour. However, being able to theorise about economic activity by instrumenting human agency solely through material self-interest limits the ability of the science to cover the heterogeneity of human behaviour. Economists have generally recognised the limitations of neoclassical economic agency outlined in the previous section. Nevertheless, it is important to outline that it has remained ingrained in textbook economics because it facilitates analysis.

II.i Early Doubts

Even one of the most emphatic proponents of self-interest failed to abstract its benefits from a social and political setting.

“So Vice (self-interest) is beneficial found; When it’s by Justice lopt and bound”\textsuperscript{30}

\textsuperscript{30} Mandeville, \textit{Fable}, i. 36-7.
Mandeville recognised justice as a necessary principle of arbitration in society; his paradox presupposed a basic regulatory framework. It parallels the view of Chang, who argues against the existence of pure markets. Markets build on a set of political and social conventions like the right for private property or possibility to commodify.  

Furthermore, Mandeville acknowledged the limitations of material self-interest when he distinguished between “self-love” and “self-liking”, separating preferences into material needs and more intangible drivers such as pity and the craving for social esteem. Similarly to Maslow’s pyramid in which separate drivers like “basic needs”, “esteem needs” and “self-actualisation” coexist, Mandeville’s agency seems to encompass separate and distinct human passions that cannot be reduced to one numeraire. Mandeville thus embraced the interconnected complexity of human drivers when investigating how individual pleasure machines could prompt social benefits.

Adam Smith regarded the blind workings of the invisible hand as something particularly applicable to the most voracious and selfish agents, who could at times also be fostering the interests of society. Yet by recognising various motives in human nature, he acknowledged that other more tractable mechanisms also upheld social cohesion. Society is not solely the result of commerce; quite the opposite, commerce relies on society. The view in Smith, to which even Hayek conceded, is that the benefits of markets require “both good institutions and good norms”, limiting the protagonist role of neoclassical economic agency.

Adam Smith’s earlier lectures also expound the heterogeneous complexity of human nature, induced by his attachment to Stoicism.

32 Tolonen, M., Self-love and Self-liking, (Helsinki, 2009)
34 Rothschild, Economic Sentiments, p.123.
35 Ibid., p.156.
“the source of human motivation… cannot be easily reduced to a few simple principles”\(^\text{36}\)

A point often disregarded in discussions about utility-maximisation is that passions and rational choice are both covered by the notion of interest. According to Force, the market mechanism of Smith relied on the supposition that actions are driven by “interest calculations” in order to establish how selfish “passions of individuals contribute to the common good”.\(^\text{37}\) Passions and rational choices are therefore melted into one concept – utility – in order to be substitutable. This standardisation of individual reason-to-act was crucial to Edgeworth’s quantification of individual motivation and means that all drivers are exchangeable with one another. Smith, however, recognised an important distinction between intangible drivers – pride and sympathy – and material drivers. For this reason, he saw the market mechanism as insufficient to uphold social cohesion.

\textit{II.ii Evolving Hesitations}

According to Amartya Sen, Edgeworth made a more substantial contribution to justifying the assumption of egoism, especially in war and contracts, than Adam Smith.\(^\text{38}\) Sen also identified the limitations in Edgeworth’s endorsement of egoism.

“between the claims of oneself and the claims of all lie the claims of a variety of groups-for example, families, friends, local communities, peer groups, and


economic and social classes… and the dismissal of (universal) utilitarianism
as a descriptive theory of behaviour does not leave us with egoism as the only
alternative”. 39

By departing from the dichotomy between either caring for the interest of oneself or of
all, Edgeworth was able to argue, within the neo-Epicurean tradition, that humans were
more likely to be selfish. 40 The possibility that people might have mechanisms in play
that allow them to care for others but to a smaller degree than for themselves remained
excluded. Edgeworth himself acknowledged this limitation before simply “deferring
controversy” on the issue. It is not difficult to sense that Edgeworth’s motivation was to
“facilitate” rather than to “justify” the “employment of mechanical terms and
mathematical reasoning in social science”. 41 In light of the limiting assumptions which
the use of mathematics required, even Edgeworth and Pareto were openly critical about
the application of economic theory to real problems. 42 It seems that they regarded their
economics more as a thought-experiment than an applicable science.

II.iii Modern Reservations

The legacy of neoclassical economists promoted the standardised and individualistic
accounting of human preferences on the basis that it facilitates complex analyses.
Nevertheless, this view has attracted criticism from a broad range of social scientists.
Skidelsky links it to a reduction in the diversity of propositions that permit covering the

39 Ibid., p.318.
40 Edgeworth, Mathematical Psychics, pp.3-6.
41 Ibid., 15-16.
42 Weintraub, How Economics Became a Mathematical Science, pp.36-37.
empirical material.\footnote{Lord Robert Skidelsky, ‘Too Much Maths, Too Little History’ in a lecture by the LSE Economic History Department (2015). Accessed 09/16, \url{https://www.youtube.com/watch?v=6rXBBqMmlP88}} Already in the 1950s, literary economists criticised the approach of making mathematically convenient assumptions, which inevitably limited the applicability of the models.\footnote{Bodenhon, D., ‘The Problem of Economic Assumptions in Mathematical Economics’ \textit{Journal of Political Economy} Vol.64 No.1, (1956), pp.25-32.} After economists failed to foresee the 2008 financial crisis, Krugman argued that the profession “mistook beauty, clad in impressive-looking mathematics, for truth”.\footnote{Krugman P., ‘How did Economics Get it so Wrong’ \textit{The New York Times Magazine} (02/09/09)} Colander and his co-authors have gone as far as to argue that a false sense of certainty derived from financial models played a role in the run-up to financial crises.\footnote{Colandar et al. \textit{The Financial Crisis and the Systemic Failure of Academic Economics} (2009)} The Black-Scholes formula provides an example, as it delivers assurances when trading options due to its assumption of large deviations being extremely rare.\footnote{Harfor, T., ‘Black-Scholes: the maths formula liked to the financial crash’ \textit{BBC Magazine} (28-04-12)}

Economic models create an abstract world that might be consistent and render useful insights without corresponding to reality.\footnote{Morgan, M., \textit{The World in the Model} (Cambridge, 2012), p.405.} 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. As a result, Joseph Schumpeter clearly distinguished between application and pure analysis. “From the standpoint of the requirements”, or what is externally expected of economic theory, it seems to have consistently underperformed in its application, even if the analytical apparatus itself is not to blame.\footnote{Shumpeter, J., \textit{History of Economic Analysis} (Oxford, 1954), pp.17-18.}

With regards to agency, von Neumann acknowledged that sophisticated analysis based on an inadequate model of human behaviour would render little understanding.
“There is no point in using exact methods where there is no clarity in the concepts and issues to which they are to be applied.”

Exact theorising in social sciences requires the difficult task of computing human behaviour. Institutional economist and laureate Elinor Ostrom recognised that “theoretical inquiry”, by its nature, “involves abstraction from the complexities of a field setting, followed by the posting of the theoretical variables that underlie observed complexities”. She argued that the power of a theoretical approach is “proportional to the diversity of situations it can explain”. What behavioural economists have shown is that, because mainstream economic agency is constructed on material self-interest, observable deviations – social preferences – cannot be explained by it. Ostrom contended that dealing with complexity is not the same as assuming it away. The trade-off between generality and explanatory power implies that increasing the variety of axioms about human behaviour could increase the economist’s toolbox without compromising the technique of modelling.

The internal discussion in economics might have evolved much since the 1980s, as McCloskey points out, partly due to the realisation that other types of arguments are as central to economics as “axiom-and-theorem”. In his 1953 essay concerning positive economic science, Milton Friedman described economic theory as a “system organising empirical material and facilitating our understanding of it”. The system, or language, aims to develop hypotheses about future phenomena that are regarded to be “valid and meaningful” rather than “truistic”. This resembles Thomas Kuhn’s argument that the conceptual frameworks which scientists accept by consensus is based on

---

52 Ibid.
paradigms defined by “collective judgement” and ultimately grounded in a shared number of criteria about what ought to be accepted.\footnote{Kuhn, T. S., ‘Objectivity, Value Judgement, and Theory Choice’ in The Essential Tension (1973), pp.357-359.} According to Kuhn, scientific criteria include factual accuracy, internal consistency as well as with other theories, broadness of scope, simplicity, and fruitfulness.\footnote{Ibid.} However, Friedman argued that in order for economics to be useful, its analytical toolbox should also be exhaustive.\footnote{Friedman, M., Essays in Positive Economics, (Chicago, 1953), p.4.} It is futile to ask one theory to explain everything, meaning that a variety of propositions will more fully account for variety. Concerning the ways in which human collaborative activity is rationalised from the standpoint of individuals, this essay postulates that the categories of understanding are not exhaustive because they are built on a limited notion of agency.
III. The Limits of Neoclassical Economic Agency

The pervasiveness of economic analysis in other disciplines, called ‘Economic Imperialism’, implies that the validity of its behavioural model should be constantly reassessed. This section attempts to pinpoint the fundamental limitations of mainstream economic agency before suggesting a way to complement the assumption of material self-interest with social approbation maximisation.

III.i Social Preferences Exist

In the last two decades, one of the main accomplishments of behavioural field work and randomised controlled experiments has been to falsify the *Homo Economicus* and as a consequence undermine the generalisation that human behaviour can be studied predominantly by reference to material incentives.\(^{59}\) The economic agent is often willing to defy material self-interest and strategies considered Nash Equilibria. One of the main insights of this movement is that economic agents also hold social preferences.

Beyond material self-interest, humans can be “conditional co-operators and altruistic punishers”.\(^{60}\) For instance, the ultimatum game has shown that a player responsible for proposing the division of money with another anonymous player, who decides whether to accept the division, fails to reveal selfish maximisation. Receiving your share of the partition depends on the acceptance of the second player, who, if materially self-interested, would accept any positive number. Assuming player 2 is self-interested too, the first player could offer the minimum positive amount to ensure

---


\(^{60}\) Fehr, *Moral Sentiments and Material Interests*, xii.
acceptance. Contrary to expectations, first-movers often offer 50% of the money and second-movers regularly reject offers below 30%, listing concerns about fairness or inequality-aversion as explanations.\textsuperscript{61} Based on similar research, Andreoni and Miller captured altruistic behaviour in rational and well-behaved preference ordering.\textsuperscript{62} Data collected from 176 subjects allowed for an ordering in which only 25% were selfish money-maximisers.

While field research conducted by List finds that experimental results do not fully extrapolate to field settings,\textsuperscript{63} indicative of a Hawthorne effect, other observations also support the existence of social preferences. 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 payoff for the individual.\textsuperscript{64} Further research into the existence of social preferences can be found in the work of Ernst Fehr, Samuel Bowles, Dan Ariely, and other behavioural economists.

\textit{III.ii Self-Interest Beyond Consumption}

Despite these findings, mainstream economic theory generally implies inexistence of social preferences. It suggests behaviour is driven by material self-interest when consumption is the object of utility-maximising individuals. Lack of other-regarding preferences arises from the scarcity of resources, which creates interest conflicts amongst consumption-maximising individuals. In most cases, an agent’s consumption of bundle $(c_1, c_2)$ is excludable and limited resources make it rival. However, the

\textsuperscript{61} Ibid., pp.12-15.
existence of social preferences exposed the model to be at odds with certain behavioural traits; rivalry is inherent only when consumption is the only goal.

As economic theory generally assumes, the individual maximisation of subjective well-being can partly be instrumented through wealth and consumption, since many components of well-being can be purchased. Contradicting the Easterlin (1974), who found no association between happiness and income, recent research finds a positive link between wealth and well-being. But the validation of material self-interest is methodological convenience rather than empirical. Unsurprisingly, wealth is revealed to be a good estimator of life-satisfaction, but its ‘effect’ is neither monotonic nor homogenous across agents. Survey data collected by Kahneman and Deaton indicates that income loses its positive effect on emotional well-being past the threshold of $75,000 per annum. It reveals an important distinction between emotional well-being, which relates to the day-to-day emotional state, and objective reflection on overall life-satisfaction. This research alludes to the possibility of diminishing marginal returns of well-being to wealth. The commonly held assumption that monotonicity of preferences applies to consumer choice is thus exposed as inadequate. More of the same is not always better.

Diminishing marginal returns to money, which Figure 1 outlines, has been proposed as an alternative to the budget constraint, one of the reasons being that it allows for the consumer's subjective valuing of wealth at different stages of their acquisition process.\(^67\) The fact that income levels vary across similar states of emotional well-being, once basic needs are covered (M1), thus partially invalidates the approach of using material self-interest as a comprehensive proxy for human motivation.

### III.iii Methodologically Indivisible

From the notions of human agency within social science, Jon Elster identifies a tension between the *Homo Economicus* and the *Homo Sociologicus*. While most economic models permit deviations from the benchmark, such as “imperfect information… (or) transaction costs”, choice is generally not “determined by forces beyond the control of

the individual”, excluding constraints on the choice set. For instance, social welfare rankings are reduced to preferences of individuals, meaning these must be measurable and comparable. The emphasis on individualistic agency in economics implies agents are isolated from each other.

Methodological individualism is, however, challenged by the argument that individuals are the products of their social contexts. This complicates the study of micro-foundations as it brings social nurture as an influencing factor alongside nature. According to Isaiah Berlin, even one of the sturdiest defenders of individuality, J.S. Mill, could not separate the individual from interactions with others; in a deeper sense, the individual is partly defined by others. Roy Bhaskar argues that whatever characteristics we designate people with, they all “presuppose a social context for their employment”, which would make social nurturing something natural to individuals. As a result, the state of nature in which humans are not yet in society, an idea popularised by certain thinkers of the Enlightenment, becomes an artificially-constructed concept that incorrectly justifies the reduction of society to individuals. The view that humans exist within a social setting is corroborated by some of the earliest archaeological evidence recorded, which suggests that people have always lived in some form of association. If the existence of a society is as old as the existence of the individual, we might ask whether they are separable, or whether we can fully account one with the other.

The documentation of social preferences by behavioural economists exposes the limitations of material self-interest. The importance of social context in defining the individual also challenges the universality of micro-foundations. Agents are

interconnected and look beyond material self-interest. This complexity of human behaviour presents a challenge to modelling it, a challenge for which Smith and Mandeville had some answers.
IV. Instrumenting Social Preferences through Social Aprobation

While Mary Morgan considers Smith’s account of human agency too multifaceted for modelling, a first step is to incorporate the social approbation motive. Both Smith and Mandeville departed from the tensions between self-interest and altruism to reach conclusions that looked beyond the dichotomy. Their thought indicates that human behaviour is driven by social approbation maximisation as well as material self-interest. For this, they presumed that people sympathise and hence approve of each other. Society approves of actions that conform with social norms, turning the desire for social approbation into a driver of social preferences.

IV.1 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. Embedded in the foundations of economics is the belief that ethical and social considerations are beyond its scope. This tradition can be traced back to David Hume’s acknowledgment about the futility of passing moral judgement on self-interested yet social agents.

    If people are virtuous they “become social by their virtue; if vicious, their
    vice has the same effect.”73

The message was that social preferences exist and are driven by a genuine regard for others or by indirect selfish motives. The resulting question is, however, whether the

73 Hume, D., A Treatise of Human Nature (1738), iii.ii.ii, 256.
selfish-selfless dichotomy is useful. It can be argued that, in its barest form, self-interest simply indicates the interest of an agent, 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 critic of Mandeville, Joseph Butler, argued that people could place self-interested value on morality, rendering morality and self-interest compatible.\textsuperscript{74} This approach reinforced one of Mandeville’s crucial contributions. It is pointless to categorise human activity by distinguishing between other-regarding and self-regarding acts if both can always be traced back to some form or another of (neutral) self-interest. According to Mandeville, “to judge a man’s performance” we need to know “the principle and motive from which he acts”.\textsuperscript{75} Yet the fact that generosity can be interpreted as a genuine impulse to help others, as with Rousseau’s pity-mechanism, but also be considered “ambition in disguise”, as expressed by La Rochefoucauld,\textsuperscript{76} 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 branded as “deliberative processes”.\textsuperscript{77} On the contrary, 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 causes impulses that relate to collaborative activity, such as “conformity of an


\textsuperscript{75} Mandeville, Fable, I.43.

\textsuperscript{76} Force, Self-Interest Before Adam Smith, p.137.

\textsuperscript{77} Bowles, S., The Moral Economy (Yale, 2016), pp.103-110.
action with a set of rules”, as well as more selfish impulses. The distinction between selfishness and selflessness breaks, as it rests on ambiguous intentionality.

This realisation is one of the cornerstones of outcome-based approaches such as utilitarianism, which economic agency builds on. It reveals the false dichotomy between egoism and collaboration, a result relevant to economics. Even Edgeworth distinguished between economics and ethics on the basis of self-interested activity and selfless behaviour respectively. As a result, through its limited conception of agency, neoclassical economics has been shaped as the science that studies only the self-interested behaviour of humans, traditionally leaving out important aspects of social interactions. Recognising that all choice is relevant to the economic inquiry of efficient allocation suggests that there is a need to look beyond the dichotomy that has justified studying all behaviour with respect to material self-interest.

IV.ii Social Approbation Maximisation

Based on neutral self-interest, I argue that the borders between the self and the rest are not that clear-cut. Humans are partly a distinctive entity and partly a component of a greater social structure. It is therefore necessary to find a way of making the assumption of individual maximisation less self-centred. Since people partially reveal social preferences, there must be a behavioural mechanism to explain it. Smith and Mandeville outlined a possible process by which individuals seek improvements in the social approbation received from others.

78 Ibid.
79 Force, Self-Interest Before Adam Smith, p.171.
In Smith’s complex conception of agency, humans are driven by the “great purpose of human life which we call bettering our condition”.\textsuperscript{80} 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”.\textsuperscript{81} In Kahneman’s prospect theory, which fits remarkably well to 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.\textsuperscript{82} Human action can therefore be defined as changes beyond the current state of an agent, which are assumed to be driven by the agent’s preferences over a set of possible changes. In other words, an improvement.

Smith depicted improvement as an “innate desire to receive the approbation of others”.\textsuperscript{83} His views on social regard bear resemblance to Rousseau’s conception of \textit{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 with the difference between that point and our own position. While Rousseau identified that “all equality vanished” as a result,\textsuperscript{84} Mandeville saw the resulting desire for social esteem as a motive for collaborative activity. Positioned in between, Smith seemed to consider that social approval can be both the object of “congratulations” and “envy”.\textsuperscript{85} As a result, 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-

\textsuperscript{80} Smith, \textit{The Theory of Moral Sentiments} [henceforth TMS], 6th Edition (1790), I.iii.2.1.
\textsuperscript{82} \textit{Ibid.}, p.277.
\textsuperscript{84} Rousseau, J., \textit{Discourse on the Origins of Inequality} (1754)
\textsuperscript{85} Smith, \textit{TMS}, I.ii.v.1.
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.86 Consumption being bounded, he reasoned, changed the nature of wealth acquisition from having intrinsic to instrumental value.

“The rich man glories in his riches, because he feels that they naturally draw upon him the attention of the world”87

Smith thus acknowledged that wealth is accumulated beyond material needs and identified it as a way of drawing the admiration of others. Nevertheless, even if wealth accumulation “constantly obtain(s)” social recognition, it is not the only behaviour conductive of distinction.88 An agency model that incorporated social approbation maximisation would, according to Smith’s propositions, be more comprehensive than instrumenting the quest for betterment solely with 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.

86 Ibid., I.III.ii.1.
87 Smith, TMS, I.III.ii.1.
88 Ibid., I.III.iii.3.
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”.  

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 of 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”.

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 thus be directly inferred from the approval and admiration received from others.

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

“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”.

---

89 Ibid., I.III.ii.1.
90 Force, Self-Interest Before Adam Smith, p.132.
91 Smith, TMS, III.1.3.
As a result, Smith believed that the self is formed within society, raising doubts about whether it 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.

Looking at 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, an idea that closely resembles empathy. Sympathy makes us relate to another by “conceiving what we ourselves would feel in… (their) situation”. 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”.

This argument challenges methodological individualism because it connects individual perception of 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

---

92 Force, Self-Interest Before Adam Smith, p.261.
93 Smith, WN, I.I.1.
94 Ibid., l.l.i.10.
with others.\textsuperscript{95} The idea is that our senses cannot “carry us beyond our own persons” but envisioning what we would feel in another’s shoes would make us “become in some measure (weaker in degree) the same person”.\textsuperscript{96}

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 “proprietary” 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 focus on our own situation. Instead, this mechanism is relevant for social collaboration because the feelings and judgements of supposedly isolated individuals can be connected and aligned.\textsuperscript{97}

The sympathy mechanism also implies that approval is triggered if we can relate – imagine ourselves acting similarly – and disprove 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 these should be.

“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”\textsuperscript{98}

\textsuperscript{95} Rothschild, \textit{Economic Sentiments}, p.27.  
\textsuperscript{96} Smith, \textit{TMS}, I.I.I.3.  
\textsuperscript{97} \textit{Ibid.}, I.I.iii.1.  
\textsuperscript{98} Smith, \textit{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; in other words, the consensus. This constitutes an incentive for pro-social behaviour because conforming with social conventions is rewarded by others with social approbation, making its pursuit individually rational. This mechanism does, however, require the assumption that humans have a natural inclination to sympathise with each other. Consequentially, they approve of what they deem deserving of applause, which is a function of what they think others would applaud.

Mandeville cultivated similar ideas on 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”. 99

As seen before, Mandeville reduced behaviour to the individual maximisation of pleasurable passions without consideration for others, 100 combinding methodological individualism and hedonism. Mandeville then proposed that humans take joy in being applauded and considering “pride” as 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” 101

100 Mandeville, *Fable*, i.42.
101 Mandeville, *Fable*, i.30.
Because humans “are solicitous about gratifying… (their) appetites”, the desire for social approbation trumps the passions which the agent knows are not considered praiseworthy by others.

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 independently from categorical ethical propositions, just like assuming consumption-maximisation. In fact, social and moral preferences, which he called *virtue*, are not universal truths in Mandeville’s account of society.

“virtue, who from Politicks Had learn’d a Thousand Cunning Tricks, Was, by their happy influence, Made Friends with Vice”.

According to Mandeville, moral rules are variant over time and space because they are arbitrary tools invented by rulers and moralist to render people “useful to one another”. He proposed 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

---

102 Ibid., i.30.
103 Mandeville, *Fable*, i.7.
104 Ibid, i.34.
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, other-regarding distribution.

“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”

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 maximizing 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,

---

105 Mandeville, *Fable*, i.72.
the more they are forc’d to approve of his Behaviour”\textsuperscript{106}

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 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”.\textsuperscript{107}

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

\textit{IV.iii Evidence}

\textsuperscript{106} \textit{Ibid.}  
\textsuperscript{107} Mandeville, \textit{Fable}, i.7.
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. 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. 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.

Further research indicates how 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, making non-confession (cooperation) the preferred strategy. 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 being punished as retribution from previously-punished low contributors. 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 behaving antisocially, caricaturising compliance with social norms as a pre-emptive measure to avoid potential punishments for society. The notion that there might be a need for such a mechanism would suggest that Edgeworth’s contractual justification of selfish economic behaviour does not cover all externalities of an agent’s activity.

**IV.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.\textsuperscript{113} 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”\textsuperscript{114}

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. Similar to Smith, Akerlof and Kranton argue that “people have identity-based payoffs” which they derive “from their own actions” and “from others’ actions”\textsuperscript{115}. 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

\textsuperscript{113} Smith, TMS, I.i.1.
\textsuperscript{114} Ibid., III.i.5.
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”.116

The resulting view is that judgement about oneself is contingent on others, a clear challenge to methodological individualism. This mechanism takes the form of the impartial spectator in Smith.

> “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” 117

The impartial spectator, according to István Hont, is the mechanism by which an agent partially imports “a balanced normalised appraisal of society… into the psychological armoury of individuals”, therefore integrating social preferences as individual preferences.118 As such, the impartial spectator enables identity-based payoffs that do not require overt social approval. It simply rewards the agent when choices conform to the expectations of 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 regard for oneself. Mandeville’s and Smith’s ideas show instead how social approbation maximisation constitutes a desire that might drive pro-social behaviour that conforms to social norms, even anonymously. Nonetheless, the social approbation incentive is not holistic and only complements material self-interest.

---

117 Smith, *TMS*, l.iii.10.
V. Proposing a Simple Utility Function

Since neoclassical economic agency has been sanctioned through its integration into mathematical language, I will formalise the notion of social approbation maximisation derived from 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.

V.1 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:

\[ u_i(c_i, u_j(c_j, u_i)) \]

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 minimize inequality, implicitly caring for the material well-being of others.

\[ \text{Bergstrom, T., 'System of Benevolent Utility Interdependence' University of Mitchigan working papar (1988)} \]
Where $\gamma_i$ captures the agent’s disutility of having less than $j$ and $\beta_i$ 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.

**V.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, methodological individualism and material self-interest are 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 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:

\begin{equation}
    u_t(c) = c_t - \gamma_i \max(c_j - c_i, 0) - \beta_i \max(c_i - c_j, 0)
\end{equation}
\[ u_i(q) = \alpha \log S_i(q) + (1 - \alpha) \log A_i(q) \]

Where utility depends on the latently separable material self-interest, \( S \), and social approbation, \( A \), of an individual, \( I \), which depend on continuous actions, \( q \). Material self-interest encompasses commercial activity in which acquisition of wealth, consumption, or investment opportunities is the aim. These should account for the utility-enhancing variables traditionally studied by 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 brings 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 deciding 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 \( q \).

(1.1) \[ \frac{\partial S}{\partial q} > 0 \]  

(1.2)
\[
\frac{\partial A}{\partial q} < 0
\]

(1.3)

\[
\left| \frac{\partial A}{\partial q} \right| > \left| \frac{\partial S}{\partial q} \right|
\]

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 extra 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 indicates that the agent puts more weight on material self-interest.
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).

\[ A_i = A_i(o_i(q), c_i^e(q)) \]
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 decomposed into:

$$c_i = c_i \left(n, \bar{k}_i(q)\right)$$

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.

$$w_i^e = w_i^e \left(k_i^e(q)\right)$$

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.
VI. Conclusions

The aim of this paper has been to identify 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 surged in the past decades, as evidenced by the partial assimilation into economic orthodoxy of the experimental and behavioural research of recent decades. This constitutes a response to the neoclassical tradition which has justified the consumption based utility-maximisation that underpins benchmark economic theory. In this paper, I have aimed to revive and modernise a more comprehensive utility-based model of human behaviour derived from the thought of two classical economists – Mandeville and Smith. The capacity to hold pro-social preferences has revealed that material self-interest is a limited proposition to explain individualistic behaviour, partly because methodological individualism ignores social dynamics. The central idea obtained from these thinkers is that a behaviour 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 motive to become an incentive for collective action, it must be proposed that time-variant social norms aim for 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. 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.
BIBLIOGRAPHY


Black, A., A World History of Ancient Political Thought (Oxford, 2009)

Bellinger, K., The Economic Analysis of Public Policy, (Routledge, 2007), pp.50-63

Bergstrom, T., ‘System of Benevolent Utility Interdependence’ University of Mitchigan working papar (1988)


Crawford, Ian, ‘Introduction to Economics 2’, Oxford Faculty of Economics Lectures, (MT 2014)


Friedman, M., *Essays in Positive Economics*, (Chicago, 1953)


Hume, D., *A Treatise of Human Nature* (1738), iii

Hundert, *The Enlightenment’s Fable* (Cambridge, 1994)

Heilbroner, R., *The Worldly Philosophers* (New Yrok, 1953)


Rousseau, J., *Discourse on the Origins of Inequality* (1754)


Weber, M., *Economy and Society* (California, 1922)

Youyou W., Kosinski M., Stillwell D., *Computer-Based personality judgements are more accurate than those made by humans* (2014)