

The Aspiring Agent: An Account of Moral Personhood for Economics

Agnieszka Wincewicz

Submitted in fulfilment of the degree of Doctor of Philosophy
Geography, Politics, and Sociology, Newcastle University

September 2018

Abstract

The Aspiring Agent: An Account of Moral Personhood for Economics

Agnieszka Wincewicz

Could modern economics benefit from understanding the economic agent as a moral person? One that is not merely a holder of ‘moral preferences’ but takes active part in self-formation? In modern economics ‘homo economicus’ is routinely employed as an adequate account of the economic agent. This atemporal and amoral creature, limited to instrumental reasoning, is used to represent both abstract entities such as firms or markets, and individual human persons or groups thereof. Theories that translate economics into policy proposals ultimately intended to benefit individual human or communities (e.g. welfare economics) tend to inherit the thin concept of homo economicus as their implicit philosophical anthropology. But homo economicus, as economists happily admit, is not a person. It cannot then serve as a reference point in determining human welfare. Limited to instrumental rationality, it does not actively participate in determining its own ends. Human persons on the other hand, in taking responsibility for the larger part of their own lives, act as moral, not merely rational, agents. Large areas of economics have proper uses for homo economicus. But for areas where economics touches on the relations and actions of human persons, wherein instrumental reasoning holds insufficient explanatory power, I offer the ‘aspiring agent’ as a remedy for deficiencies of homo economicus.

The aspiring agent is the aspect of the human person that strives for self-betterment through ongoing, conscious attempts to refine her own preferences. She exercises practical reason in line with who she wants to become and her vision of the good. This ‘aspiring agent’ involves anthropological assumptions of the virtue ethics tradition of Aristotle through the Scholastics, and Adam Smith’s ‘impartial spectator’.

I develop the concept of the aspiring agent in analysis of certain recent attempts to correct for homo economicus’s deficiencies. Most attempts ultimately end up reverting to it. These attempts are found in the growing research trying to explain preference formation and change, including behavioural welfare economics and the related policy tool of nudging. I argue that the aspiring agent is a better concept for achieving their own stated research or policy goals.

Table of contents

Abstract	i
Table of contents.....	iii
Chapter 1. Introduction: The Human Person in Economics, Pre-Modern and Modern.....	1
1.1 Criticism of the Modern Abstraction	3
1.2 The Argument.....	5
1.2.1 Contribution	8
1.2.2 Structure.....	12
1.2.3 Methodology	13
Chapter 2. Neoclassical Economics' Account of Agency: 'not human but important'	17
2.1. The 'dogma' of Given Preferences	20
2.2. Homo Economicus and Changing Preferences	23
Chapter 3. Behavioural Account of the Economic Agent	27
3.1. Moral Preferences	28
3.2. Accounting for Preference Formation and Change	31
3.2.1. Multiple-self models	33
3.2.2. Sour grapes.....	38
Chapter 4. Towards the Concept of an Economic Person: Insights from Premodern Economics	43
4.1. Self-Formation and Virtue Acquisition in Aristotle.....	45
4.1.1. Oikos & phronesis	46
4.1.2. Eudaimonia & autarkes.....	47
4.1.3. Deliberation & choice	48
4.1.4. Habituation & virtue	50
4.1.5. Human action & moral responsibility.....	52
4.2. Aquinas and the Scholastics: The Economic Realm as Space for Personal Development.....	54
4.2.1. Scholastic certainties	54
4.2.2. The moral end of man	57
4.2.3. 'Prudentia' and moral action	59
4.3 Adam Smith: Moral Maturation, Responsibility, and Praiseworthiness	62
4.3.1. Self-interest, rightly understood	65
4.3.2. The impartial spectator	67
4.3.3. Moral maturation	69
Chapter 5. The Aspiring Economic Agent	74
5.1. Definition of the Aspiring Agent	75
5.1.1. Practical reason: proleptic and expressive	76

5.1.2. Aspiration as a learning process	79
5.1.3. Autonomy and responsibility	82
5.2. Alternative Interpretation of Preference Change	84
5.2.1. Multiple selves – not self-command but self-creation.....	85
5.2.2. Sour grapes: adaptive preference change as value learning.....	88
5.3. Aspiration vs. Welfare: The Good of Aspiring.....	91
Chapter 6. Economic Policy Illustration: Nudges	98
6.1. What Are Nudges?.....	101
6.2. Circling Back to Homo Economicus.....	103
6.2.1. Preventing mistakes	105
6.2.2. Nudging for the good.....	108
6.2.3. Skipping levels	112
Chapter 7. Nudging for the Aspiring Agent	116
7.1. Virtue-Developing Nudges.....	118
7.2. Opportunity for Discovery of One’s Ends	123
7.3 Responsibility for Choice	128
7.4. Nudging Entrepreneurs.....	133
Chapter 8. Conclusions	140
Bibliography	147

Chapter 1. Introduction: The Human Person in Economics, Pre-Modern and Modern

The way we tend to understand social, political, and economic institutions is to a significant extent determined by an underlying conception of the human person, whom we understand to be the subject involved in the creation and functioning of these institutions. From its very inception as a subject of enquiry, economics has been studied as the product of human agency under conditions of scarcity. These were understood to be real actions of real human persons.

It was the focus on the person that located political economy among the moral sciences in the tradition of Western political philosophy. The mark of this tradition, from Plato and Aristotle, throughout Christian antiquity, the Middle Ages (especially following Aquinas), and the Scottish Enlightenment of Adam Smith, is its placing of economic thinking within the area of moral actions of human beings. What are now called the human sciences (economics, sociology, etc.) were traditionally referred to as the moral sciences. The word ‘moral’ in this context refers to human rational agency which encompasses one’s ability to reason, to formulate one’s goals and principles of action, and to pursue one’s goals in line with those principles. The moral sciences were distinguished from the natural sciences in recognition of the fact that human action is purposeful and volitional, while the objects of the natural sciences are causally determined.

One important dimension of the pre-modern analysis of economic life was a sustained reflection on the complex interplay between conceptions of the human person and human well-being on the one hand, and social and political arrangements adequate to satisfy and sustain those conceptions on the other. Conceiving of the person as moral agent was crucial for these intellectual efforts. It was a particularly important tenet of the pre-modern tradition. The traditions I specifically refer to were based primarily in the normative framework set by virtue ethics. They assumed human economic activity to be part of a more profound endeavour and of a conscious effort to become a better person. The thesis before you explores the plausibility of this line of reasoning as a criticism of, and corrective to, contemporary economics.

With its emancipation from the theological and philosophical vision of man, modern economics gradually lost interest in the moral aspects of human action and agency. Already in the seventeenth century, economics and moral philosophy separated into two distinct paths of enquiry. Nonetheless, the human person as a moral agent continued to be treated in one way

or another as an important factor of economic analysis. The famous definition of economics proposed by Alfred Marshall in the late nineteenth century retained the spirit of that approach. In his *Principle of Economics* he called it ‘a study of mankind in the ordinary business of life. It examines that part of individual and social action which is most closely connected with the attainment and with the use of material requisites of well-being’ (Marshall 1890, 1).

Since around the middle of the twentieth century, however, there has been a tendency in economics (as well as in other social sciences), to avoid explanation of phenomena, which originate in human action, by appealing to the conscious decisions of persons in favour of impersonal factors (Evans 2002, 14). As the subject of economic science and analysis, human beings have increasingly been losing their distinctiveness as persons, as moral agents, being replaced by impersonal economic agents, necessarily non-moral. It is the latter, not the former that is considered the core subject of today’s economic analysis.

In post-WWII economic theory, the economic agent acquired a distinctly abstract character.¹ One important result of this has been mainstream economics’ gradual loss of interest in individual human persons as a normative and conceptual focus. Unlike persons, the neoclassical economic actors lack depth of motivations, of reflection, of commitment to values or ideas. These traditional constituents of personhood are no longer considered relevant for economic analysis. ‘Economic person’, or the older ‘economic man’ disappeared in the abstractions known under the currently heavily-criticized concept of ‘homo economicus’.

Economists agree that their science is about human activity and conduct. But they do not find explanation of all factors determining human actions as such—and specifically human actions as moral actions—to be necessary in their explanations of economic phenomena.² Economists of the formal neoclassical school grant that they abstract from the over-determined conditions of social reality, by disregarding all but the economic motives of actors in constructing theoretical models. They do not consider it necessary that their models should correspond with concrete human natures or even a general extra-economic idea of ‘human nature’. It is only in behavioural and in experimental economics that the causal phenomena identified with agents are usually said to be individual persons, not abstractions.

Even there, however, the person is portrayed in a simplified picture, which consists in an abstract construct of preference orderings. The origin of preferences themselves is not enquired into, and the relations between different orders of preferences within an agent are

¹ Cf. Ross (2012).

² The agents in most applied economics are aggregate. Ross (2012, 710) acknowledges, however, a standing pattern to be of disaggregating down to people when aggregate agency ‘hits trouble’.

conceived of within the normative precepts of rational choice theory. Neoclassical economics, with its focus on abstract rational choice theory, no longer needs a distinctive notion of a person to explain economic phenomena. Or so we hear from its practitioners.

1.1 Criticism of the Modern Abstraction

Philosophers, sociologists, and political scientists have long criticized economists' conception of agency as being, among other things, unresponsive to people's everyday experience and their own findings about the nature of human choice.³ Economists deem this criticism unpersuasive. It is based on the mistaken assertion that economic models are to be illustrations of real people, when in fact they are merely simplified accounts of phenomena that may or may not result from human actions. They are agnostic about whether the human person is the ultimate economic actor. Abstracting from irrelevant details of the real world, they say, is essential to their study, just as it is to any other scientific endeavour.⁴

Abstraction as such need not be a suspect activity. Generalizations or idealizations are certainly useful and have their place. Assumptions can be powerful explanatory tools, and they are needed when all the information cannot reasonably be known because of, say, the scale of the analysis. Problems related to human choice are so complex that it would be impossible to analyse them without some simplifying assumptions.⁵ As Hausman aptly put it, partly in defence of the simplifications used by economists, 'the real story concerning what people buy, how they invest, or what careers they choose is not one that a single discipline can tell' (2011, 118). It is crucial, however, that the simplifications are interpreted and applied as such in the right context.

For one important function of economic analysis is to serve as a toolbox for normative analysis of policy makers, which involves both predicting behaviour and assessing welfare impacts in relation to persons. It is not *about* abstract entities. Economic analysis, in turn, is

³ One of the most ardent critics of rational choice approach to human action is Amartya Sen. His famous 1977 paper 'Rational Fools' is a good starting point for those interested in details of his critique. Coming from an economist, his voice in the debate has been particularly significant and continues to be widely engaged with by both economists and philosophers.

⁴ On the role and use of unrealistic assumptions in economics see Mäki (1992, 1994, 2000, 2009).

⁵ What Churchland notes about modelling of brain activity could just as well be said by enthusiasts of economic models: 'We can learn things from the models that we might never have learned from the brain directly. And we can then return to the biological brain with some new and better informed experimental questions to pose' (Churchland 1996, 245). See also Danielson (1998) and especially Irvine's paper as a good example of the power of abstraction. The problem with economic modelling of agency is that it pretends that it need never go back to the human agent. It would be like the brain scanner denying that there need be neurons present in the brain, after basing analysis on neurological activity.

not the same as observation or description of specific economic activities or phenomena. It is not natural objects but constructed artefacts that are used to build economic models. When applied in policy, these artefacts are not to be confused with the real subject matter that they (purport to) represent. However, the complexities of theoretical abstractions on the one hand and of the social reality on the other, make the translation of results obtained in modelling into social and political practice difficult and confusing.

It is not uncommon that models or their elements are employed in an unreflective fashion as an aid in the formation of policy. It may also happen that ‘the original abstractions enter social policy as suppressed normative presuppositions’ (Myrdal 1969). This happens when an abstracted understanding of the economic agent is translated (implicitly or explicitly) into an actual assumption about persons, meaning probably citizens of some land, in policy formation. Tacitly presupposing, for example, that the theoretical assumptions about economic choice are a true reflection of how persons in fact choose, may lead to a policy design which does not ultimately promote the intended ends. It might even cause unintended negative consequences.

Recent developments in welfare economics, partially exemplified by the growing popularity of the so-called policy of nudging, are a case in point. Traditional welfare economics relies on the premise that welfare equals the satisfaction of preferences revealed in people’s choices, when these conform to assumptions of rational choice theory. Since the growing body of studies in behavioural economics shows that people rarely make choices that meet these abstract criteria, the standard way of defining welfare as preference satisfaction has been seriously challenged. Furthermore, since people’s behaviour has been shown to diverge from that of the theoretical homo economicus, welfare policy advice can no longer comfortably rely on this assumption.⁶

With the increasing popularity of behavioural economics, both among scholars and policy makers, many aspects of human choice, which were traditionally regarded as insignificant in standard neoclassical theory, have been gaining importance. This is challenging the prevailing view that there is no need for economists to delve into the complex issues determining human choice. It is now more readily admitted that better understanding of how people choose can help develop better ideas about what contributes to their welfare and design better policies.

Nevertheless, characteristically and perhaps unsurprisingly, there is a tendency among economists to incorporate the various discoveries about human decision-making into the

⁶ For a detailed argument see Hargreaves Heap (2013).

standard framework of economic explanation, based mostly on the tenets of rational choice theory. Various aspects of the theory are modified accordingly, but the overall structure of explanation remains largely intact. Despite the modifications, the economic agent continues to be defined in the language of preferences which limits the explanation of human choice to types of reasoning exercised by the abstract homo economicus (calculation, means or ends-based reasoning, utilitarian calculi). Important aspects of human agency which are not conceivable in the rational choice framework are left outside of the picture. With an increasing role of economics in informing public (not merely economic) policy, like nudging, it seems pertinent to revisit some of the aspects of the mainstream economic theory which are finding new application.

1.2 The Argument

Addressing some of these recent developments, this thesis argues first, that a better understanding of how people choose and what constitutes their welfare requires that human agents are understood and explicitly treated as moral, and not merely economic, agents. For, as I shall illustrate, moral considerations are inseparable aspects of economic activity. They co-determine the choice behaviour of economic agents and thereby direct the outcome of economic activity. Abstracting them away or forcing them into the strictures of rational choice theory limits the extent to which many contexts of human choice can be meaningfully explained and translated into policy.

Second, I develop an account of humans as aspiring agents, rather than either rational, self-interested actors in full knowledge (neo-liberal) or class-interested, pre-determined servants of the material forces and structures of economics. I offer the aspiring agent to show that moral agency cannot be treated as an extra ingredient in the existing framework of economic explanation. I argue instead that moral agency is an element which challenges the very fundamentals of this novel framework, just as it provides the source of explanation of human choice.

Third, I show how this view of human agency can help explain some contexts of economic behaviour and inform policy by offering a better understanding of the concept of well-being. I argue that simplifying assumptions concerning human agency used in the standard economic approach are an insufficient basis for economic policy, especially policy designed to increase welfare.

The traditional way of approaching economic analysis relied on the common-sense

view that human economic activity has its source in the human need and desire for improvement of one's well-being. Included in that were one's family or household, however broadly or narrowly these were variously construed. This is the original reason why people buy, sell, employ and invest, borrow and lend, innovate and create. Often such decisions are at least partially driven by social and moral considerations. This, in turn, opens our economic behaviour to both economic analysis and moral scrutiny.

Recent revival of the problem of morality of economics has assumed a relatively narrow focus. On the systemic level, the question of morality of market institutions has been posed; whereas on the individual level, the problem of self-interest vs. other-regarding sentiments has been addressed. Characteristically, these problems have been discussed mostly by non-economists. Economists of the neoclassical school, who hold the rational choice theory as the only relevant framework for the analysis of choice, continue to regard them as unimportant for their purposes. When the importance is conceded, they are nonetheless rejected as unformalizable along the lines of assumptions of rational choice theory.

In this thesis, the way I approach the moral aspects of human conduct is intended to engage critically some core elements of modern economic theory of preferences. It thus challenges the prevailing view that economists need not engage in the analysis of moral agency. For purposes of explanation and prediction, we are told, neoclassical economic theory requires that that behaviour of the economic agent under investigation conforms to a set of rational choice axioms. More specifically, we are asked to assume the agent's preferences are given—that is, complete—and stable. To the (limited) extent that the behaviour of persons corresponds with these assumptions, economic activity can be explained in the framework of rational choice theory.

However, unlike the abstract *homo economicus*, persons have the capacity to form, revise, and change their preferences as they see fit.⁷ They can even have no preference at all! While the mainstream theory assumes complete and transitive preferences which implies that all aspects of a choice setting can be weighed off against each other and all alternatives can be ranked in terms of 'betterness', people can sometimes refuse to do this (and for good reason). The mother in the famous novel 'Sophie's Choice' who has to choose whether her daughter or her son lives is an example of a dilemma which makes formation of a well-defined preference impossible. A less dramatic but not less present in everyday life is an example of norms, rules, principles and values people embrace. When we uphold them, we generally act non-instrumentally and sometimes somewhat automatically, which is opposed to the instrumental

⁷ This is view famously formulated by Rawls (1993).

and deliberate nature of forming and satisfying preferences. What motivates people in most of these circumstances is not some conscious and deliberative calculation but rather a disposition to act according to some value.

These capacities are together often described as the defining characteristic of the person. In a positive light, it is the ability to form one's own will, one's free will. Danielson (1998) puts it in a rhetorically powerful phrase, that could also be less obviously positive: 'The special rationality of persons, then, consists essentially in the capacity to be irrational'.⁸ An economic model of man that cannot deal with the irrational or non-rational—or even defines them away—would seem to have clear limitations when dealing with agents who so obviously fall short of rational action.

Let me reiterate that in some fairly trivial contexts, rational choice theory can also explain choices, even those revealing incomplete and unstable preferences. But in many other situations, and in the most important and intimate ones at that, this approach is insufficient. In such cases a different framework of explanation is called for. One thing to remember is that, although in such situations the agents in question do not meet the strictures of economic rationality, they could nonetheless serve as important objects of economic analysis. These are actual situations characteristic of many consumer, investment, and entrepreneurial decisions, as well as choices involved in our work relations and relationships with ourselves and the environment and other people. These situations call for decisions which are not—or at least not exclusively—driven by instrumental means-end reasoning. They may be better characterized by the agent's openness to unforeseen possibilities and learning. They are driven by many things, but often by what I follow other scholars in calling 'aspiration'.

In what follows I call an agent engaged in this open-ended reasoning an 'aspiring person'. This image of human agency guides the arguments developed in the thesis. In the language of economics such a person can be described as someone who wants and tries to improve her existing preferences without having a specified understanding of the improved preferences themselves. One chooses in partial knowledge, and might continue to do so as the choice, and the *meaning* of the choice for her self-development, become clear. Often the choice is not once-for-all, but the repetition of the same choice (continued purchasing of vintage wines), or the reaffirmation of a standing choice (celebrating marriage anniversaries as reaffirmations). Insofar as this effort can be seen as an element in the process of self-formation and development of one's idea of the good, it is to be regarded as part of human

⁸ Similar ideas are exercised by Charles Taylor (1985) in his notion of human identity which is defined by our fundamental evaluations, that is articulations of what is worthy, more fulfilling, etc.; and Harry Frankfurt in his essay 'The Importance of What We Care About' (1982).

moral agency. Obvious general examples are quitting an addiction or taking up a challenge to become physically fit. An aspiring agent is in this way always also a moral agent. It is someone who strives for self-betterment through conscious formation and change of preferences—meaning, constant refinement of the preferences, and the reasons for (continued) preferring of something or someone over other possibilities.

This view of human agency is ultimately rooted in the Aristotelian-Thomistic tradition of virtue-ethics. As I try to show, the most important parts of that tradition were also shared by Adam Smith. The analytical framework of aspiration provided recently by the philosopher Agnes Callard (2018), lends scholarly rigour to reflections which might otherwise lack sufficient analytical structure.

1.2.1 Contribution

The growing body of research in preference formation and change, within behavioural economics and in decision theory, has exposed the analytical significance of various factors informing people's preferences. Relaxing of the assumptions of fixed and given preferences has been shown to be a useful strategy to better understand why human choice is not like that of the abstract homo economicus. Openness of people's ends is now more readily acknowledged in the study of choice, including economic choice.

All of this creates an encouraging research environment for the economic study of a more common view of human choice. It involves the capacity to form, revise, and pursue one's own conception of the good, a view that contrasts with the standard neoclassical approach of one single rationality of utility maximisation.⁹

Yet, with all their appreciation of various cognitive and psychological factors as determinants of choice behaviour, these recent studies do not take the step of connecting their conclusions to the moral aspects of human agency. There is very little concern about the moral capacity for self-formation. This thesis offers one way to fill this gap. The understanding of the human economic agent as an aspiring person links the economic and moral discourse in a meaningful way that is also relevant for economic theory and policy.

On the level of economic theory, my thesis contributes to exploring ways in which the growing evidence of preference instability can be addressed in economic research. The concept of an aspiring person challenges the assumption of the well-informed and decisive

⁹ "Choice", as John Dewey (2008, 96) put it, "signifies a capacity for deliberately changing preferences."

economic agent as the standard unit of analysis. In so doing, it shows how this assumption narrows the scope of explanation of choice, that is available in economic experiences and data, to what is merely formalizable in the language of rational choice theory. It shows that the formal strictures of this theory make it impossible for it to capture an image of an evolving self, indicative of a conscious change in one's values and preferences. Like applied maths before calculus, the standard models can only understand change by imagining each state to be static.

The claim that standard rationality theory does not adequately explain human choice has been addressed in much of the research in political theory, sociology, and philosophy. Various aspects of the complexity of human economic agency have been explored there. Among them, the human striving for self-betterment and self-formation—*aspiration*—which appears to be a crucial factor of moral agency, has received only some piecemeal (and otherwise marginal) treatment. This study is a closer analysis of aspiration as one particular element defining persons and makes a more explicit statement of its relevance for economic theory and policy.

It borrows especially from valuable insights offered by Frank Knight and James Buchanan. They criticized modern economic theory for insisting on patterns of thought that make elementary recognition of many aspects of our behaviour very difficult to analyse and easy to neglect. Knight was one of the few twentieth-century economists who openly confronted the inadequacy of 'rational economic man' to explain human choice. He recognized that rigorous and useful models of economic maximization required that human beings be described as purposely and consciously utilizing means to attain predefined ends. At the same time, he contended that 'there is no such man', because human beings do not know what they want and what is good for them, and they 'do not act very intelligently to get things which they have decided to acquire'. And he did not presume that human beings know even their most basic wants: 'We propose to suggest that these wants which are the common starting point of economic reasoning are from a more critical point of view the most obstinately unknown of all the unknowns in the whole system of variables with which economic science deals' (Knight 1922, 20).

In another famous observation Knight states: 'the chief thing which the common-sense individual wants, is not satisfactions for the wants he has, but more, and better wants' (Knight 1922, 42). Consequently, Knight always cautioned against the over-zealous application of economic theory to phenomena, the explanation of which extends beyond formal economic theory, and argued that economic theory is not the explanatory science of all reality. That is a form of scientific modesty. I would go a step further. Economics is not even the explanatory

science of economic activity. Ethics and moral science are. Economics is the describer and systemizer. But it is rightly part of a larger moral science.

Similar reflections to those of Frank Knight underlie much of Buchanan's work. In a relatively unnoted essay, 'The natural and artefactual man', Buchanan (1999) talks about one's sense of becoming as a central, if not the most important, fact of human life. He is critical of the teleological foundations of modern economic theory, with its tendency to force all analyzable behaviour into the straitjacket of 'maximizing a utility or objective function under constraints'. In contrast to this approach, he suggests that it is useful to think of man as an imaginative being, to allow for some recognition that persons imagine themselves to be other than they are, and that they take action to achieve imagined states of being.

My thesis builds also on the recent study of the individual in contemporary economics offered by John Davis (2011). He argues that the rise of scientific formalist economics is to be blamed for the 'destruction of the individual as a moral centre of the economy'. I do not directly defend that broad, and plausibly true, argument. But I do rely on its authority, as I make use of its conclusions.

Insofar as the thesis proposes to explore a neglected dimension of personhood in economics, it could be seen as yet another attempt to enrich the formal concept of the economic man with one more property that cannot be found in the standard assumptions—thus solving the problem of *homo economicus*.¹⁰ Previous attempts of this kind have been mostly shown to be reducible (or to return) to the standard account of economic agency. They were therefore not useful toward their intended end of solving the problem.¹¹

Since the concept of moral agency proposed here cannot, I argue, be captured in the standard analysis, this is not a matter of simply extending a concept of *homo economicus*. I am showing the limitations entailed in this abstract construct. My goal is not to expand the lengthy list of alternative terms which have been suggested to account for the allegedly neglected qualities of the economic man, by expanding the catalogue of his preferences or giving him a heart full of universal empathy. Instead, the notion of agency developed here is intended to bring out of the shadows the anthropological reason why the economic understanding and explanation of human choice in the very terms of preferences cannot account for what it sets out to. The focus of this thesis is not on economic methodology itself as much as on the implications of the mainstream economic thinking about preferences for policy and possible ways to better inform policy expertise with a richer anthropology of economic agency. The critical account of the *homo economicus* is intended to show how

¹⁰ Previous ideas involved the extension of the concept into *homo socioligus*, *homo politicus*, etc.

¹¹ I discuss this in detail in Wincewicz-Price (2016).

inadequate mainstream theory is for capturing the moral dimension of human decision making.

On the policy level, the thesis sheds light on the problem of the application of the abstract reasoning which is ascribed to idealized economic agents and then applied to policy. Policy deals with less ideal and more complex persons, so much so that the policy results are routinely hard to steer predictably toward the good of their intended beneficiaries. One new policy innovation informed by behavioural welfare economics, the policy of nudging, is a good example of how important it is for policy makers to understand human choice in a broader framework than that of the abstract homo economicus. This innovative policy framework has been discussed from many different angles, both theoretical and practical (e.g. Bovens, 2009; Hausman and Welch, 2010; Goodwin, 2012; Niker 2017).

My own approach is to some extent related to the existing criticism, but I put it in a slightly different perspective. I ground my scepticism about the policy of nudges in the methodological criticism of the concepts of preferences, rationality, and welfare implicit in its theoretical underpinnings. One important problem with the policy of nudging is diagnosed in the existing literature. Although proponents of this strategy openly recognize that human persons are not like the theoretical homo economicus, proponents simultaneously promote policy instruments which propose this idealized agent as a normative benchmark. Nudges are meant to improve people's choices by helping them to choose as if they had much better understanding of their goals than they in fact do. It is suggested here that nudging can be less controversial and more effective if its normative objectives are based on the concept of the aspiring person who does not fully know his ends. The aspiring person is not however ignorant of his own ends. He knows enough of them to act according to a general life plan or style, and adapt along the way. Policies intended to nudge aspiring persons would necessarily be more contextual than those deployed against. This I explain in the substantial chapter on the aspiring person.

Despite the undeniable dominance of neoclassical economics in the mainstream of economic research, many 'heterodox' approaches, schools, and traditions exist, challenging the rational choice orthodoxy. They have emerged in recent decades, offering interesting alternative analytical frameworks for the study of economic phenomena. The notion of the aspiring agent could be quite comfortably incorporated into some of these propositions, especially in the rapidly developing area of evolutionary economics.¹²

¹² Some evolutionary economists go as far as to claim that selection and ordering of ends proves to be the most fundamental economic problem and lament the abandonment of the problem by orthodox economics. See, for example, Fudulu (2014).

1.2.2 Structure

In this study, I claim that the understanding of economic agents as moral agents remains relevant, if not imperative, both for economic theory and practice. Abstracting away from the important aspects of what makes us persons, while useful in many leading areas of economics, cannot readily be justified in discussion of certain economically-informed policy issues. Policy aimed at increasing personal welfare is an obvious example.

The thesis consists of three parts. Part I, consisting of chapters 2 and 3, identifies major images of human agency in modern economics.

Chapter 2 sketches the standard image of the neoclassical, self-interested economic man constrained by the assumptions of given and stable preferences. It shows that the abstract and axiomatic nature of this account makes it possible to capture human agency only to a very limited extent. Important factors of moral agency are considered as external to the economic model of choices. I indicate that some of these factors, such as the openness of human behaviour to change and novelty, deserve closer attention in the economic analysis of choice.

Chapter 3 reviews the main lines of criticism of homo economicus developed in behavioural economics. It discusses some important attempts of behavioural economists to challenge the neoclassical assumptions related to self-interest and fixed preferences. It argues that the image of human agency retained in behavioural economics is also limited and insufficient to capture the dynamic and open-ended nature of human choice. For, it is by and large that of homo economicus *mutatis mutandis*. It suggests, however, that their work in the area of preference formation creates space for further elaboration of some complexities of human choice behaviour. This can serve as a starting point in the discussion of preference formation understood as an important aspect of moral agency. Two examples of preference change models are analysed to support the argument of this chapter: multiple selves and ‘sour grapes’.

Part 2, consisting of chapters 4 and 5, explains the notion of aspiration as a constitutive element of human agency. It draws on the virtue-ethics tradition, as well as the philosophical framework of aspiration developed by Callard.

Chapter 4 consists in an historical study of the fundamental tenets of the virtue-ethics tradition. In this tradition, self-formation is seen as a crucial element of human agency, distinguishing human from non-human creatures. My guides in this enquiry are Aristotle, the scholastics, with Thomas Aquinas as their most prominent representative, and, perhaps most importantly, Adam Smith. His views on human agency have been profoundly and unjustly

neglected by most economists.¹³ To see Smith interpreted through the lens of this tradition may surprise some economists, but section 4.3 should justify his inclusion in the company of Aristotle and Aquinas. The shared moral philosophy of these thinkers will serve as building blocks for the conception of an aspiring agent, which is developed in the following chapter.

Chapter 5 gives an account of person as an aspiring agent. This being is shown to be a proleptic and expressive reasoner, rather than merely instrumental as in rational choice theory. She is open to learning, and responsible for her own self-formation. This anthropology is then used to provide alternative interpretation of two models of preference change from Chapter 3: multiple-selves analysis and ‘sour grapes’. They were previously shown to be problematic when seen in the rational choice framework. Some preliminary policy implications are suggested in the concluding section of the chapter.

Part 3, which is chapters 6 and 7, suggests possible applications of the idea of an aspiring agent.

Chapters 6 and 7 offer an illustration of how the concept can inform policy. They discuss ways in which it can help the debate around the justification of the use of nudging. It is claimed that, insofar as nudges interfere with the agent’s process of preference formation, nudges can be seen as potentially welfare-reducing. For, they inhibit virtue-acquisition, leaving the agent always biking with training wheels. However, insofar as they create opportunities for the agent to aspire to a good or value he might have not otherwise discovered, nudges can be seen as potentially welfare-enhancing.

The closing chapter also concludes the argument, putting it in a broader context of a recent approach to economics that attempts to reclaim it as a moral science. I argue that ethics and certain moral capacities are a relevant and inseparable aspect of economic activity. I also indicate future trajectories for research, particularly in economic methodology and the emerging discipline of behavioural welfare economics.

1.2.3 Methodology

The argument advanced in this thesis necessarily crosses the boundaries of several disciplines. It tackles the important problem of conceptualisation of the human in economic theory and

¹³ While philosophical ethics, especially that of Aristotelian-Thomistic variety, has been widely used in business and management literature, it does not seem to have provided much inspiration for non-business-related problems of economic theory. Likewise, Smith’s *Theory of Moral Sentiments* has been treated somewhat superficially by economists who usually construe his concept of sympathy and other morally-driven feelings or attitudes as, in the end, conducive to the pursuit of one’s self-interest.

policy with a mix of tools borrowed from philosophy of economics, history of economic and political thought as well as political and moral philosophy. Such multidisciplinary approach seems necessary given the narrow confines of modern economics and the research strategy assumed in the thesis, namely one of informing contemporary problems with insights from pre-modern theories.

Broadly speaking, the thesis attempts to integrate elements of several approaches and discourses concerning the place of human person in economics. Individual chapters can be seen as specific contributions providing contexts to a more elaborate conversation about various ways of portraying human economic agent that runs through the whole study. The thesis is not meant to imply that the entire economic discipline needs to be reformed or overturned. It does not take on a revolutionary goal of displacing the prevailing ways of doing economics. It merely identifies important explanatory limitations of the narrow framework of the rational choice theory on which the discipline relies so heavily. It also offers important qualifications which should be considered when applying economic theory of choice to policy. Finally, it suggests that economic analysis can be more meaningful when its abstract mathematically tractable concepts are reintegrated into the philosophical and moral discussion about human actions and human goals. Simplistic moral philosophy of economics deserves special attention in questions of well-being which are ethical in nature and need to be considered in a broader context than that of instrumental means-ends reasoning.

In the diagnosis of the research problem, the neoclassical depersonalised and “disembedded” (to borrow the classic term coined by Karl Polanyi) homo economicus is contrasted with the premodern “embedded” human identified in works of Aristotle, Aquinas and Smith. In another context of the discourse, the neoclassical rational agent is criticised from the perspective of behavioural economics and contrasted with some of its extensions which have been proposed in this growing sub-discipline of economics. In the central discussion, an alternative account of human economic agent is proposed to point out the limitations of the standard Homo economicus and its extensions developed by behavioural economists. This alternative concept is called an aspiring person. It is the main character of the story that is being told here. His (or her) defining characteristics are informed by the features that are identified as common for the premodern accounts of human agent, and a recent insightful conceptual analysis of the agency of aspiration offered by a contemporary moral philosopher, Agnes Callard, who has developed some of these premodern ideas in a framework which seems ideal for the analysis of the problem discussed in this thesis.

Homo economicus came out of debates in philosophy, theology and the moral sciences, finally settling in the 20th century as the accepted understanding of the economic

agent. As I see it, one can take two routes to analyse the concept. One pushes further into economics: further up to the structural level or further down, by means of dissecting the Homo economics the way it has been tried in behavioural economics. The second route involves a move away from economics in order to gain some perspective. There are, in turn, at least two ways to achieve the latter. One can either apply contemporary scholarly tools to the problem borrowing from other social sciences (coming up with concepts like homo sociologus, homo politicus, homo psychologicus, etc), or rediscover the distillation of the concept of the economic agent, noticing what was evaporated along the way. Taking this route as I do in this thesis necessitates a multidisciplinary engagement with the major arguments identified here.

The strength of this approach lies in the opportunity it creates for discovery of new insights into economic theory and policy. Homo economicus promises much in terms of explanatory power. However, as an oversimplified image of human, it cannot serve as a standard for normative theories that inform policy which is made for actual persons or any given group of persons. It can only deliver within the confines it is itself defined by which excludes many of the reasons persons tend to hold and change their preferences and reveal inconsistent choice behaviour. By design it excludes the moral life of humans. Taking moral personhood in the conception of economic person does not overpromise its uses but neither does it underdeliver.

One of important weaknesses of the multidisciplinary approach could be identified in its eclecticism which is not always seen as a virtue in modern scholarship. Insofar as this thesis is an attempt to systematise a problem that has been touched upon in multiple disciplines, I think this approach can nonetheless be justified. The chapter developing the concept of an aspiring person provides the justification.

Why choose welfare policy and specifically the tool of nudging for the analysis of the limitations of Homo economicus as a normative standard? For one, these are now some of the most popular areas of inquiry in applied economics, with devotees in policy circles up to the highest levels of government (primarily in the US, the UK, and The Netherlands). More importantly, however, they are particularly relevant illustrations of my argument since, as Chapter 6 shows, although they purport to move away from traditional welfare policy based on the standard rational choice theory approach, they in the end return to use the model of homo economicus as the economic agent.

A few clarifying points regarding terminology used in the thesis are in place. In what follows I look at the moral agent as an aspiring person who strives for self-betterment through a conscious formation and change of preferences and whose choice behaviour cannot

therefore be fully accounted for by the tools of the rational choice theory. “Moral” here refers to the tradition which treated economics as part of a broader area of inquiry called moral sciences. In this tradition to be moral was identified with distinctly human capacity for the choice between good and evil. Throughout the thesis, unless stated otherwise, I will use the term “moral” in this sense. “Moral” should not therefore be confused with the actuality of goodness (or lack thereof) but the possibility for goodness (evil). This capacity is not present in Homo economicus which cannot choose between good and evil. Homo economicus is merely rational or irrational in a narrow way that does not account for moral concerns.

Self-betterment, or self-formation are terms used here for the experiential side of welfare policy. Some aspects are considered to be shared by everyone and are easily measurable (e.g. concerns of health and safety), whereas others are bound to specific groups (e.g. definitions of some mental illnesses, questions of what is found to be honourable and dishonourable) or even individuals (e.g. issues of lifestyle). In this perspective, the question of whether self-betterment is objective or subjective is beside the point. A more relevant question is what do we do with the large area of self-betterment that falls outside of that which is universally understood to serve the person’s good, such as general concepts of health and safety.

It should be noted for clarification that even though it may seem from the foregoing that the aspiring agent is called for to account for an entirely separate and separable aspect of personhood, this would be a mistaken impression. Such separation is analytically useful, and it is with this aim that it has been discussed here as if it could mean an independent aspect of the human. An actual person, however, integrates, of course, these various aspects in ways which are often hard to discern. The neatest way to assert the existence of the aspiring agent is to claim that human action, even when conceived of in a narrow context of economic rationality, cannot be wholly explained by assumptions made for the abstract homo economicus, because unlike the latter people take active part in self-formation.

Virtue ethics approach which is taken here dictates the redefinition of the overarching question of ethical theory from ‘What I ought to do’ to ‘What sort of person ought I to strive to be’. This ethical tradition is less concerned about duties to others and addresses more directly questions of personal excellence and societal flourishing. It understands moral self-formation to require not only self-less beneficence, but also other-regarding and self-regarding virtues. Personal excellence is directed at better social coexistence. It is built around enlightened self-interest – not altruism.

Chapter 2. Neoclassical Economics' Account of Agency: 'not human but important'

This chapter explores the neoclassical economist's understanding of agency that is implicit in the standard theory of rational choice, and shows where and why it departs from a realistic account of human agency.

Economists use the concept of agency in order to ground formal theories of practical rationality (Ross 2002). The place of agency in economics should thus be understood in the context of the logical structure and goals of a given theory. The logical structure of neoclassical economics has been founded on rational choice theory. Thus, the concept of agency employed in standard micro- and macroeconomics is one of the rational agent. Rational agency in the economic sense originates directly from the axioms of rational choice theory. It should not be confused with other concepts of rationality employed especially in philosophy or ethics.¹⁴

The neoclassical concept of the economic agent derives from the work of Samuelson (1947). Samuelson's fundamental project was to drain the claim that economic agents maximize utility of its psychological content. A Samuelsonian economic agent, as Ross (2005, 245) defines him, is simply 'any system that observes certain consistency conditions in its behaviour, such that it can be interpreted as if it is maximizing the value of a function that maps a system of preferences over commodity bundles onto the real numbers'. 'Economic' here is defined by reference to an axiomatic theoretical system for identifying equilibria among behavioural dispositions or strategies of agents.

Standard rationality theory, or expected utility theory, is axiomatic in nature, in that it is formulated around a specific set of logical assumptions (completeness, transitivity, independence, and continuity) chosen in order to produce well-defined preference orderings (von Neumann and Morgenstern 1944). The 'von Neumann-Morgenstern utility function theorem' states that 'if these four axioms are satisfied, any set of well-ordered preferences can be represented by a distinct (monotonic) individual utility function'. The utility function is the theory's representation of the agent. Just how crucial the assumptions of rational choice theory are for economics has been concisely summarized by Gary Becker in his well-known statement that 'The combined assumptions of [utility] maximizing behaviour, market equilibrium, and stable preferences, used relentlessly and unflinchingly, form the heart of the economic approach' (Becker 1976, 5).

¹⁴ Decision theory, game theory, and even much of political theory, in turn, approaches rational agency in a way that is analogous to neoclassical economics.

In what follows I will discuss what these assumptions mean for the concept of economic agency, focusing especially on the two most substantive assumptions: preference completeness and transitivity. The discussion is intended to show that it is not important for rational choice analysis that the economic agent is either a human being or a person. The issue of what constitutes the individuality of this agent never arises. It is not a problem that economists find interesting for their purposes. The economic agent is an abstraction, a useful but merely theoretical construct of the smallest unit of activity under analysis. As Becker (1993, 402) writes,

while the economic approach to behaviour builds on a theory of individual choice, it is not mainly concerned with individuals. It uses theory at the micro level as a powerful tool to derive implications at the group or macro level. Rational individual choice is combined with assumptions about technologies and other determinants of opportunities, equilibrium in market and non-market situations, and laws, norms, and traditions to obtain results concerning the behaviour of groups. It is mainly because the theory derives implications at the macro level that it is of interest to policymakers.

The abstract notion of economic agency is usually justified on grounds internal to the goals of economics. Ross argues that nothing in economic theory requires that economic agency be identified with individual persons, because economists use it to build abstract models of firms, nations, labour unions, consortia in auctions, lineages in evolutionary games, and other feedback-sensitive, incentive-driven systems that have no evident psychological properties at all. Economic theory can be about persons, or it can be about emergent systems of production, consumption, and exchange, where the ultimate units of these activities are not necessarily to be identified with persons. Most economic theories study markets and networks, employing for this purpose an idea of agency that is related to the concept of the person only by historical semantic tradition. Economic agency is a theoretical construction elaborated as part of the development of a family of models. Even the majority of applications in the area of economics most directly related (at least in principle) to the theory of choice, consumer theory, make no necessary working use of the supposed identity of economic agents and persons.

It is usually argued that any requirement that these agents be individual persons requires an extra-economic motivation. That is why most economists never estimate the utility function of an individual person. Rather, they measure elasticity coefficients of aggregate demand and production functions from changes in prices, interest rates, income

distributions, national savings rates, and other index quantities.¹⁵ In game theory, it is not even agents but strategies themselves that are the players of the games, with agents merely standing in to play their brief turns in a competitive process that continues beyond their individual lifespans. Economic agents are ‘representative’ optimizers, whose ontological status is indeterminate. Davis (2006, 70), for example, notes that there is nothing in the current understanding of choice that requires it to apply to any type of agent in particular:

The representation of choice now dominant in economics, which involves nothing more than maximisation of a formal objective function, applies equally well to any type of agent, whether it be a single individual, a collection of individuals, an animal, a collection of animals, or even a programmed machine.

Although nothing in the theory requires that the agents in question be persons, much less human persons, this fact alone does not mean that economics is uninterested in individual behaviour. Its interest in individual behaviour, however, is usually limited. And it needs to be seen in a broader perspective. Most economists study human choice because it is a building block in the analysis of social phenomena, rather than because they are concerned with explaining individual persons qua individual behaviour. It relies on the abstract conception referred to as *homo economicus* which co-defines rationality and individuality, not in connection with human behaviour, but by reference to the logical-mathematical properties of equilibrium analysis. As Debreu stated: abstract individual conception refers to nothing in the world whatsoever (1986, 1265). *Homo economicus* is thus not to be understood as ‘person’. It is more accurate to conceive of this theoretical construct as the ‘representative agent’ (e.g. repeated markets) and not a generalization of any potentially real person. It does not provide an empirically relevant model of individual behaviour. Yet, *homo economicus* is a convenient model for abstract economic analysis, the bigger the better. That is the reason why these economists are committed to its notion of underlying coherent preferences, among other empirically unverified (and often unverifiable or false) positions about the ‘agent’.

Philosophers at times hold the mistaken view that rational choice theory is psychological and individualistic, and applies to the person’s preferences. These preferences

¹⁵ Most applied economists pay lip service to the idea that all of these things somehow ‘boil down to’ decisions by individual people. For example, textbooks in international economics admit that so-called ‘community indifference curves’ used to represent national welfare cannot be disaggregated into individual indifference curves without destroying the point of using them; most books cheerfully suggest this as a cautionary note and move on without further ado, assuming that the idea of ‘national welfare’ makes sense in its own right.

are then understood by philosophers as the person's mental states. In this perspective, only individuals have preferences. Economists, however, despite some notable exceptions (e.g. G. Becker 1976) have been sceptical about rational-choice theory as a psychological theory.¹⁶ Instead they have embraced purely behaviourist interpretations, and generally seem to endorse what is often referred to as a reductionist program (e.g. Satz & Ferejohn 1994). The theory considers only the formal or mathematical properties of the agent's preferences while the content of those preferences is considered irrelevant to the theory. It is also irrelevant why the agents hold certain preferences and not others.

The theory of formal rationality examines only the relations among the agent's preferences. It does not judge the rationality of these preferences. The conception of human rational agency in terms of maximizing over a complete and consistent set of preference orderings is not psychologically realistic. Rational choice explanations are not therefore psychological explanations. Insofar as they are applied to human agents, they explain patterns of action within well-defined environmental constraints and looks for regularities among those patterns. On this account, rational-choice theory can be applied wherever there is a patterned behaviour. Studies have shown, for example, that pigeons do reasonably well in conforming to the axioms of rational-choice theory.¹⁷ But it is not fit to explain behaviour of any particular agent, without being accused of trivializing it. All that formal rationality can posit about an individual agent's action is that it is explicable, *as if* the agent is maximizing satisfaction of his preferences.

Below I discuss two substantive and heavily criticized assumptions that are fundamental in the modelling of a representative economic agent's preferences. I do so in order to illustrate in more detail why homo economicus seems to be a particularly misidentified representation of a person. The assumptions hold that the economic agent's preferences are given, and that they remain stable over time. These assumptions operate in rational choice theory to render it fit only to explain choice in a very restricted sense. 'Choice' here should not be confused with the way real people decide, choose, and act.

2.1. The 'dogma'¹⁸ of Given Preferences

Rational choice theory takes preferences as exogenously given. The standard theory does not venture to explain how they are formed. In the traditional approach, the assumption of given

¹⁶ Cf. Samuelson (1938).

¹⁷ Cf. J. Kagel (1987).

¹⁸ McPherson (1983).

preferences is often justified, quite legitimately, by the fact that economists have very limited knowledge about what people's preferences are, how they are formed, and what influences them, with one exception. Preferences proceed unambiguously from a single source, namely, self-interest. Preferences are therefore also assumed to be given unambiguously, referred to in the literature as the completeness axiom. In this, they do not reflect any conflicts of loyalties or interests. If such conflicts existed, they need to have been resolved prior to preferences formation. More thorough understanding of this assumption is crucial for the purpose of this study. For, economists in the neoclassical tradition begin their analysis of choice at the point where ends are (or are assumed to be) set. Their task is to specify structures that are appropriate means of achieving those ends, whatever they may be. When human persons are the focus of the analysis, individuals are expected to determine their own ends. Economists are, rather, addressing the next step in the process toward those ends, beginning from wherever they find the persons.

Hausman (2011) has recently offered a clarifying definition of preferences, including how they should be understood in the context of economic modelling. He shows that the assumption of given preferences can only make sense and be useful under a very specific notion of preferences. Many non-economists (such as the philosophers mentioned immediately above) and some economists, too, speak somewhat misleadingly of preferences as if they were psychological phenomena. But what the term should be taken to mean is a formal ordering relation, constructed so as to allow equilibrium analysis. Hausman defines them, then, as 'total subjective comparative evaluations'. He illustrates this with the example: "To say that Jill prefers *x* to *y* is to say that when Jill has thought about *everything* she takes to bear on how much she values *x* and *y*, Jill ranks *x* above *y*' (2012, 34, italics mine). So: a preference is comparative (*x* is ranked above *y*); the comparison is in terms of value; the valuation is subjective ('how much she values ...'); and it takes account of the totality of factors that the individual thinks relevant to the comparison ('everything she takes to bear on ...').

Hausman claims that this definition 'matches most of current practice' in economics. He thus urges economists to reserve 'preference' for this usage (2012, 34–35). Unlike the ordinary usage of the term, which permits some factors that influence choice to compete with preferences in determining choices (e.g. moral commitments or values), Hausman's usage regards these factors as elements influencing preferences. That is to the advantage of economists who want to continue to use the term. As total or overall evaluations, preferences are already informed by reflection on what there is reason to do. They are outcomes of comparative assessments rather than inputs into deliberation. The existence of an all-things-

considered preference ordering presupposes that the economic agent has been able to compare the available options from different points of view or according to different reasons or motivations that she may have had, in order to choose among those options. Models that explain and predict choices in terms of constraints, beliefs, and total comparative evaluations do not deny that human agency and motivation are complex matters. They instead locate those complexities in theories of preference formation and change and not in theories of choice. Only this way of defining preferences allows them to take them as determinants of choice.

But that only applies to deliberations wherein objects of preferences can be unequivocally compared. And therein lies the rub. In such deliberations the agent has some single, well-defined goal or function (a dominant end); or the differing goals which he pursues have some common factor, such as the satisfaction of a known desire. This approach has no use in the—not uncommon—situations in which one needs to combine two or more different goals, virtues or standards, which he feels he cannot ignore or downplay but which seem to demand incompatible things of him. In the standard economic approach, a person experiencing different motivations, who may find it impossible to compare all options, and also be unable to make a clear-cut decision between them, would be considered a case in which the agent's preferences are not fully formed (incomplete). And so he or she is incapable of a rational action. An example of such a case is a choice one has to make between integrity or charity, on the one hand, and some other incompatible goal one might have, say, the pursuit of a suitable retirement, or subsistence, or approval of others (Taylor 1985, 236-7). One can compare the strength of one's desire to have a cup of coffee now with one's desire to have a cup of tea now, and the degree of respective enjoyments and satisfactions. But how can either of those desires and their satisfaction be compared with one's desire to be a fine scholar, a good father, a true friend? Such alternatives are incommensurable, much like vacationing at the beach vs. vacationing in the mountains (Finnis 2011, 424), vs. not vacationing at all and having more money in the bank. Each possibility has some intelligible appeal not found in what makes the other appealing. It is often impossible coherently to combine all the demands that we might consider important and valid. An agent can see one reason as outweighing another only if the two reasons are contained within a single evaluative perspective. This is one reason why not every kind of human choice can be explained in the strict terms of rational choice theory. It is also why humans are considered economic agents only to the extent that their choices can be expressed in those terms. This concept of preference is sometimes referred to as synchronic.

While Hausman's definition of preferences as total comparative evaluations seems to

adequately capture mainstream economic practice, it reveals a problem. By assuming such a strategy, economists are free to ‘explain’ away whatever seems anomalous. People who do X are simply said to have a preference for X.

In fact, it could be argued that the rational ideal embodied in the concept of preferences as total comparative evaluations eliminates human choice in the ordinary sense (Shackle 1979). Said differently: commensurability makes rational choice impossible (Finnis 1990, 236), since the choice is then obvious. Finnis notes that ‘if one option dominates the others (i.e. offers unqualifiedly greater good or lesser evil), the others cease to have rational appeal. However, a morally significant, rationally-motivated choice is between alternative options, each of which has rational appeal. Therefore, where one option is dominant, deliberation ends not in choice but in insight (into the unqualified rational superiority of that option) and action’. Choice informed by practical, and not merely instrumental, reason requires the presence of uncertainty for its very meaning (Buchanan 1999, 257). Virtually the only realistic context in which an option can rationally be regarded as dominant is within the confines of ‘technical’ reasoning and action, e.g. competitive games. ‘Game theory’ is the name of a vast and sophisticated body of reasoning about situations of ordinary life (e.g. ‘bargaining’) conceived *as if* they had the simple, unitary-goal, and typically the self-interested structure of a competitive game, where there is no ambiguity about what counts as winning and losing, and the problem is for the individual player: how to win (or avoid defeat, or minimize one's defeat).

The standard model of choice posits fully articulated comparative evaluations that cannot be taken for granted in the analysis of individual behaviour. It can refer to decisions made by humans only insofar as they can be taken *as if* they were informed by known preferences. The limited scope of application of the ‘as if’ reasoning will be explained a little later. I now turn to the second substantive assumption of rational choice theory: preference stability.

2.2. Homo Economicus and Changing Preferences

Another problematic assumption of rational choice theory is the requirement of internal consistency of choice. This means preferences are to remain stable over time. It is assumed that the agent never changes his fundamental preferences over fully-specified outcomes. At most, the agent may change its derived preferences over actions after learning new information about their likely outcomes. These shifts in preferences are deemed rational as

long as they occur against a stable background of some core, stable preferences.

Stigler and Becker (1977) have famously interpreted the theory such that it condemns any change of preferences as irrational, insofar as preference change introduces inconsistencies into the preference set. This has become the common approach to preferences in neoclassical economics. The exponential discounting model (Samuelson 1937) forbids time-inconsistent behaviours. It also deems irrational all behaviour which reveals inconsistent preferences. For observed choices to be seen as rational, preferences need to be well-behaved, as it were. As much as this assumption looks unrealistic if applied to choice-behaviour of an individual, its meaning becomes clearer if it is remembered that no particular conception of the individual is presupposed in this theory. Is it a human being? A corporation? A koala bear? It is merely a tool for studying behavioural patterns, irrespective of whether these behaviours are conducted by actual persons, groups, or even some narrowly-defined aspects of a person.

Insofar as economic agents are the unit of analysis in economics on the basis of their rationality-as-consistency properties, real persons are not the unit of analysis. Intertemporal consistency of preferences is not a requirement posed to an actual person, or even a generalized definition of a person, but to a 'representative agent'. Just what it is representative of is a matter for the economist to determine. This assumption shows again that, for the economist, when using the representative agent framework to predict demand functions on the market, it is irrelevant whether there are persons that make choices – or even agents, for that matter. The proper objects of economic studies are economic systems, not individual persons. These studies are possible only because and to the extent that systemic selection pressures in markets control idiosyncratic preference drift (Ross 2005).

This approach does not deny that people change their preferences over the course of their lives. What it means is that in the context of an economics based on revealed preference theory, persons whose preferences change in violation of the assumptions of this theory are not economic agents. At best, they can be modelled as sequences of economic agents.

In the dynamic approach to the standard rational choice view, any agent at time t is treated as an independent rational decision maker and, over time, as no more than a collection of transient agents that might happen to have certain common preferences (Gold and Sugden 2006, 191). Neo-Samuelsonians go one step further by decoupling agency from personality altogether. On this view, thoroughly explored by Ross (2005), persons are best modelled as sites of competition between multiple agents (or 'selves'), each endowed with complete preferences. Ross (2005) strongly emphasizes that, far from being 'prototypical economic agents,' persons are rather marketplaces where competing selves bargain over the allocation

of scarce resources.

One of the most commonly-discussed problems of this dynamic choice framework, based on the rational choice approach, is that the account of independent transient agents exhibits a problematic absence of an enduring agent (Bacharach 2006). This is particularly problematic for a sufficient explanation of many choices made by real persons. In this approach it is impossible, for example, to account for long-term commitment of any kind, be it a professional career or personal relationship or some other value.

Following both Dennett (1991), and metaphysicians of morality such as Charles Taylor and Bernard Williams, we might think of a typical human biography as instantiating a general project of creating and sustaining a coherent narrative self. As Rawls (1999, 260) says, we should see our life as one whole, the activities of one rational subject spread out in time. Mere temporal positions or distance from the present is not a reason for favouring one moment over another. Similarly, Buchanan (1999) notes that we move through time constructing ourselves as artefactual persons. We are not and cannot be the same person in any utility maximization sense.

This is not, however, the account of agency that interests most economists. They are concerned with the general basis of agency, on which its idiosyncratic expression in humans can be analysed as a ‘derivative case’. They acknowledge that humans show ubiquitous preference-reversal over time if modelled as economic agents. Only short time-slices in their biographies are (well-)modelled by sets of classical utility functions (Ainslie 1992). As Ross observes, whole people approximate stable agency just to the extent that their

narratively constructed selves inherit relative behavioural stability from the network of community pressures on them to do so. To predict and explain their behaviour well, therefore, one must model them as (internally complex) nodes in complex dynamic systems, rather than as agents (Ross 2005).

The above discussion seems to support Davis’s view of rational choice theory as ‘a free-use tool with no particular connection to the concept of human agency and no particular association with the idea of an individual human being that is supposed to exercise this tool’ (Davis 2015, 76).¹⁹ The complexities of human behaviour are not thought to be the proper object of this exercise. Economists operating within the strictures of rational choice have not expressed much interest in a more realistic account of persons and usually for good reasons—

¹⁹ As Gerhard Debreu so clearly put it: ‘an axiomatised theory has a mathematical form that is completely separated from its economic content’ (Debreu 1986, 1265).

or at least for reasons that are considered good enough for the state-of-the-science. One can still wonder, however, whether economists really do not need criteria for distinguishing human individuals as distinct and independent beings, who do not dissolve into social aggregates and patterns. Is there really no use for economic theory to acknowledge that persons count *because* they are not reducible to economic agents? That their choices also count as their own precisely because the free will cannot be reduced to economic agency? My yet-to-be-argued postulate is that it is one thing to say that human agency cannot be captured by economics' instrumental, synoptic approach, and another to claim that it is not an issue economists need to take into account.

While successful in domains which do not need to take account of complexity of human choice (e.g. basic consumption decisions) this approach has significant costs for the study of the morally constituted person in the economic realm. Its limiting assumptions have deprived the human economic agent of almost everything that could make it a person. It has denied the economic agent the ability to learn about what is good for him and to make value choices. In short, it has neglected his capacity of practical reason. For purposes of analytical rigor and predictive power, it has abstracted this capacity away, as if it were not there. As shown above, the notion of 'preference' fails to capture specific aspects of moral motivation and action. The forthcoming chapters will show, however, that these 'hidden' aspects of choice are important in the analysis of welfare and indeed they should not be overlooked.

That the understanding of economic choice qua human choice does matter for economics has been the major premise of research in behavioural economics. In an attempt to account for some important choice situations which remain outside of the purview of rational choice theory, behavioural economists proposed to expand the notion of homo economicus so that it could be more directly associated with the real human. The next chapter discusses the nature and value of their approach to economic agency.

Chapter 3. Behavioural Account of the Economic Agent

The concept of economic agency summarized in the previous chapter has met with much criticism. Its abstract and axiomatic character has been especially difficult for behavioural economists to accept. They focus their analysis on flesh and blood humans. Scepticism concerning the unrealistic assumptions of rational choice theory was what initially gave rise to behavioural economics. It is, contrarily, concerned with how real humans actually behave.²⁰ Toward that end, it draws largely on laboratory and field experiments, and also on data from individual reports of subjective well-being. ‘Happiness literature’ is where such things are studied. Given the unrealistic economic agent dominant in the standard theory—at least for thinking about human persons as economic agents—it is not surprising that the experimental results of behavioural economics turn out to be at odds with results that are routinely expected based on the assumptions of neoclassical economic agency.

In fact, the biggest contribution of behavioural economics seems to lie in showing the extent to which people’s behaviour departs from the assumptions made about economic agents in the standard approach. It has identified multiple contexts of choice that the standard approach abstracts away, and exposed obvious limits for the application of the rational choice model. The use of the abstract approach may be justified as sufficient for the systematic analysis of aggregate objects such as markets. As shown in the previous chapter, in very limited contexts it can also explain the behaviour of individual agents.²¹ But, as behavioural economics research shows, the generality of the rational choice model has few uses at the level of individual behaviour. That gives behavioural economists reason to enquire into more ‘embedded’ accounts of individual agency.

The insights of behavioural economics have turned out to be particularly challenging for the prescriptive advice generated by the standard approach, of which welfare economics is probably the best example. Exposing the limits of the rational choice theory account of economic agency in this area is perhaps its largest contribution to the discipline. In later chapters I show how this contribution has recently become highly influential in thought about policy and actual policy making.

Since people’s behaviour departs from the abstract assumptions of economic agency,

²⁰ The formal groundwork for behavioural economics was laid by Kahneman and Tversky (1979; 1984) along with various other collaborators who established the existence of a number of anomalies in human decision-making that deviate from the textbook model of rationality used by economists.

²¹ In studying behaviour, economists normally manipulate prices, income, and product selection and then observe or predict how people react – all the time assuming their subjects’ preferences are fixed. But if preferences can change, and change in real time, economists will find themselves without solid ground to begin on, namely, without any basis upon which they can understand changes in behaviour.

behavioural economists have suggested various ways to expand the *homo economicus* concept. The rest of this chapter summarizes and evaluates a few of these suggestions, and outlines some preliminary implications they have for the concept of economic agency and its use in policy.

3.1. Moral Preferences

One of the most important sources of criticism of the standard approach to economic agency is its apparent agnosticism about human moral faculties. Recent findings of behavioural economics research led economists to account for neglected moral characteristics of the economic agent, such as altruism and so-called social preferences, which resulted in interesting expansions of the *homo economicus* concept.²² The proposed alternative notions of the economic agent often suggest that the ‘social’ or ‘moral’ aspects of human nature should be made more explicit in theories, and not merely hidden in the black box of factors influencing preference formation (which economists are not interested in exploring anyway). That personal motives and moral purposes cannot be ignored in the analysis of economic choice has been the claim voiced by many behavioural economists.

While their diagnosis of the problem seems accurate, the proposed solutions are problematic in their selective and narrow approach to what they take to be moral in humans and how they render it in a formal economic language. They remain attached to the idea of preferences as the defining characteristic of economic agents and, contrary to the expressed intention, they do not form a far cry from the notion of the economic man they propose to correct.

Much of the recent theoretical work by behavioural economists, which attempts to incorporate some moral aspects of economic activity, almost exclusively defines morality as either selflessness (e.g. ‘altruism’) or cooperation. They present the latter formally under the description of ‘social’ or ‘other-regarding’ preferences (Bowles and Gintis 2011; Camerer and Fehr 2006; Fehr and Fischbacher 2002; Fehr and Camerer 2007). Sometimes moral behaviour is also identified with rule-following (Vanberg 2006). Whatever these moral features of the human agent are found to consist in, they are usually treated as part of the agent’s preferences which are then incorporated into consumer choice models.²³

²² I discuss this more broadly in Winciewicz-Price (2016).

²³ Vanberg’s conception of preferences over rules is an exception, as is Sen’s notion of ‘metapreferences’. They all, however, continue to speak the language of preferences as the defining framework for the study of human moral nature.

Khalil (1997) offers a useful perspective on some of the early work done in the field, contrasting two ‘extreme’ positions to conceptualizing morality as preferences. One is taken by Becker (1976) who considers economics as the study of all ends, both material and moral, which are subject to scarcity. He embraces the view that one can simply substitute a moral good for another bundle of goods, generating equivalent utility and being psychologically indifferent. At the other end of the spectrum, Etzioni (1988) argues that ‘moral utility’ and ‘pleasure utility’ are not commensurable, and therefore capable of substitution. The limitation of Etzioni's model lies in the implied disconnectedness of moral and pleasure-based utility. By extension we then have both a ‘moral self’ and ‘pleasure self’ in decision-making. The implication of Becker's model is that morals, just like other goods, can be valued in monetary terms, whereas Etzioni's model says moral values are not commensurate with price. An informative overview of some more recent achievements of modelling morality in economics has been provided by Hodgson (2014), and so I will not repeat it needlessly here.

The extension of the scope of preferences by pro-social and other ‘moral’ preferences can be criticized on two important fronts. Firstly, this approach does not challenge the structure of the standard rational choice model in any fundamental way. The economic agent, now with a broader spectrum of preferences, can still be assumed to be a clear-headed maximizer of his clearly-defined preferences. What changes is the scope and character of the preferences themselves. But it can nonetheless be argued that someone with ‘other-regarding’ preferences is still maximizing her own utility, and may in fact be regarded simultaneously as selfish (Hodgson 2014).

McCloskey argues that the reduction of the person to an economic agent with social or moral taste trivializes human moral faculties, making them analogous to a taste for chocolate ice cream:

Most academics and other intellectuals nowadays, without giving it much thought, adhere to the emotivist, chocolate-ice-cream theory. They view the ethical person as maximizing utility function with respect to the doing of good deeds, just as he or she does in the eating of ice cream. No duty, love, faith, or persuasion carried weight (2009, 26).

This approach is very much related to the neoclassical conception of the individual. There the individual is defined as no more than a collection of preferences (Davis 2006, 70). These novel approaches use more sophisticated distinctions, but insofar as these ‘human’ properties discovered by behavioural economists can be subsumed in the rational choice

framework as preferences of a special kind, they do not differ from accounts which explain events ‘as if’ they were brought about by economic rational agents. For this reason it has been doubted whether requiring the standard model to be descriptively accurate in this sense is at all necessary, or merely complicates matters without any added benefit. It appears that the numerous elaborations of the concept of economic man by behavioural economists do not add much to the standard approach, if the latter can explain the ‘moral’ behaviour without the need to modify its major tenets.

Secondly, it can be doubted whether defining moral preferences as matters of how we treat other people is a sufficient account of what makes people moral. While it certainly is a very important aspect of moral conduct, it is not the only one. Compare, say, the ancient tradition of virtue ethics. It claims that moral behaviour is not only about fulfilling one’s duty or obligation to others and following rules. It is just as much a theory of the formation of the one’s will, and a development of one’s desires and aspirations. Even behavioural economists’ extended account of the moral economic agent does not capture this important aspect of human agency.

As a consequence of adhering to the standard preference-based framework, prominent models of what is taken to depict moral agents follow the approach which takes preferences as given and values the norms of coherence between them, but neglects how preferences are arrived at and how they change. Insofar as they operate within the standard utility calculus and continue to speak the language of preferences, they, too, like the standard approach, are blind to an important aspect of the moral nature of the economic agent: the possibility of exercising deliberations and change of heart by the economic agent. Relying on the standard framework, they do not seem to be able to provide sufficient expansion of the notion of apersonal and amoral economic agency to account for what I will later call ‘aspiration’. In this regard, behavioural economists have nothing to offer to the end of ‘re-personalizing’ the economic agent.

I argue instead that what makes persons distinct economic (and moral) agents is not to be found merely in the content or character or solidity of preferences, but rather in their instability and process of formation. Unlike nonhuman agents, humans possess a reflective capacity which enables them to form their preferences and make judgments about their preference orderings, as well as change these preferences as a result of a change in their values and purposes. In less formal language we can say that only human choice is preceded by attempts to sort out: (1) what is worth doing (choice of ends), and (2) what sort of person to be or become (as a distinct end). This shows that moral agency is not to be defined only by its capacity for altruistic or other-regarding sentiments but also by one’s striving for self-

betterment through a conscious formation and change of preferences in line with chosen ends. As such the human agent (the person) is more complicated and less certain about his preferences than the traditionally-conceived economic agent. Such agent can better be portrayed not by more elaborate interpretations of what brings utility, but by accounting for the uncertainty and instability of his preferences.

This implies, however, the need to relax the two above-mentioned economic assumptions: (1) that preferences are given and (2) that they are stable. In fact, behavioural economists have also questioned these assumptions, without, however, taking the step of connecting their conclusions to the moral nature of human agency. Their research in this area could nonetheless be treated as a starting point in the discussion of the place of human agency in economics. The following section details some relevant insights which behavioural economics research contributes to this discussion.

3.2. Accounting for Preference Formation and Change

This section summarizes the criticism of the concept of given preferences based on research in behavioural economics. It introduces the notion of ‘true preferences’, and shows problems involved in their determination.

Again, the assumption of fixed preferences means having a complete and stable rank ordering of what persons want that dictates their choices. The assumption has been criticized from several angles as contrary to the experience of actual persons. That preferences are not the predefined sets of indifference curves, as they are represented in microeconomics textbooks, has been the common finding of psychologists, behavioural economists and, more recently, neuroeconomists. Human agents have preferences that frequently change over time, but standard economic theory of choice says nothing about this phenomenon.

Experimental results, on which behavioural economics relies, provide a great deal of evidence revealing how slippery, ill-defined, changeable, and distorted preferences can be (Camerer and Loewenstein 2003; 2005). Rather than being stable, well-defined, and discernible, preferences have been shown to be constructed (Markman et al. 2000; Slovic 1995); context-dependent (Tversky and Thaler 1990, 210); reversible, and obscure (Payne, Bettman, and Johnson 1992). As Slovic writes, ‘preferences appear to be remarkably labile, sensitive to the way a choice problem is described or “framed” and to the mode of response

used to express the preference' (Slovic 1995, 365).²⁴ The axiom of preference invariance is violated when the same options are presented under different descriptions (Tversky and Kahneman 1986), or when options are presented in different contexts (Shafir, Simonson, and Tversky 1993). Real-world persons often do not, cannot, or will not reach judgements about some things. Furthermore, there is little plausibility to the claim that rationality requires them to do so.

Empirical studies of behavioural economics and cognitive psychology show that people often do not act very 'rationally'. They rarely adhere to 'rationality' in classical rational choice theory with its utility-maximizing orientation on clear, stable, and well-ordered preferences (Kahnemann et al. 1982; Kahnemann and Tversky 2000). According to research in behavioural economics, this results from people suffering, on the one hand, from weaknesses of cognition and, on the other hand, from weakness of will. Not only does the formation of preferences depend strongly on the many factors that do not affect choice according to the rational choice approach (the framing-effect of the problem, the predetermined default rules which prefer certain selection decisions, the way risks are conveyed), but it is also affected by conflicts of interests, values, wishes, desires, etc. Moreover, clear long-term preferences, say, good health, can come into conflict with short-term preferences, such as yielding to temptation and having another cigarette.

If it is assumed that preferences depend on the particulars of each choice problem, then, according to behavioural economists, they are better understood as constructed when the decision calls for them, rather than being 'ready waiting' all the time. They resist the idea that agents 'could look up their preferences in a book, and respond to situations accordingly' (Tversky and Thaler 1990, 209).

Much of this criticism has been rejected by mainstream economists as inapplicable to economic theory, which is not at all interested in studying actual persons. As Ross (2012, 704) observed: 'none of these putative discoveries undermine the standard model of economic agency unless it is supposed that the paradigmatic economic agent is a natural (including socially constructed) person'. The fact that humans fail to demonstrate sufficient stability of goal-specification over time is after all the primary reason why economists are reluctant to associate individuality with economic agency. Even welfare economists who cannot deny that it is persons, not abstract agents, that they want to treat as the proper objects of welfare concern, may find this account of depersonalized agency acceptable. That is acceptable at least insofar as they want to explain choice, measure well-being, and design policy proposals

²⁴ The original sources for many cognitive biases and heuristics are collected in Kahneman, Tversky, and Slovic (1982). Other collections include Kahneman and Tversky (2000), and Gilovich, Griffin, and Kahneman (2002).

with regard to society as a whole or some well-defined, relatively uniform social groups. If, however, policy advice is to be based on the promotion of individual well-being, the advice that comes from assuming that people are well-behaved, group-defined preference satisfiers cannot apply to a world where individuals fail to *reveal* well-behaved and well-defined preferences (whether they have such preferences but do not reveal them is another, epistemological matter). Since real persons fail to behave as predicted by expected utility theory, because their choices reveal incoherent preferences, their behaviour cannot be meaningfully represented by a utility function.²⁵ Welfare economists thus lose evidence concerning that which promotes their welfare, in all but the most basic or general preference sets (water, shelter, food, clothes...). In other words, they do not have an obvious way to determine what people's 'true' preferences are, beyond those dictated by the possession of a carbon-based body.

There are many examples of conflicts among preferences, and of preference change, that lead to behaviour that is very puzzling from the point of view of received economic theory. These also turn out to be particularly difficult for welfare policy. They show that richer and less rigid conceptions of agency are needed to account for the fact that a person may have not just a single set of preferences, but competing sets; and even contradictory preferences (or preference sets). Moreover, meta-preferences which rank the preference sets, and a person's preferences, may oscillate between alternative states. In what follows I discuss two examples of how some of these preference inconsistencies have been captured in informal models of preference change.

3.2.1. Multiple-self models

One attempt to understand and expose the implications of the apparent anomalies of choice, detected by the theoretical and experimental studies of behavioural economics, has been offered in the increasingly popular models of multiple selves. These models address internal conflicts of preferences.²⁶ They are built to capture empirically significant behavioural 'anomalies' or 'failures', leading, for instance, to intertemporal inconsistency. They do so by modelling the person as a community of agents with partially conflicting interests.

²⁵ Hammond (1976) shows that when choices reveal this type of inconsistency between time periods, then the underlying preference structure is not consistent with the standard axioms of consumer theory and the agent cannot be construed as maximizing a well-behaved, intertemporal utility function.

²⁶ Thomas Schelling (1980, 1984), Jon Elster (1979, 1985), Klein (1992), Thaler and Shefrin (1981), Rubinfeld (2001).

Thomas Schelling (1984), one of the most influential authors of this approach, has aimed to show that the phenomenon of rational strategic interaction among alternating preferences is a significant part of most people's decisions and welfare. He argues that it cannot be left out of the economic account of the consumer. He further argues that many important purposive behaviours are ignored if the consumer's preferences and values are treated as fixed and given, and, all the more, given unambiguously. It is important that, in fact, they do change, even over short periods of time.

An example he gives is the smoking behaviour of a person who intends to quit smoking. The vacillations of this state cannot be captured by rational choice theory based on stable preferences. An alternative he offers to account for the dynamic phenomenon is an image of multiple selves taking turns in commanding the agent's behaviour based on their respective values, almost like a multiple personality disorder becoming the standard account of the human will. A smoker may be committed to giving up his addiction for most of the week, but cannot resist temptation to reach for a cigarette at a Saturday party with friends. Schelling holds that we cannot fully understand all of such a person's behaviour if we see her as an integrated individual. We must treat her, rather, as a collection of single-minded individuals (Schelling 1984, 60-61, 84-87, 93, 95-96).

His other famous illustration of the problem is a woman who instructs her obstetrician not to make anaesthetics available during delivery. The woman anticipates asking for an anaesthetic when feeling great pain accompanying labour. But she would rather persist in her determination and deliver the baby without any pain relief. At point-A, she precludes the possibility of following what she believes will eventually be her preference at point-B. A parallel example in a well-known book by Elster (1979) presents Ulysses giving an order to his crew to tie him to the mast so that he would not succumb to the Sirens' song. Both cases show a deliberate constraining of one's future actions due to the recognition that one's preferences at a later time will be different from those prevailing currently, and that those preferences will lead one to do something his or her current preferences do not favour. There are other models that might account for dynamism of preferences and a prophetic warning that allows the changing preferences not to be the ultimate guide of action. I deal with this below in 'aspirational personhood'.

Models of multiple selves treat human choice as analogous to social choice. The individual here is not a unified self, but is understood as having more than one self, or more than one personality, whose interests may conflict. This is not positing multiple selves in a romantic sense of many deep selves, but in a descriptive sense of different, discernible centres of desires. In other words, the different selves have different preferences, which is an

interesting attempt to get inside the black box of preferences employed in traditional economics. Unlike the notion of preferences as total comparative evaluations, this approach drives a wedge between one's self and one's preferences, where preferences are understood as the desires that govern one's actual behaviour. Behavioural inconsistencies across time or across decision frames are illustrated to result from the coexistence in one person of several selves, each endowed with coherent preferences satisfying the expected utility axioms. Depending on the time one makes a decision, or depending on the framing of the decision problem (e.g. the reference point), or depending on the choice menu, all the actual choices of a person will not necessarily reflect the preferences of the same self.

These dilemmas are sometimes referred to as 'weakness of will', incontinence, or 'being in two minds'. In such cases the agent feels like several people and some reflective capacity is required to integrate the competing voices. This problem illustrates a case where the simple model of rational action from given preferences appears to be of little use. The purpose of a utility function or preference ordering is to represent a process by which competing desires and impulses are more or less rationalized. It applies to situations in which any potential conflicts of factors determining preferences has been resolved. Multiple-selves models account for cases where resolution of such conflicts is not an obvious thing. They are a useful illustration of various circumstances in which it is difficult to determine what one's actual preference is and what, if any, reasons should be given to prioritize one preference over another. This is a crucial problem for economics and economic policy concerned with people's welfare, which operates on the principle of preference satisfaction.

Since welfare economists take persons' revealed preferences as evidence of their well-being, the relevant question here is whether there is a genuine, or legitimate, or representative, or right-thinking self that can be identified by examining the conflicting preferences of the person. Characteristically, Schelling does not reach any conclusions. Rather, his method is to describe a multitude of cases, each offering a different challenge to possible ways of defining this self. The multiple-selves literature is inconsistent on the question of whether one self is somehow superior to the other(s).

Sometimes, however, there is no implication of superiority, only conflict (Cowen 1991). In other cases, the analysis explicitly favours one self over other, as in Thaler and Shefrin's (1981) model, which pits a far-sighted 'planning' self against many myopic 'doer' selves; the former self is self-evidently superior. But, again, it depends on the specific case. And much of it is bound up in knowledge gained from experience. Sometimes we feel as though the planning and doer selves and any other selves are equal(ly) (our)selves; other times we feel like one self is somehow superior to the other(s). This could be because, say, the

planning self is more foresighted or more capable of adopting a seemingly disinterested perspective. Being less swayed by circumstances, and more general in its function, it seems like a stable self. (But again, *who* is choosing the real self? Is that another self, and is it a doer or planning self?)

Cowen (1991) makes a related point when he observes that the rule-oriented self is not necessarily any more ‘rational’ than the impulsive self. Cowen adopts this terminology to distinguish his approach from the more typical long-run self versus short-run self. There is no *prima facie* reason for believing that victory for the rule-oriented self is desirable. Many cases, such as the abuse of alcohol or hard drugs, may require the victory of the rule-oriented self for personal welfare, but too many victories for the rule-oriented self can be injurious to mental health. A person who continually thwarts the impulsive desires of his other self may become frustrated and overly rigid and lose his capacity for spontaneity (Cowen 1991, 365). As possible instances of individuals who err on the side of too little spontaneity, Cowen offers tightwads, workaholics, and compulsive exercisers (1991, 367).

By definition, each self in the multiple-selves framework is rational in the sense of utility theory since a person is divided into selves such as to make sure that each self has consistent preferences. It follows that the weighing of selves cannot depend on whether selves satisfy the axioms of utility theory. Rather, it will depend on the content of the selves’ preferences. The perspective of weighing two or more options in search for the best outcome is of course reminiscent of the standard instrumental calculus. Appealing to standard rational choice theory based on the calculus of expected utility does not enable us to select which of these portrayals of instrumental reason is correct. The theory has usually been interpreted as working from preferences that one has at a particular time, and ignoring any possible preference changes that one expects to occur in the future.

It appears that resolution of an internal conflict between multiple selves stands in need of something other than preference orderings. To the extent that the multiple-selves approach merely reifies impulses within a person as separate persons, it adds nothing useful to our perspective. Internally conflicted individuals need something which stands in judgement over ordinary preference orderings and thus enables the privileging of one self over the other. It is often suggested in the literature that self-command can help resolve such conflicts. For Schelling (1980) self-command plays in such cases a role of a management skill. It is doubtful, as I explain later, that self-command can be treated so simply (and just which self is doing the commanding?). It cannot just be a device which helps satisfy preferences better: it must stand above and outside the normal idea of preference. Without a more substantial theory of what is the preferred self – the ‘real self’, as it were – ‘self-command’ cannot

answer the question of which preferences are to be served by it. In Hargreaves Heap's words: 'what is required is a reason for giving weaponry to one side rather than the other' (1989, 107). Such a reason will not be found in preferences alone because both selves are equally placed on this score. Since each self has a set of preference orderings, and this is all that defines an agent, if there is nothing more to a person than such a set of orderings and agency, then there is no reason to choose between the two selves.

Some authors propose that 'the weighing of selves' requires a substantive theory of rational preferences²⁷ or a good-defining theory.²⁸ Intuition or common sense may help here, but the point is that some theory of the good is ultimately required. Scientific results may also give indications regarding what is the right thing to do. For instance, the fact that the regular consumption of fat food increases the risk of cardiovascular fatality indicates that positive and normative considerations are deeply entangled.

Sometimes it might be obvious what the actual or true preference is (or should be). Sometimes it is only the person in question who can and should answer this question for himself. Sometimes still, the question of which of the identified preferences is good-defining is the wrong question to ask. What is often forgotten in the models of multiple selves is that the identified selves are not, in fact, different selves with different preference sets, but aspects of a single self, or better, of one person. To characterize any one of these selves as the 'true' self and thus the 'true' source of preferences is an error if the beginning assumption of that each of them stands in at times for the true self. As the next chapter shows, if the true self can be identified at all, it is not to be found in any one of these separately personified desires and personalized preferences, but rather in the interaction among them – an 'emergent self', or as I call it an 'aspirational self'.

The example of the multiple-selves framework shows that the standard theory stands in need of a broader perspective to account for preference conflicts and ways to resolve them. In Chapter 5, I suggest an alternative interpretation of such conflicts. Meanwhile I turn to my second example of an informal account of preference instability.

²⁷ Defining instrumental reason must involve taking a stand on certain issues about the nature of the good of a rational person. But, according to the instrumentalist thesis, reason should be defined independently of the good. Jean Hampton argues, however, that there are substantive constraints on the good associated with instrumental reason that enter at the point of specifying the appropriate sources of good-defining preferences: 'The instrumental reasoner, in order to specify his conception of reasoning, has to say what a good-defining preference is, and doing so means that he must take a stand on the sources of the set of good-defining preferences' (Hampton 1998, 188).

²⁸ 'What counts as a good-defining preference is not self-defining. Determination of such a preference requires a theory that picks out such a thing, and this theory must take a stand on the nature of the good associated with rational action. Instrumentalist rational choice theory is certainly insufficient for this task because it aims to be neutral among alternative positions regarding what a preference is, what the sources of good-defining preferences in a situation are, how to cope with preference change, whether or not it is rational to generate preference change, and so forth' (Hampton 1998, 203).

3.2.2. *Sour grapes*

In the book *Sour Grapes* (1983), Elster explores another interesting context of preference change in his analysis of the phenomenon of preference adaptation. He addresses the problem of preference formation influenced by the conditions of one's action. This is a phenomenon not accounted for in the standard rational choice theory, which does not enquire into how preferences are formed.

An adaptive preference formation occurs when the agent changes a preference for something in response to the content of his feasible set. A change in constraints leads to a change in preferences, which is a violation of the axiom of context-independence of the standard model. Elster illustrates the phenomenon with the story borrowed from La Fontaine's *Fable of the Fox*. The fox discovers some sweet ripe grapes which he wants very much, but since he cannot reach them, he says to himself that they are sour and therefore not worth having. The story shows the fox's preference change between attempting to grasp the grapes and walking away, from desire to indifference.

Adapting like this to the available options is a psychological mechanism to reduce cognitive dissonance, that is, to reduce the frustration caused by situations where what one most wants is not something one is likely to get in the circumstances that prevail. The fox's frustration is reduced by his unconscious downgrading of the preference for self-creation unavailable grapes—it could just as well have been conscious, but not according to Elster—so that what he now wants most ends up being merely something he can get, strawberries. This causal adjustment of desires to the possibilities, shaped by non-conscious drives, Elster calls 'sour grapes'.

Elster contrasts adaptive preference formation with what he takes to be an alternative mechanism for dissonance reduction, character planning, as advocated by Stoic, Buddhist, or Spinozistic philosophies (1983,110,117). Character planning refers to the adaptation of preferences to what is possible, as a result of one's own conscious decision concerning which preferences to have. It is a process of 'engineering' of 'conscious strategies of liberation' (1983, 117). Rather than subconsciously convincing himself that the inaccessible grapes are really not all that good after all, the fox exercising character planning would reform his desires by deliberately upgrading the value he sees in the accessible strawberries, in order to get just as much satisfaction from them as he would have from the grapes.

Elster's 'distinction between the causally induced and the intentionally engineered

adaptation of preferences to possibilities' (1983,110) is particularly interesting for the argument advanced in this thesis since it poses a useful challenge to the assumption of stable and complete preferences. Elster finds them 'too remote from the real world'.

Elster's explicit twofold aim in his analysis is to extend the purely formal conception of rationality to allow a scrutiny of the substantive nature of desires and beliefs (1983, 15), and thus to address the question of rationality of preference formation. He distinguishes rational preferences from irrational ones based on how they came about. Preferences causally induced by non-conscious drives, which cannot be construed as part of some plan for intertemporal utility maximization, are rendered irrational. Think of the fox's preference initially to walk away from the unreachable grapes. Elster characterizes preferences which are formed in this way as 'objectionable' (1983, 22). For the processes through which they are formed he counts among what he calls 'disreputable psychic mechanisms' (1983, 129). By analogy, rational preferences should not have been forced upon the person from without, but should instead result from one's active, conscious and deliberate planning to change one's preferences by means of determining goals, strategies, and choices, so that they become adapted to the possibilities that one has. Accordingly, a central presupposition for actions to be rational in a wide, substantive sense is that the desires or preferences which lie behind them are *autonomous*. I will return to the relation of preferences to autonomy in a later chapter.

Elster's point in making this distinction is to show that the notion of adaptation of preferences to possibilities is 'crucially ambiguous' (1983,117), because it may be read either as something which is causal (irrational) or as something which is intentional (rational). This ambiguity, in turn, creates problems for welfare evaluation based on revealed preferences. Regarding the example of the fox, Elster notices that for the utilitarian, there would be no welfare loss if the fox were excluded from consumption of the grapes, since he thought them sour anyway. His new preferences are in accord with the options open to him, and he has other sources of sustenance, at least in the strawberries. But since his new preference resulted from his conviction that he would be excluded from consuming the grapes, it would be difficult to determine his true preferences by invoking those he revealed in his decision to walk away.

The problem this radical context dependence poses for ordinal welfarism is often illustrated with the example of preference for city life versus preference for life in the country. If, while in the city, I would prefer city life to life in the country, but when in the country, I would prefer the country to the city, it is indeterminate which state of affairs is ultimately preferable and so which should be brought about. Adaptation renders ordinalist evaluation

incomplete. It is therefore impossible to establish, on the basis of the agent's ranking of preferences alone, which state of affairs is preferred overall. Put another way, adaptation of this sort (sour-grapes, irrational) can make the desire for satisfaction a misleading metric of welfare. The problem is analogous to the search for 'real' or 'true' preferences in the above discussion of multiple selves.

Elster's discussion of the problem of adaptive preferences is an interesting illustration of why it can be useful for economists to study the details of preference formation. It seems, however, that—much like the multiple-selves framework—it does not offer an unambiguous method to decide when preferences should be seen as correct or true, or welfare-enhancing. The reason for this lies in the problematic conception of character planning.

The problem with Elster's definition of rational preference adaptation being possible only as a result of character planning is that it is based on goal or outcome-oriented reasoning typical of a standard rational choice theory approach. Character planning is defined in exclusively instrumental terms. Realizing one's purposes is, of course, part of being rational, but exactly in the thin sense of rationality which Elster wants to extend. Still, it does not seem rational to approach every situation with the attitude that one has certain pre-defined goals that one tries to realize. Imagine approaching a marriage that way.

Genuine change and development, where one does not control the process and does not know where it will lead, is not uncommon. And yet, somehow that very way of being in the world—the way sane, adult humans almost everywhere act—seems to be regarded as incompatible with being rational, as expressing that one is controlled by some sort of alien force working 'behind one's back.' On Elster's account, character planning seems to mean that one should explicitly set up a known outcome in advance as a goal for oneself, and then devise means, in the form of self-manipulation and manipulation of others, in order to reach this goal. While the agent's first-order preferences change in response to a change of one's options, it is assumed that his higher-order preference that is guiding the process of conscious adaptation remains given and stable.

That is basically the approach of the standard theory, *mutatis mutandis* for Elster's purposes, the applicability of which to real-life choice situations is very limited. Elster himself admits that 'conscious character planning is a relatively rare phenomenon' (Elster 1983, 132). The conclusion seems to be that in most cases the adaptation of preferences to possibilities will be an irrational phenomenon, rendering satisfaction of such induced preferences a misleading metric of welfare. The problem is in a basic assumption: character planning takes a person to be a clear-headed maximizer over clearly-defined preferences. Sandven (1999a) argues that it makes for a very static conception of the personality, since it

does not allow for a person to be regarded as rational, who does not have full control over change or development of his or her preferences, namely, who does not control the process and does not know where it leads.

What Elster is recommending as an evaluation criterion of preference formation brings us to an unrealistic picture of a person, analogous to and sharing in some limitations of the standard rational choice approach. In particular it cannot account for action open to learning and receptive to new experience, action aimed at growth and development, action that involves readiness to be changed. Elster's attempt to go beyond the limitations of instrumental rationality does not succeed in addressing this challenge. While intending to go beyond the unworldly assumption of given and stable preferences, he arrives at an account that takes this assumption to a higher level of preferences. That is then a view entirely compatible with the standard rational choice approach (cf. Stigler and Becker 1977). In fact, the very language of character planning has instrumental connotations ('engineering' of conscious 'strategies').

It seems, however, that an alternative interpretation of preference change through character planning is possible, better described perhaps as 'character formation'. It gives a more realistic account of human agency, by better capturing the not uncommon fact of human choice, namely that is often not a result of known and well-understood preferences, but rather a reflection of preferences being formed through choice.

As I shall show below in the chapters that follow, adaptation to change in one's feasible set could also be interpreted as a process of self-discovery or self-creation. An obstacle on the way to fulfilment of one's initial preference can bring about a change of preferences as a result of change in one's value criteria. Such a change could not be interpreted as sour grapes since the agent is fully aware of his new values and actively responds to them. It could not, however, be rendered character planning in Elster's sense, for it does not constitute a strategic upgrading of existing options in line with a given higher-order preference. Rather, it entails an adjustment of one's values (i.e. change of the higher-order preference structure). Such a process requires the agent to be open to form new goals and new reasons to pursue them, thus enabling him to become a new person in substantial and morally-relevant ways. Instead of strategically manipulating his beliefs, such an agent finds new meaning in the options that he has. This alternative interpretation allows for the possibility of genuine learning, which is not obvious in Elster's character planning. Real character formation consists in one becoming a different person with new values, evaluations, and preferences that one did not have before—it entails a character-transformation. I shall discuss details of this approach after specifying some basic characteristics of this account of human agency in the next chapter.

Both the multiple-selves and sour-grapes approaches show that preference instability cannot be ignored in the analysis of human choice. And both provide clear examples of situations in which preferences cannot be taken as given. They expose the limitations of the rational choice instruments in accounting for preference conflict and adaptation, and in determination of true preferences. This, in turn, challenges the welfarist doctrine which identifies welfare with the satisfaction of known, stable, and beneficial preferences. It seems however that they are mistaken in trying to arrive at some theory which would help to establish objectively which of the conflicting preferences are correct, or what it means for preferences to be formed rationally. In fact, the problems involved in identifying the agent with a particular well-defined (static) aspect of his preferences or meta-preferences proves that consistency of behaviour over time is not only hard to achieve but hard to define (Winston 1980; Shefrin and Thaler 1981; Schelling 1980). I argue that this approach misses an important element of what makes us human, and in doing so they become less effective at their ostensible theoretical goals. Both frameworks discussed above could instead be interpreted as illustrations of self-discovery and self-experience, or a process of preference learning, albeit of different kind than that which the standard rational choice theory allows. Contrary to the instrumental means-ends approach of the latter, this process calls for the exercise of practical reason by the person in question. The view of human agency I want to offer for the consideration of economists is one that is rooted in the Aristotelian-Thomistic tradition, important parts of which, as I will try to show, are shared by Adam Smith. It is a tradition marked by virtue ethics and development of practical wisdom. This tradition's anthropological claims about human action expose that which is missing in the neoclassical account of the human economic agent.

Chapter 4. Towards the Concept of an Economic Person: Insights from Premodern Economics

Many important normative concepts associated with human choice and decision making were left outside of economic analysis as the discipline emancipated itself from a broader discourse of moral science. In this sense, it was only with the arrival of the modern world that it was possible for modern economics to be born. Premodern economics, embedded in a normative framework, saw man's economic pursuits in the broader context of his moral decisions. Langholm suggests that the assumptions of human conduct used in modern economics originate in the ideas about human rationality and reason developed in the Enlightenment period (Langholm 1992, 1-10) and she is not isolated in this view. On the other hand, however, Hollis notes that 'there was another voice in the Enlightenment, a voice that argued that the power of reason allows humans to master nature, manipulate society, change culture and, indeed, shape our own selves' (Hollis 1977, 12).

The capacity for shaping our own selves is a distinctly human feature, and an important one, too, as the thorough investigations of it in classical philosophy and especially in the virtue ethics tradition show. Virtue ethics is the oldest and broadest stream of ethical theory in the study of human conduct. Introduced by Plato and Aristotle, it was taken up by the Stoics, used by Cicero, and finally adopted into Christianity in its most lasting form by Thomas Aquinas.²⁹ In this tradition, right action is defined as that which leads to the 'well-being' of the individual. The term '*eudaimonia*' is a classical Greek word, commonly translated as 'happiness', but perhaps better described as 'well-being' or 'human flourishing' or 'good life'. *Eudaimonia* characterizes the well-lived life. At the centre of the eudaimonistic ethics is the view of the person as a being conscious of oneself and one's own will, concerned not only with the question of what one should do but also of what one should and could desire. In the Aristotelian-Thomistic tradition, moral behaviour is not only about fulfilling one's duty or obligation to others and following rules. It is just as much a theory of the formation of the will, and a development of desires and aspirations (Koslowski 2006). One does not need to accept Aristotle's theory of the will or even regard the will itself as a meaningful concept (there is a powerful modern trend to treat it as obsolete) to recognize some truth in this observation.

A similar idea appears to be present in Adam Smith's *Theory of Moral Sentiments*, where he shows that people derive self-worth from actions that are agreeable to the impartial

²⁹ McCloskey (2009, 44) notes that until the late eighteenth century, virtue ethics was seen by 'most Europeans as the only sensible way to think about good and bad character'.

spectator. It is rarely noticed that Adam Smith saw that human well-being is to be derived not only from acquisition of material goods, as *The Wealth of Nations* suggests, but also from pursuit of goals that are hard to conceive of as material, such as moral self-betterment.

Political economy, like other modern social sciences, arose to a considerable extent in the process of emancipating itself from these traditional moral teachings, around the time when virtue ethics was losing its prominence in the academic theories of philosophers. In the aftermath of the modern dismissal of metaphysics, discourse on moral philosophy began to exclude the problem of identifying the *telos* or ‘end’ of man and the notion that the virtues are a necessary means of helping him reach that end (Gordley 1991).

The approach of the premodern thinkers differs significantly from present-day attempts to introduce morality into economics. I propose that the premodern thinkers who did not have a separate notion of an economic agent or even ‘economic man’, but analysed all human decisions and actions in broader ethical terms, can offer important insights into what is involved in economic choice qua human choice. They show that the former cannot be completely and cleanly separated from the latter without some meaningful loss. Aristotle, Thomas Aquinas, and Adam Smith are three particularly important representatives of this tradition. What connects the three philosophers is an explicit account of human nature characterized by man’s striving for the good and self-formation (‘man’ was then used, but especially in Aquinas’s case, ‘human being’ is what was meant). Their insights offer a potentially valuable addition to what has been only partially recognized by economic utility theory, decision theory, and other rationality studies. They provide building blocks for my concept of the economic agent as an aspiring person whose preferences are not fixed and stable but forming and changing. The concept of aspiration, in turn, highlights the dynamic aspect of human agency and suggests that it is not merely satisfaction of preferences but the very process of their formation which makes human agency different from non-human agents and constitutes part of people’s welfare.

Pre-modern virtue ethics tradition and its account of the place of economic activity in a broader context of human moral life can provide some valuable insight into how various elements constitutive of the human could be better integrated. Promoting economics as a moral science today cannot, however, be reduced to unreflective borrowing from the thought of Aristotle, Aquinas or Smith. Just as their work was produced in specific historical contexts, the relevant elements of their scholarship need to be meaningfully incorporated into the very different context in which contemporary economic theories are made with a careful understanding of the differences between the old and the new approaches. One needs to be cautious not to over-interpret historical texts. One way to assess the credibility of an attempt

to inform contemporary questions by historical ideas could be found in its capacity to disclose and illuminate ardent problems in contemporary study. My use of the ideas borrowed from pre-modern economic thought should find justification in its capacity to identify and provide possible solutions to the problems related to obsolescence of the understanding of human economic agents as moral agents in contemporary economics. Insofar as these ‘old’ ideas provide building blocks for the concept of the aspiring agent, they are instrumental in exposing the deficiencies of homo economicus in accounting for human choice and provide insights into possible correctives of this incomplete image of human economic agent.

4.1. Self-Formation and Virtue Acquisition in Aristotle

I begin with Aristotle and his image of the human being as an agent which emerges from the economically relevant parts of his ethics and politics. For, it cannot be said that Aristotle was an economist, but that an economics can be construed in his broader philosophy. Aristotle’s anthropology stresses the importance of man’s capacity and propensity to develop and practise virtue, and accounts for a person’s responsibility for acquiring a good character. The maintenance and governance, the ‘rule’ of the household – literally, *oikos* + *nomos*, which became ‘eco-nomics’ – by human action stands central in his thought. Society is made of multiple *oikoi*, eventually forming a city (*polis*). The city is the only place where humans can reliably flourish. For Aristotle, a particular practical ethics is the basis of politics. The habits of hearth and home, husbandmanship and farming, and so on, form the basis of the human good that flourishes collectively in the city. Aristotle’s economic thought is also formed in the same practical ethics.

Aristotle makes the exercise of agency the logical basis for ethics and links it tightly to human distinctiveness. For Aristotle, an agent is roughly defined as a being that can entertain, and take steps to fulfil, ‘reasonable desires’. Since reason is taken by Aristotle to be the defining property of the fully human, it follows for him both that only human beings (and perhaps divine spirits) can be agents, and that all morally competent human beings must be agents. For him, the rationality that is the distinguishing mark of agency is neither merely formal nor only instrumental. Aristotle also takes certain kinds of goals as necessary for agency, but they are not as well-defined as rational choice theory requires goals (read: well-formed preferences) to be. The goal of truth-seeking is one such example of a general goal.

4.1.1. *Oikos & phronesis*

The great influence of Aristotle not only on ethics, but also on what later came to be called economic thought, makes his works a natural starting point of enquiry into the moral nature of an economic agent.³⁰ Aristotle is the first thinker to analyse economic life systematically, so that it is seen as ‘embedded’ in the extended fabric of the community. For Aristotle, economic affairs do not have a free-standing status but properly belong to practical philosophy. His ‘economic analysis’, which is usually identified in Book V of the *Nicomachean Ethics*, and Book I of the *Politics* is only a sub-section within an enquiry into other, more essential subject-matters (Finley 1970, 5). The most important criterion of evaluation of these various matters appears to be the opportunity for each agent to realize his own potential and thus to fulfil himself in a flourishing life.

What are today referred to as economic pursuits, though important, are for Aristotle simply one among other means toward human flourishing and excellence. One’s endeavours related to securing some level of material status have a functional, not finalistic, nature and it is a necessary, but not sufficient, instrument for attaining a good life. It is true that a good life cannot do without the possession of material goods. Nonetheless, it is not exhausted by the material component, but depends, rather, on a plurality of human dimensions. Present in Aristotle but missing in modern perception of the economic agent is the motivational complexity of every human choice and thus of economic choice too.

It is also in this perspective that contemporary commentators who look for an equivalent of the ‘economic man’ in Aristotle associate him with a person engaged in management of a household (*oikos*). As part of Aristotle’s broader philosophical system, economics is based on prudence or ‘practical wisdom’, and economy is a prudential art or craft. Prudence is the crucial characteristic of the ‘household manager’. Happiness (*eudaimonia*) or ‘health and well-being’ of all those who individually and collectively people the household are the formal objectives of wealth acquisition (*chrematistics*). Since the two central attributes of *eudaimonia* identified by Aristotle are that it is *teleion* (final, perfect, or complete) and *autarkes* (self-sufficient, lacking in nothing), in pursuing a good life one strives for self-fulfilment and self-sufficiency.

When reading Aristotle’s reflections on what we call today economic or market behaviour, one cannot resist the impression that economics for him is mostly about human

³⁰ The Scholastics used the *Nicomachean Ethics* as one of the leading textbooks. Some important contributions of such influential modern thinkers as Marx and Keynes were derived from Aristotle’s works.

values, concerned more generally about *eudaimonia* (Pol I, 9, 1257b, 40-1258a, 2)³¹. Unlike modern economics, he treats of economic phenomena such as production, consumption, exchange, in the context of virtues. He also deals with a broader notion of human rationality than the instrumental approach of modern rational choice theory. His is the rationality understood as prudence (*phronesis*).

The '*phronimos*' in Aristotle is the person who has *phronesis*, practical wisdom (as compared to theoretical wisdom), full reasonableness (as opposed to perfect rationality). This includes the ability of discerning what things are good for himself and for mankind at the point of choice, rather than theoretically or 'from nowhere'. Such a person is the norm of action, both economic and non-economic (NE II.6, 1107a1; VI.11, 1143b15; VI.5, 1140b8-10).

Virtue, however, is not something we are born with but something we acquire. In book II he points out that we possess neither the virtues themselves, nor the capacities for them, by nature. The Aristotelian perspective implies 'a shared though differentiated human nature', namely, 'a shared human tendency to become virtuous and not a virtuous human nature' of an essentialist kind (van Staveren 2001, 10).

4.1.2. Eudaimonia & autarkes

According to Aristotle, *eudaimonia* is constituted neither by honour, wealth nor power, but by 'rational activity in accordance with virtue over a complete life' (NE). This might be described today as productive self-actualization. This process of self-completion and all that it entails reveals the human person to be a potential and not a fully actual being. This perfection cannot be actualized outside of a moral context, for to act contrary to the good amounts to non-fulfilment. Fulfilment is only reached through morally good acts (Koslowski 2006). Indeed, in Aristotle (and, as will be shown below, also in Aquinas) a person is conceived through his actions as *en route* toward some completion (Schenck 1976). In a way, the whole *Nicomachean Ethics* is Aristotle's reply to the question of how we grow up to become the fully adult and self-sufficient rational animal that is the end toward which the nature of our species tends. That it was written to his son, as the title says, would lend credence to this interpretation.

Why is self-sufficiency (gr. *autarkes*) important for a good life? Because in order to

³¹ According to Campbell (1987), the ancient Greeks seem to have thought only in terms of ethics, for ethics was the only source of reason for public decision-making.

live well, according to Aristotle, one must possess a certain independence from the exigencies of material existence and from others; for instance, by not being a slave or permanently under parental control. And the household is the institutional realm in which those exigencies are dealt with through the art of household management. The proximate end of household management, that is, of economics, is therefore a degree of independence from the demands of material existence. That proximate end, in turn, subserves the final end of household management, living well. Individual households are found not to be sufficient for this end, and thus the village and eventually the city are formed to meet it.

What self-sufficient agents are not immune to, however, is contingency. *Eudaimonia* is not to be understood as an ideal which we can only approximate but never attain. Self-sufficiency with respect to humans, whose pursuit of the good is fragile, is always imperfect. For Aristotle, it is possible to know the good, and fail to act on it because something intervenes between the general knowledge of the good and the knowledge of the particulars involved in pursuing the good. *Eudaimonia* can therefore be considered a state of relative, and not absolute self-sufficiency (Patrone 2005, 41).

A resolution of the practical problem of pursuing the good life cannot be achieved solely with a general account of human nature. An account of the nature of particular individuals and their contexts is also required. The constituents of *eudaimonia* vary for those who are incapable or ill-situated to pursue the ideal of contemplation. Self-sufficiency must be interpreted with respect to the particular facts about us. Self-sufficiency is not freedom from contingency because freedom from contingency is not humanly achievable.

4.1.3. Deliberation & choice

Aristotle's account of deliberation indicates that the human choice situation cannot always be treated as forming a closed, complete, consistent system. Indeed, a person's reflection on a new situation that confronts her might disrupt an order that has previously existed. It could bring a change in her evolving conception of one or many points of living or acting.

Aristotle's account of human choice and deliberation indicates the unfinished or indeterminate character of our ideals and value structure, which are constitutive both of human freedom and practical rationality itself.

Given the above-mentioned space for particularities and contingencies of human choice permitted in the Aristotelian framework, Aristotle would not entirely oppose the subjectivity of one's views regarding realization of one's potential and fulfilling oneself in a

flourishing life. Yet, according to Aristotle, subjective notions of the good life are not enough to meet the goal of living well. People may differ in their particular vision of what it means to flourish, but all those visions are accurate only insofar as they are consistent with the life of virtue. The life of virtue consists in intellectual and moral virtue, i.e. in the ability to discern what is best and the ability to act on this judgment. In other words, moral virtue serves to bring our actual choices and preferences into line with what our better judgment recommends to us (Veatch 1962, 101).

We find in Aristotle's account of human choice and deliberation, contrary to the standard economic approach as developed in rational choice theory, that deliberation about man's ends by definition does not benefit from the means-end paradigm (Wiggins 1980). For Aristotle, it is the mark of the man of practical wisdom to be able to 'deliberate well about what is good and expedient for himself, not in some particular respect (health, strength) but about what sorts of things conduce to the good life in general' (NE, Book VI). Deliberation in Aristotle is not primarily a search for means, but rather a search for the best specification of the good that one wants. A nontechnical deliberation is one in which man has a vague description of something he wants—a good life, a satisfying career, a fulfilling hobby, etc. The problem is not to see what will be causally efficacious in bringing this about, but rather to see what really qualifies as an adequate and practically realizable specification of what would satisfy this want. As long as this specification is not available, there is no room for means. This observation is not given enough consideration in the mainstream economics account of human choice.

While the comparatively trivial technical problems (means-end cases) may often be resolved by recourse to rules, this is not the case with deliberative situations. "Matters concerned with conduct and questions of what is good for us have no fixity (...) The agents themselves must in each case consider what is appropriate to the occasion" (NE II, 2, 1103b 26).

Wisdom in action can never be reduced to general rules that apply smoothly across circumstances because of the uncertain and idiosyncratic nature of human action. Further, wise decision making does not assume that we already have a fixed and absolute set of preferences or values. Instead, one of the purposes of wise decision making is to discover what we prefer or what is really of value. For it is only through wise and morally imaginative deliberation that we gain clarity about those objects we have previously deemed to be of value; it is only through envisaging our life as transformed by different possible courses of action meant to achieve that object that we come to more vividly understand the features and consequences related to that object.

Given this specific nature of practical knowledge in general, particular cases are even more difficult to determine, since they do not fall under any rule or precept, but are decided by the agents themselves in each case whenever they reflect on what is appropriate to the occasion. From the nature of the case the subject matter of the practical is indefinite and unforeseeable, and any supposed principle would have an indefinite number of exceptions.

For Aristotle, deliberation is the identifying characteristic of the person possessing practical wisdom, and it is action (not contemplation as it is sometimes mistakenly thought), that makes a person virtuous: ‘... we become just by doing just acts, temperate by doing temperate acts, brave by doing brave acts’ (NE VI, 5; II, 1). This linkage between virtue and action is relevant to my aim to represent the economic agent as a person forming and developing herself and her values through action. Aristotle explains it thoroughly in his theory of virtue acquisition which is one crucial building block I use later to form the concept of an aspiring agent.

4.1.4. Habituation & virtue

Aristotle is well known for asserting that virtue is habituation of a certain kind. He does not simply say that we inherit or acquire virtue from our parents. In book II of NE he points out that we possess neither the virtues themselves, nor the capacities for them, by nature. He claims that one becomes virtuous by doing virtuous actions. Thus, Aristotle’s good man develops over time (Burnyeat 1980, 69). This development takes place in a sequence of stages with both cognitive and emotional dimensions. Since the noble and the just do not admit of neat formulation in rules or traditional precepts (1.3.1094b14-16, 2.2.1104a3-10, 5.10.1137b13-32, 9.2.1165a12-14), it takes an educated perception, a capacity going beyond the application of general rules, to tell what is required for the practice of the virtues in specific circumstances (2.9.1109b23, 4.5.1126b2-4).

The development of correct ideas about which actions are noble and just requires an upbringing in good habits. Habituation for Aristotle is therefore a way of learning which he puts on a par with, though differentiates from, induction, perception, and other modes of information acquisition (e.g. intellectual intuition and experience). We first learn what is noble and just ‘by learning to do noble and just things, by being habituated to noble and just conduct’. In this sense practice has cognitive powers. The ultimate goal towards which the beginner’s practice is aimed is that he should become the sort of person who does virtuous things in full knowledge of what he is doing, choosing to do them for their own sake, and

acting out of a settled state of character—that is, from habit.

Thanks to good formation, ‘by doing the things you are told are noble and just, you will discover that what you have been told is true (Burnyeat 1980, 74). You will become able to make judgments of your own. To learn to do what is virtuous is also to learn to enjoy doing it, to come to take the appropriate pleasure in doing it. What the virtuous man enjoys is the practice of virtue undertaken for its own sake. Practice leads to knowledge. To understand and appreciate the value that makes certain action enjoyable in itself, I must learn for myself to enjoy it, and that takes time and practice. In short, it takes habituation. In so far as I do not realize this, and I desire and perhaps do the opposite of such things, I feel ashamed of this failure.

In the ordinary meaning, the word habituation refers to a process in which the parent or teacher *acts upon* the child or pupil. To habituate someone is to do something *to* them, and to be habituated is to undergo some experience of training at the hands of another. Aristotle is very clear that the process as he conceives it does not mirror this linguistic sense. For he says in many places that it consists of actions (*praxis*) performed by the person acquiring virtue:

...we acquire the excellences through having first engaged in the activities, as is also the case with the various sorts of expert knowledge—for the way we learn the things we should do, knowing how to do them, is by doing them. For example, people become builders by building, and cithara-players by playing the cithara; so too, then, we become just by doing just things, moderate by doing moderate things, and courageous by doing courageous things. (1103a31-b2)

Aristotle is cautioning his reader against the overly passive conception of virtue-acquisition, which he feels attends an intellectualization of virtue. His point is not to stress the importance of parents in shaping or even creating a child’s character. His point is that we become the people we make ourselves. Even sick people, Aristotle points out, need to take their well-being into their own hands at some point. They cannot expect just to be made healthy by the work of the doctor. The same holds for those of us sick in soul: we should not think that our own justice is anyone else’s doing. This process of becoming is encapsulated in the distinction that Aristotle makes between *doing virtuous actions* and *being a virtuous person*.

In so far as a person is learning to be virtuous, he is performing a quasi-virtuous action, an aspirational imitation of what the fully virtuous man is doing. It is a kind of action of which only another person—the fully virtuous person—is a model. In this framework

(which I will use later to suggest an alternative interpretation of preference change) a temporally prior self can be normatively posterior. The agency of the earlier self exemplifies that special form of change that is *learning*.

When the person who is only learning to be virtuous does not perform virtuous actions for the sake of those actions, that is not because he does them for the sake of any base motive, but because he does them *in order to acquire virtue* (without yet knowing what he is to acquire, since that knowledge only comes through doing it rightly). The ‘trainee’ does not act from a fixed disposition to justice not only because he fails to be just, but because his disposition is, more specifically, in flux. He has a different kind of character from the vicious (or virtuous) person: an unfixed, changing character. His disposition is being formed in his learning of the virtue. He uses his dim awareness of value to come into contact with the value which, he can then see at the end of the process, has guided his pursuits all along.

4.1.5. Human action & moral responsibility

In the *Ethics* Aristotle searches for the distinguishing characteristics of human actions for which we hold the agent morally responsible. He finds it to lie in the fact that a man is the source and originator of his actions as he is of his children (NE 3.5):

The natural bodies, as opposed to things with souls, have a source not of causing movement or of acting but of being acted on. It is absurd to blame external objects rather than oneself as being too easily caught by such attractions, and to take the credit for one’s good behaviour but blame pleasant objects for one’s bad behaviour (NE 3.1. 1110b9-15).

External objects are not in themselves sufficient causes for the voluntary movements.

That the origin of human actions is in the agent means as much as to say that people are moved to act by what appears desirable to them; what appears desirable to them depends on their character; and their character in turn depends on their actions, and is therefore in their power. Each of us desires what appears good but each of us is also responsible for his disposition toward goods: what one finds to be good depends on one’s character. The responsibility for man’s action lies in his character and cannot be shifted to an external object of desire. The well-brought-up and well-(self-)formed person avoids unjust actions for insofar as they are unjust or ignoble, they do not appear to him as pleasant or enjoyable. Good

upbringing makes the noble a part of the pleasant for us.

Aristotle's lectures are designed to take their reader to the next step and make the noble a part, perhaps the chief part, of one's conception of the good. That is why he makes it a condition of virtue that virtuous actions be chosen for their own sake. Choice which is reached by deliberation from a conception of the good includes a desire for them as good in themselves as well as noble and pleasant.

The (fully) virtuous man would, then, be the one who performs virtuous actions (he returns what he borrowed, rushes into battle, and refuses pleasures) only when they are in fact right and good things to do. He will also *not* repay a debt, *not* rush into battle and *not* forswear pleasures, when *those* inactions are the right thing to do. In order for one's action to be (in any sense) just, one must do it for the right reasons. When someone justly returns what he owes, he does so not in order to avoid a more painful outcome (much less from evil ends) but from some the recognition that it is the right thing to do. If the learner's *ethismos* (habituation) is to be a moral education, he must be learning precisely to answer the question, 'When is it right to charge forward?'

Aristotle is not blind to the fact that the process he is describing has a dependent character, relying on external support. He makes reference to the need for a teacher whose job is, presumably, to guide, assist and correct the learner: 'If it were not like this, there would be no need at all of anyone to teach them, and instead everyone would just become a good builder or a bad one.' (1103b12-13). It seems he can hardly learn this by repeated blind charges (or nudges?). Of course, parents and teachers play a significant supporting role³², but it is nonetheless of the utmost importance to Aristotle that the agent is in control of his own virtue-acquisition:

But if without being ignorant a man does the things which will make him unjust, he will be unjust voluntarily. Yet it does not follow that if he wishes he will cease to be unjust and will be just. For neither does the man who is ill become well on those terms—although he may, perhaps, be ill voluntarily, through living incontinently and disobeying his doctors. In that case it was then open to him not to be ill, but not now, when he has thrown away his chance, just as when you have let a stone go it is too late to recover it; but yet it was in your power to throw it, since the moving principle was in you. So, too, to the unjust and to the self-indulgent man it was open at the beginning not to become men of this kind, and so they are such voluntarily; but now that they

³² Although Aristotle seems to think the most important supporting work is done by the laws and the lawgiver (see NE X, 9).

have become so it is not possible for them not to be so. (1143a12-a21)

Aristotle holds that what makes you just is what you (voluntarily) do, not anything that is done to you. Habituation consists in work done by the habituee.

4.2. Aquinas and the Scholastics: The Economic Realm as Space for Personal Development

Influenced by Aristotle, the scholars of medieval universities in Europe known as the Scholastics adopted a similar approach to the study of human action. The aim of this section is to present major principles of their system of thought and the place of human economic dealings in it. As in the previous section, the goal here is to focus on the conception of human nature and human conduct and the way they served both as the background and the starting point of the scholastic analysis of economic endeavours. Scholastic anthropology claims that persons are beings capable of moral (self-)rule, whose sense of duty is ‘an important but not the only determinant of their conduct’ (De-Juan, Monsalve 2006; Decock 2013, Langholm 1998). The moral dimension of human behaviour served as a basis of what was later called scholastic economics.

4.2.1. Scholastic certainties

The name ‘Scholastics’ usually designates a number of thinkers who lived and wrote between the eleventh and mid-seventeenth centuries in Western Europe. They were mostly theologians and jurists who, among many other topics, treated of the question of the morality of various social interactions, including economic exchanges. Often, they were treating these questions in the broader discussion of salvation. For instance, can a good Christian engage in market activity for the sake of self-enrichment, with the knowledge that Jesus said the love of money is the root of all evil. Their moral and philosophical approach to economic problems is a most distinctive feature of their scholarly efforts.

Scholastic economic thought has received some fairly detailed attention by a few recent scholars, primarily Langholm, but also Dempsey (1948), Schumpeter (1954), Grice-Hutchison (1993), and Lapidus (1994). Schumpeter famously traces the early evolution of economics as a field of study from its origins in scholastic theology. According to Langholm

(2003), *The Guide for Confessors* became the first text of economic analysis. For, priests would use its generalized stories about the good and the bad to solve cases of conscience, to determine objectively whether one was in a state of mortal sin because of one's immoral (sometimes economic) activity. According to Schumpeter, it was the 'Scholastic doctors' of the Middle Ages who deserve the title of founder of economics:

It is within their systems of moral theology and law that economics gained definite if not separate existence, and it is they who come nearer than does any other group to having been the 'founders' of scientific economics. And not only that: it will appear, even, that the bases they laid for a serviceable and well-integrated body of analytic tools and propositions were sounder than was much subsequent work, in the sense that a considerable part of the economics of the later nineteenth century might have been developed from those bases more quickly and with less trouble than it actually cost to develop it, and that some of the subsequent work was therefore in the nature of a time and labour-consuming detour (Schumpeter 1954, 93).

Just how important their achievements were, is illustrated in Schumpeter's juxtaposition of the (today mostly neglected) Scholastics with (the still widely recognized) Adam Smith. According to Schumpeter, 'the Scholastic doctors fashioned all the basic analytical tools that Smith found at hand when he wrote the *Wealth of Nations*'. In Schumpeter's view 'the scholastics and the natural-law philosophers had worked out all the elements of' economic analysis, while Smith simply undertook 'the task of co-ordinating them' (Schumpeter 1954, 178).

Among these Scholastics it was particularly Thomas Aquinas who, in Schumpeter's view, contributed most to establishing the grounds of modern scientific analysis (1954, 8).

While it is true that the Scholastics made important contribution to formal economics, the focal point of their analysis in this area is not technical economic analysis but moral philosophy and theology. In all the well-known matters discussed by the Scholastics (usury, just price, market conditions, market price, monopoly, labour and wages) one can trace a distinctive 'scholastic paradigm' (Langholm 1998, 57) in which every aspect of economic life was viewed from the perspective of man as a morally responsible agent and child of God. It was through this sort of study of moral philosophy that Scholastic economics emerged (Langholm 1979, 11-36). When questions of an economic nature emerged, the Scholastics saw them as a subordinate part of the broader theological and moral concerns (Gordon 1975, 157).

Like Aristotle, the Scholastics did not approach economic problems in the same way as we do today. Their terminology was also different from the abstract language of modern economics. As in Aristotle, their concepts associated with economic activity did not constitute a separate field of enquiry but rather formed an appendix to ethics and law. Most medieval writers who ventured into the discussion of economics saw it as a form of behaviour which like all others must be considered within the discussion of human beings as of God's creation and, as such, morally or law-bound actors. Although economics was not yet acknowledged as an independent discipline, it formed a consistent body of doctrine according to which economic relations ought to be ruled, by the laws of distributive and commutative justice, always finally by the 'law of charity'.

The anthropology underlying the scholastic paradigm in economics is not that of the homo economicus whose goal is the satisfaction of given preferences. Their famous doctrines of usury or just price are ultimately grounded in the vision of a morally responsible man, responsible to one another because ultimately responsible to God (De-Juan, Monsalve 2006). The idea that the just price is not just an outcome of a whimsical decision of the parties involved or of social convention highlights the embeddedness of scholastic contract law, for instance, in a specific conception of man. Man's will is thought to be autonomous, capable of creating its own laws, or at least ruling itself. That is one way to speak of 'free will'. But within the hierarchy of laws, the will remains subject to a plethora of heteronomous laws. Those laws include positive human laws, divine law, and natural law (Decock 2013, 523). This vision of human nature implies the possibility to act contrary to right reason and to positive law.

The frame of reference for the analysis of later Scholastics was 'a moral universe that obliged any buyer or seller to act for the common good and agree to terms of exchange accordingly, regardless of the advantage granted him by the forces of the market' (Langholm 1998, 85). The economic realm is just one among other spheres of human activity where learning of virtuous conduct takes place.

The *moral* universe of the Scholastics is uniquely Christian, with charity at the top of the virtues, seeking to expound and justify divine law (found in the Bible) and Christian standard of morals. The *ethics* of Scholasticism, however, is not to be found in the Hebrew or Christian Scriptures. Its broad framework is Aristotelian. As we learn from the *Catholic Encyclopedia*, it rests on the conviction that the supreme good of man is happiness, that happiness is the realization, or complete actualization, of one's nature, and that virtue is an essential means to that end'. What the Scholastics added to this Aristotelian system was the replacement of Aristotle's vague notion of the meaning of happiness (with its occasional

inclusion of particularities of ancient Greek habits and prejudices), with life according to the Divine purpose in creation and the dignity which is man's birthright as a child of God, made in the divine image (*imago dei*).

As Christian theologians, they held the position that man's ultimate goal is happiness or *beatitudo* which consists in the vision of God. One could only have such a vision if one's vista were cleared of other distractions and idols, that is, false images of god, as well as false forms of worldly happiness (*felicitas*, pleasure, honour-seeking).

Making this vision an explicit goal for every person does not change the basic similarities between the Scholastic and the Aristotelian thought. Aristotelian thought was subsumed within the broader vision of ultimate happiness as one of the means – necessary but not sufficient – which is shown in the four cardinal virtues (justice, prudence, courage, and moderation, which are partially derived from Aristotle, but also Cicero et al.) being always lower in importance than the theological virtues (faith, hope and love/charity). For the Scholastics, as for Aristotle, human happiness consists in living a distinctively human life. This life realizes, so far as possible, one's potential as a human being. It is a life unlike that of other animals because a human being can be caused to act, not only by appetite or habit, but by reason and will. Man can understand that an action contributes to the distinctively human life she should live or would like to live, and she can choose it for that reason. The moral virtues, namely, those virtues which pertain to action, discipline the senses and train the desires according to practical reason. They set limits on what is moral based in divine revelation and natural law. They attempt to ensure that human actions are oriented not only toward other persons, but equally if not at times more importantly toward oneself. Both Aristotle and the Christian scriptures base care for others on care of oneself.³³

4.2.2. The moral end of man

The moral dimension of human conduct which formed the basis for six centuries of scholastic doctrines was most thoroughly studied by their best-known representative, Thomas Aquinas. In his magnum opus, *Summa Theologiae* (Aquinas 1947: II, I, 1945: vol. II), he presents a

³³ Cf. Matthew 7:12 and Luke 6:31: 'Do unto others as you would have them do unto you', and the Aristotelian virtue of liberality, with prodigality and meanness as the excess and deficiency. Liberality is about the right use of wealth, which involves gift-giving. If one gives too much away, it cannot be a virtue, for he is left unable to be self-sufficient. He has then lost not only the ability to give more but even the ability to abide without taking too much from others. To do justice to others (Aristotle) or to do charity for others (Jesus) each require that you have sufficiently taken care of your own self. I return to this below in the discussion of self-interest rightly understood (NE, Book IV).

basic account of the human person as a moral agent with divine provenance.

Summa Theologiae sketches a broad view of Thomas's understanding of human being in which he further develops Aristotle's insights concerning moral agency. In Thomas's vision, as in Aristotle, human action is teleological or purposive, in that man's goal is to achieve the virtuous life. The intellectual and moral virtues perfect the human intellect and appetite in proportion to human nature (ST Ia IIae., q. 62, art. 2). Unlike Aristotle, the ultimate goal is set beyond human nature in sharing in the divine nature which was described above as the beatific vision.

In Aquinas's formulation, the human person is naturally ordered toward the true and the good through the operation of the intellect (reason) and the will, even as sin has corrupted this nature. When functioning properly, these two faculties interpenetrate and cooperate with each other in both seeking the true and achieving the good. Whether man seeks good or ill, there is a general orientation toward perceived goods. He does not choose what he knows to be fully bad:

Whatever man desires, he desires it under the aspect of good. And if he desire it, not as his perfect good, which is the last end, he must, of necessity, desire it as tending to the perfect good, because the beginning of anything is always ordained to its completion; as is clearly the case in effects both of nature and of art. Wherefore every beginning of perfection is ordained to complete perfection which is achieved through the last end. (ST I-II, q.1)

Being human means that we can be very much mistaken in our judgements and decisions. Thus, Aquinas's approach, similarly to Aristotle's, focuses on man's imperfect knowledge of the self and his fundamental uncertainty with reference to what he wants and what he should value. For Aquinas, the *telos* of persons is grounded in self-consciousness and self-mastery. Man always strives for increasingly more complete self-fulfilment: 'all desire the fulfilment of their perfection, and it is precisely this fulfilment in which the last end consists, as stated above' (ST I-II, q.1, art. 5).

He is not born perfect (neither in the sense of complete or serene) but comes increasingly closer to perfection through his rational and free action. Such action aims to realize his human potential at its maximum scope; it is a means: 'since everything desires its own perfection, a man desires for his ultimate end, that which he desires as his perfect and crowning good'. Hence Augustine, whom Aquinas relies on as a central authority, earlier said (De Civ. Dei xix, 1): 'In speaking of the end of good we mean now, not that it passes

away so as to be no more, but that it is perfected so as to be complete.’ It is therefore necessary for the last end so to fill man’s appetite, that nothing is left besides it for man to desire. Which is not possible, if something else be required for his perfection.

As human beings, we are governed by both mind and body, and individually, we have a responsibility to be true to ourselves and to strive to become the best that we can be in this our earthly existence. The assumption that everyone has a goal in life—or many goals more or less connected to an ultimate goal—is so basic to our nature that we pass most of our days unaware of its importance to our well-being and happiness. But a goal, doing something because we believe it is intrinsically good in itself and desired by us as such, gives us a perspective and forces us to reflect on what is important and meaningful to us. And, although we may not think in terms of an ultimate goal in a theological sense, nevertheless, every step we take to reach our human potential will ultimately lead us to completeness.

In short, Thomas thought that the *eudaimonia* of this world must be a perfection of the person (Koslowski 2006). His focus was on the actualization of persons, not the actualization of lawfulness or rationality (Lemmons 2011).

4.2.3. ‘Prudentia’ and moral action

In order to become a good person, one needs to acquire the virtue of prudence, which is covered in Aquinas’s notion of ‘*prudentia*’ (ST II-II q.47 aa 1-7). Prudence enables us to reason well towards the choice of commitments, projects, and actions, and to apply general practical principles to concrete circumstances in order to choose rightly. Prudence involves both the choice of the end sought and the means to attain that end. It calls for reason rather than impulse, and takes counsel from others in the selection process of ends and then means to achieve the chosen end (Elmendorf 1892, 4). In the *Summa*, *prudentia* is described as superior to theoretical knowledge, for prudence perfects the cognitive faculty:

The Philosopher [Aquinas’s title for Aristotle] says (Ethic. i, 3) that ‘everyone judges well of what he knows,’ so that judgment would seem to belong to the cognitive faculty. Now the cognitive faculty is perfected by prudence. Therefore judgment belongs to prudence rather than to justice (ST II-II, q.58, art.4).

Prudence provides the moral agent with the directive principle of virtuous actions. It combines diligence and care. As in Aristotle, it is acquired through training and experience,

and is developed into a stable habit by repeated use. It is not surprising, then, that the virtuous person according to Aristotle resembles Aquinas's conception of the person characterized by *prudentia*. For the former, 'the virtuous person performs the right action in the right way at the right time on the right objects' (Rorty 1980, 380), and 'really knows what one is doing, being aware of the circumstances and consequences of one's actions, with the right conception of the sort of action one is performing'. For the latter prudence means 'rectitude of reason in things to be done' (ST I-II, q. 61, art. 3 and 4). In both accounts the source of action is the human himself. According to Aquinas's theory, an action is voluntary if it is due to an internal principle. The internal principle by which a human being acts, and which makes his action that of a human being, is 'reason and will', taken together:

Now man differs from irrational animals in this, that he is master of his actions. Wherefore those actions alone are properly called human, of which man is master. Now man is master of his actions through his reason and will; whence, too, the free-will is defined as 'the faculty and will of reason.' Therefore, those actions are properly called human which proceed from a deliberate will. And if any other actions are found in man, they can be called actions 'of a man,' but not properly 'human' actions, since they are not proper to man as man. (ST I-II, q.1).

Prudence, like all virtues, is acquired by voluntary action, by taking into account the consequences of one's actions into account, and by doing so over and over until it becomes habitual. Of course, not all people are able to become similarly prudent even if they try just as hard to be. A negligent person especially fails to exercise prudence. The Late Scholastics—those practising in and around Salamaca, a few generations before the modern age—concluded that a person should be liable for failing to exercise the degree of prudence that he could be expected to possess. That is reminiscent of Aristotle's view that man is responsible for the character he develops, since he is responsible for his desires: 'that good is most complete which the man with well-disposed affections desires for his last end.'

Among numerous similarities between Aristotle and Aquinas, there is an important difference in the conception of the human in both thinkers. Aristotle's claim that the human being is a product of nature became an impediment to the development of an understanding of certain human activities, such as immaterial conceptual cognition or reflective cognition.

It was Aquinas who made a revolution in the understanding of real beings (a certain kind of 'first substance', in Aristotle's terms), among them also human beings. Aquinas appealed to the experience of the existence of the subject called "I". The human being

experiences that it itself is one and the same subject who sensually feels and intellectually cognizes. The human being is a person who experiences its own personal way of existing. It is through this personal experience and personal activities that a person constitutes oneself as the determined source of action and strives for one's own full development (Krapiec 1993). The process of man's self-constitution is enabled by voluntary action, as Aquinas says:

it is proper to the rational nature to tend to an end, as directing [agens] and leading itself to the end: whereas it is proper to the irrational nature to tend to an end, as directed or led by another, whether it apprehend the end, as do irrational animals, or do not apprehend it, as is the case of those things which are altogether void of knowledge (ST I-II, q.1, art.2).

Persons *act* as persons by acting through themselves. That might sound like a riddle, but it simply means that their action is radically their own, rather than anyone or anything else's. My acting as a person is not an undergoing, or an enduring, or a transmitting of what originates outside of myself. It is I, I myself who act when I act as a person, and no one else. Aquinas expresses it well: 'persons are not acted on but act through themselves.'³⁴ Aquinas does not mean that persons can never or should never be acted upon, but only that being acted upon does not reveal them as persons. A person strives for a goal that is hers independently of other forces, and if a creature is unable to do so, it is not properly a person:

When a man of himself acts for an end, he knows the end: but when he is directed or led by another, for instance, when he acts at another's command, or when he is moved under another's compulsion, it is not necessary that he should know the end. And it is thus with irrational creatures (ST I-II, q.1, art.2).

Beings that can do this are persons, meaning moral agents, and capable of the economic activity that is a subset of this sort of intentional activity. Creatures that cannot do this necessarily exist outside of personal moral responsibility. As non-moral creatures, they are also non-economical creatures.

According to Mueller (2010), 'Aquinas's synthesis contains the first complete statement in history of what is involved in any human economic action' (Mueller 2010, 17). I will demonstrate that by its uses in economic thought. As the next section will show,

³⁴ ST I, q.29, art. 1. Cf. Crosby (1996, 26).

Aquinas's vision of human action was indeed, if unconsciously, taken up by the author of the most famous modern account of man's economic dealings. It was worked out five hundred years later by the father of modern economics, Adam Smith.

4.3 Adam Smith: Moral Maturation, Responsibility, and Praiseworthiness.

Adam Smith may not be the most obvious thinker to be associated with the Aristotelian-Thomistic tradition. Yet, his thought is very much in line with the concept of the human agent as a moral person developed by the two great philosophers. The sources of his moral enquiry, especially, go back to the ancient tradition of the Greeks and the Romans (Szulczewski 2015, 91).³⁵ Smith's essential premise, which he shared with the ancient thinkers, was that moral science could only proceed from a thorough understanding of human nature. Morality is anthropologically anchored. Therefore, the moral philosophical realm analysed in his work was humankind (Evensky 2007, 5), wherein human agency was the root of his economic understanding (Oakley 1994, ix). His understanding of man was much broader than the one that is often mistakenly ascribed to him today. Assuming his economics is anthropologically anchored, if one misunderstands his anthropology, it is no wonder that his economics is also routinely misread.

This section discusses Smith's conception of the human agent with a special emphasis on his treatment of human desire for self-improvement. Smith's views about the economic agent have often been misunderstood, distorted, and abused – particularly by those who read him least or least well. The way Smith is presented to most economic audiences today is entirely oblivious to his moral concerns. His ideas seem fully in line with modern positivistic economic analyses (Billet 1976, 296), with a footnote sometimes appearing that he also wrote on moral philosophy. The very fact of Smith's treating of the subjects of the wealth of nations on the one hand, and human moral sentiments on the other, in two separate works, has been interpreted as signifying a radical departure of the Smithian approach from the scholastic paradigm in which economic dealings and moral responsibility constituted a more integrated subject of enquiry. But Smith was a systematic thinker of a different kind than his predecessors.³⁶ Thus, the first claim has been made linking his economics and moral philosophy: that he is a systematic thinker. Young argues that 'the structure of Smith's

³⁵ McCloskey (2009) argues that Smith, in emphasizing a system of multiple virtues, was 'a half-conscious follower of Plato and Aristotle and therefore of Aquinas, and therefore the last of the ancient virtue ethicists'.

³⁶ Alvey 1997.

research programme gives a strong presumption, not only of consistency, but also of the relevance of his moral theory for understanding his political economy' (Young 1997, 24).

Many philosophers and economic historians also argue that Smith's economics cannot be separated from his moral thought.³⁷ Smith believed that many behavioural regularities can be explained better by understanding of people's attitude toward actions, rather than their valuation of final outcomes. He notes that many of our choices are 'not so much founded upon [their] utility,' but reflect primarily 'the great, the noble, and the exalted property' of the action or activity itself. Related to this, he thought that we should never investigate human nature as such, in a void, namely merely theoretically. Even if such an investigation were possible, it would prove disappointing, for it could teach us little about the origins of men's attitudes. He, in a likewise manner to Aristotle and Aquinas, puts the practical above the theoretical in most things.

When seen in a more modern perspective, Smith marks the culmination of the emphasis on the motives and actions of the individual in early British political economy. He did not accept the static view of man as coming into the world already fully equipped to make correct moral decisions. To Smith, such a conception of man neglects the everyday process by which moral decisions are actually arrived at. His *Theory of Moral Sentiments* (*TMS*, henceforth), in particular, shows man's devotion to self-improvement. It accounts for the crucial process to which all people are subject, that 'of moral and psychological development from the early stages of childhood to that of the mature moral agent' (Brown 1994, 95; Montes 2008, 45).³⁸ Indeed, Adam Smith's *TMS* is devoted to the question of how people form moral sentiments. The Smithian individual agent forms his moral sentiments through an informal learning process that is realized in social interactions (Thoron 2016). Smith makes an important connection between material and moral well-being. In *The Wealth of Nations* (*WN*, henceforth), he depicts outside interferences on people's economic pursuits as assaults on man's soul.

The apparent inconsistency of *TMS* with *WN* gave rise to the famous 'Adam Smith

³⁷ Small, for example, claims that, contrary to present-day view of economics as a moral-free science, Smith held the view that 'economic processes are incidents of larger moral processes' (Small 1907, 201). Billet reminds us that Smith, not unlike the older Greek thinkers, assumed that the 'connecting principles' of human social relations and the 'happiness and perfection' of mankind are essentially moral in character, that is to say, resting on ideas about 'right' and 'wrong' (Billet 1976, 298). In his later book, which was more focused on the economic aspects of human pursuits, he saw all economic phenomena originate in choices, decisions and actions of human agents. This is why Viner admonishes Smith scholars to 'resist the ahistorical trap of abstracting Smith's economics and social theory from his moral philosophy and theology' (Hill 2001, 22). For a recent account which makes an interesting link between 'the Wealth of Nations Smith' and 'the Theory of Moral Sentiments Smith' see Herzog (2013).

³⁸ Otteson claims that the development of each person's moral awareness and virtue from infancy to adulthood is among Smith's major contributions in *TMS* (2002, 9).

Problem'. This is so, even as a consensus has been reached that this was a false problem based on faulty exegesis, and that there is no formal contradiction between the two works. Nevertheless, the Adam Smith Problem is still simplistically stated: since *TMS* focuses on 'other-centeredness expressed chiefly in terms of generosity, benevolence, and sympathy' and *WN* focuses on self-centeredness expressed in terms of net personal advantage (Etzioni 1988, Brown 1991), there is an unbridgeable divide.³⁹ They might as well have been written by two different, and opposed, authors.

WN should not, however, be seen as a shift from moral philosophy based on the 'principle of empathy' to economic science based on selfish behaviour of individuals, and thus alien to moral considerations (Young 1992, Levine 1998, Walsh 2000). Smith posits in *TMS* and again in *WN* a common human 'desire of bettering our condition' (*WN* II.iii.28; *TMS* I.iii.2.i); and implies, though he nowhere states it explicitly, a broad notion of 'betterment', including 'a moral kind that involves improving our character and has nothing to do with acquiring material goods' (Fleischacker 2004, 63; see also Macfie and Raphael 1984, 9; Heyne 2008, 59-63; Griswold 1999, 130-136; and Otteson 2002, 196-197).

Smith made it clear in *WN* that wealth lay in the well-being of the people. What is rarely noticed is that his notion of well-being covered not only people's material prosperity but also their moral welfare. In *TMS* he focused primarily on man's moral perfection, and in *WN* his major concern was with the betterment of his material condition in such areas as productive powers of labour, manufactures, the net revenue of the society, capital, mechanics, and even 'dexterity of the workman', which 'necessarily increases the quantity of the work he can perform'. The latter, however, is not treated in entire separation from the former, but rather completes it. Smith conceives of man as a divided creature who has outwardly-directed passions that want simply to be satisfied, but also has higher-order passions about what sort of person he would like to be. The desire for self-improvement, both material and moral, which Smith takes to be natural for human beings, seems incompatible with the posited behaviour of the creature called rational economic man. The latter is he who systematically pursues his self-interest along the lines of the rational choice theory (Winch 1992, 105).

I will focus here on the main components of Smith's rich theory that are particularly relevant for the purpose of this thesis.

³⁹ See (Brown 1994, 165–66) and footnotes for a discussion of such an 'overly economic' reading of *WN*; and (Evensky 2005), chap. 10, on the 'Chicago Smith' versus the 'Kirkaldy Smith'.

4.3.1. *Self-interest, rightly understood*

According to Morrow, the crux of the question that Smith wanted to answer was ‘the place of self-interest in human excellence’ (Morrow 1923, 7). Unlike many of his contemporaries, Smith did not consider self-interest as an unequivocally negative passion, which should always be curbed by other, selfless sentiments. Like his ancient and medieval predecessors, he saw virtue in one’s being justly benevolent toward oneself, and not merely (or primarily) toward other people. He stresses that an ‘imperfect creature’ like man, whose existence requires so many things external to himself, must often act from many other motives ‘than pure benevolence’ (TMS VII, ii, 89). Self-regarding behaviour (as distinct from selfish behaviour), though less desirable than benevolent behaviour, is therefore not vicious, as was held by advocates of the ‘benevolent system’ like Hutcheson. It can be virtuous as long as such conduct does not injure others. As pointed out by Young (1985,120), Smith criticizes those systems of moral philosophy in which virtue derives solely from benevolence concluding that:

Regard to our own private happiness and interest, too, appears upon many occasions very laudable principles of action. The habits of economy, industry, discretion, attention, and application of thought are generally supposed to be cultivated from self-interested motives, and at the same time are apprehended to be very praiseworthy qualities, which deserve the esteem and approbation of everybody. (TMS VII, ii, 87)

For Smith, being wholly altruistic, and disregarding one’s own needs, about which one is usually best informed, is equal to making the same mistake as being wholly selfish, disregarding other people’s needs.⁴⁰ He saw virtue in one’s self-improvement, not only in material but also, moral aspect. He says, ‘upon many occasions...

[what] prompts us to the practice of those divine virtues is not the love of our neighbour, it is not the love of mankind (...). It is a stronger love, a more powerful affection, which generally takes place upon such occasions; the love of what is honourable and noble, of the grandeur, and dignity, and *superiority of our own characters* [italics added].

⁴⁰ McCloskey (2009) notes that ‘the Earl of Shaftesbury in the sixth edition of a work he had first published in 1713, took an evolutionary line to arrive at praise for such prudence: the affection toward private or self-good, however selfish it may be esteemed, is in reality not only consistent with public good but in some measure contributing to it’.

Moreover, in *TMS* he notes that acting on the basis of self-interest is one of the few passions that could be taken for granted: ‘We are not ready to suspect any person of being defective in selfishness, this is by no means the weak side of human nature, or the failing of which we are apt to be suspicious’ (TMS VII, ii, 87). That self-interest is natural he puts rather graphically in the following passage: ‘Every man is, no doubt, by nature, first and principally recommended to his own care; and as he is fitter to take care of himself than of any other person, it is fit and right that it should be so.’ (TMS II, ii, 11). Again, moral reasoning would begin in the particular and practical, for that anthropological fact about who is best fit to care for oneself.

Smith prompts us to endeavour to give the level of regard to our own concerns that we give to the concerns of others. Yet, since our perceptions of our interests are frequently faulty and our behaviour, even in economic settings, is capable of being blown off course by other motives such as ‘love of dominance’ and ‘love of ease’, to use Smith’s vocabulary, it becomes all the more important to be conscious of what the proper scope of self-love should be. If properly restrained, self-interest is a positive virtue (Winch 1992, 106).⁴¹ This is summarized nicely by Smith in an extended passage:

Though it may be true, therefore, that every individual in his own breast, naturally prefers himself to all mankind, yet he dares not look mankind in the face, and avow that he acts according to this principle. He feels that in this preference they can never go along with him, and that how natural so ever it may be to him, it must always appear excessive and extravagant to them. When he views himself in the light in which he is conscious that others will view him, he sees that to them he is but one of the multitude in no respect better than any other in it. If he would act so as that the impartial spectator may enter into the principles of his conduct, which is what of all things he has the greatest desire to do, he must, upon this, as upon all other occasions, humble the arrogance of his self-love, and bring it down to something which other men can go along with (TMS II, ii, 11).

⁴¹ Smith does not neglect to note, however, that this desire for honours and preferments must be balanced by an appropriate measure of the sentiment of justice (Evensky 2005, 117). Smith’s notion of justice deserves a separate study all to its own. For the purpose at hand, it is enough to mention that, as we learn from *TMS*, the laws of justice in Smith’s theory arise from the general rules of conduct with which the well-informed, impartial spectator approves. Consequently, pursuing self-interest to the point of the spectator’s disapprobation is precisely the same thing as violating the laws of justice (Young 1985, 121). An example of just economic conduct in *WN* is individual pursuit of self-interest under competitive conditions, when all the rules of fair play and strict justice are being observed. A counterexample of just practices is collective pursuit of self-interest through combination, monopoly, and extra-parliamentary pressure-group activity.

Indeed, to have an appropriate measure of self-love inspiring one to better one's condition is not only proper but even admirable. We respect the person who '[i]n the race for wealth, and honours, and preferments (...) run[s] as hard as he can, and strain[s] every nerve and every muscle, in order to outstrip all his competitors' (TMS II, ii, 11). It follows from this that lack of proper self-regard 'would undoubtedly be a failing' (TMS VII, ii, 87).

Smith shows that self-interest can be a virtue consisting in the combination of a concern for the self and for the reaction of the affected other, together with the capacity of self-command that would enable the concerns to be effected. At the same time, while self-love is a necessary condition for the unleashing of humankind's productive energy and creativity, it is not sufficient (Evensky 2005, 110). Its proper scope is to be determined by sympathy.

Sympathy, in turn, is considered a fundamental doctrine in which Smith develops a highly original theory of the moral consciousness (Wood 1984, 177). Man's ability to sympathize with fellow beings enables him not only to seek approval of others but also of his own self. It enables him to make and pass judgments. As the following sections will show, Smith's moral agent is someone who has developed the capacity for deliberation and judgment, and learned to view his actions with the gaze of a disinterested spectator. Moral judgment in Smith's theory appears then to be a prudential judgment. However, it is not prudential in the limited sense of, say, frugality but rather in the fullest sense of Aristotle's *phronesis* or Aquinas's *prudentia*. It is not so much the calculation of potential gains, as the desire to do proper things, to do things that would meet the approval of a disinterested bystander, the impartial spectator (Choi 1990, 293); God, even. Choi (1990, 294) posits, in a Smithian vein, that it is the awareness of the presence of others like oneself and moral judgments made with this awareness that make a person a human being. We return then to the anthropological anchor of Smith's ethics, which also has economic consequences.

The role of the impartial spectator in the formation of one's self-reflective judgment is briefly discussed in the next section.

4.3.2. The impartial spectator

The notion of the impartial spectator—namely, 'reason, principle, conscience, the inhabitant of the breast, the man within, the great judge and arbiter of our conduct' (TMS III.3.5, 137)—was, in Smith's own estimation, his main contribution to ethical theory. The concept of the

impartial spectator is introduced in the third part of *TMS*. Although Smith does not give any unified definition of the impartial spectator, we do learn that the spectator is a hypothetical observer who passes positive or negative judgments upon the actions of those around him, directed by virtuous considerations, whether of the intellectual or some other kind. All men place themselves in such a position to judge strangers, and later in the same impartial spectator's role to judge their friends, family, and finally their own actions (Lamb 1974, 675).

In the latter case, the impartial spectator is an imaginary construction that individuals call upon as they evaluate their own sentiments and conduct. It is in many respects a stand-in for the judgments of our behaviour that others in society would render if they were unencumbered by their own particular biases and were fully cognizant of the rationale behind our actions, as well as the circumstances in which we acted. As Smith explains, '[w]e suppose ourselves the spectators of our own behaviour, and endeavour to imagine what effect it would, in this light, produce upon us' (TMS III, I, 5).

Smith's concept of the impartial spectator emphasizes the desire for approbation as a key fact of the human condition. For Smith,

the man of real constancy and firmness, the wise and just man who has been thoroughly bred in the great school of self-command, in the bustle and business of the world (...) has never dared to forget for one moment the *judgment* which the impartial spectator would pass upon his sentiments and conduct. (TMS III, I, 37)

Smith acknowledges that people also endeavour to direct their conduct in terms of its effect upon the actual, live spectators observing them: 'The all-wise Author of Nature has, in this manner, taught man to respect the sentiments and judgments of his brethren; to be more or less pleased when they approve of his conduct, and to be more or less hurt when they disapprove of it.' (TMS III, I, 38)

By reversing the role of the impartial spectator in respect to others, and applying it to ourselves, each individual can measure the rightness of his own actions and passions. We endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it. The impartial spectator signifies one's ability to distance oneself from his particular situation and subjective point of view, and see oneself from a perspective than he can share with others.

By substituting the moral gaze for the social gaze, the individual moral agent attempts to seek approbation from within his own interior self. Our moral consciousness demands not so much that we receive the approval of our fellow-men as that we feel worthy of their

approval. This allows Smith to argue in *TMS* that it is not the love of praise and attention that motivates moral actions, but the love of praiseworthiness; the former is the object of the social gaze, but it is the latter that is the object of the moral gaze. Though we care about the opinions of others, we care much more about the opinions of the impartial spectator. Above all, we wish to seem proper in his eyes. We seek after his approbation and hope to avoid his disapprobation. The ‘man within the breast’ is thus often seen as the key to Smith’s account of the faculty of conscience:

The jurisdiction of the man without, is founded altogether in the desire of actual praise, and in the aversion to actual blame. The jurisdiction of the man within, is founded altogether in the desire of praise-worthiness, and in the aversion to blame-worthiness (*TMS* III, i, 31–2).

It is important to note that the impartial spectator is not ideal in the sense of being perfect in respect of his judgments. The agent asks himself what the judgment of the external spectators would be if they knew what he knows. The information he has is all that is available to the impartial spectator, whose judgment therefore is not indefeasible. Hence, we can never say categorically that the impartial spectator’s judgment is true. Despite Smith’s reference to ‘the judgment of the ideal man within the breast’ (*TMS* III, I, 68), the impartial spectator is not ideal as omniscient, but instead is ideal as idealized: the best, given the limited information. That information is always admixed with error and an affective nature that can yield to pressure from outside forces and, in yielding, distort the agent’s moral judgments. As a moral lens, it is coloured but not opaque, but it still does the work of focusing the moral knowledge on one point.

In Storr’s phrasing, the impartial spectator is ‘the imaginary figure that each of us constructs to offer us moral guidance as we *negotiate* our lives’ (Storr 2018, 3, emphasis added). The use of the word ‘negotiate’ is of great importance here. Many of our decisions need judgment. Judgment, in turn, is an outcome of a particular type of dialogue we exercise with the impartial spectator as we negotiate various aspects of our lives. The dialogical nature of judgment will be explained in the next section. But the language that borders on economic language should be duly noted.

4.3.3. Moral maturation

Smith has hinted at how the judgments of the impartial spectator, that is also judgments

concerning the value of our interests, might be altered as we experience new situations and as we meet new people. Although Smith believes that the impartial spectator is a universal figure (i.e. no one is without being able to reference him), he does not appear to believe that we are born with fully formed representations of the impartial spectator who are able to give us counsel in every and all circumstances that we might encounter as we go about our daily lives. Instead, the man within the breast becomes increasingly refined throughout our lives. As we encounter new situations, we observe what conduct and sentiments prompt approbation from others and which elicit disapprobation, we imagine how an independent arbiter in possession of the relevant facts would respond to our passions and our behaviour in that new context. As such, the impartial spectator is likely to be altered by our being exposed to new and varied situations. The impartial spectator is both a guide as we experience the world and is affected by our experiences in the world. The refinement of the impartial spectator marks our maturation as agents. That is why, in practical reasoning, it is not assumed that we act from a well-defined base of knowledge but rather that this knowledge is constantly in a process of formation, which in turn depends to a large extent on our experience—for instance, as to whether it is a more or less wise spectator.

It is important to note that, in his role as our guide, the impartial spectator considers many motives and desires. In Brown's (1991, 1994, 1997) interpretation, this interplay of the different voices (cf. multiple-selves analysis) should be seen as a 'dialogic' process. One important aspect of this dialogic nature of the impartial spectator's judgment is its 'openness'. It is open to contingencies, to the indefinite circumstances that constitute a particular situation in the ever-changing real world. Unlike instrumental, or theoretical reason, it does not operate only with universals and does not follow general dicta or axioms. Whatever guideline we obtain through his judgment cannot be reduced to a straightforward instruction, e.g. to maximize utility, or buy low, sell high. As an outcome of a dialogic process, such judgment is 'freely subscribed to and freely acted upon, and cannot be predetermined or rule-bound'.

Although judgment of one's conduct or interests is commonly understood to involve some evaluation with regard to generally-accepted rules, Smith concedes that these general rules of morality are formulated in such a loose way that they do not give unambiguous direction to conduct. For Smith, general rules are a second-best substitute for true moral behaviour. They are devices by which insensitive and fallible human beings can have a rough measure of their actions, and by which men either through the force of public opinion or through the enforcement of the courts can control their antisocial tendencies and behaviour (Campbell 2004, 352).

Therefore, individual discretion must play an important role in interpreting them

(Anspach 1972, 185). The essence of moral judgment involves freely taking account of the particularities of any case with its own contingencies and private motivations, and the impartial spectator is ideally placed to do just that, as his location guarantees impartiality while his knowledge of the case ensures that the decision is well informed. The agent imagines how he would feel if he were to view events from a position that is both impartial and well-informed (which also encourages him to seek out more knowledge of the facts). The impartial spectator is based on the faculty of the imagination, here as moral imagination. This process occurs neither by somehow transcending emotion nor by assuming a view from nowhere, but by setting both the facts of the case and the ‘feel’ of the case into the relief of an impartially empathetic guide to one’s own actions (see Campbell 1971, 133; Card 2002, 33; Fleischacker 1999, 125; Goodin 2000, 93).

Smith, like Aristotle, understands moral judgment as the exercise of ‘practical wisdom necessary for right actions in particular situations’ (Hanley 2009, 87). He discusses on a number of occasions how our knowledge of our own best course of action comes from our experience as opposed to solely abstract calculation. Smith’s moral agent is not someone who calculates choice based solely on a set of motivations. His capacity of calculation is imperfect and his motivations are dynamic and constantly under formation. The inclinations to choose one concrete goal over another and the activity of evaluating others’ and one’s own behaviours are both formative processes for Smith, not initial assumptions. This can be seen in his extensive discussions of the formation of knowledge and the formation of desires, values, and interests.

Economic theory treats the formation process as an aside or an afterthought to the judgment about action, while for Smith it is the judgment itself that forms our objectives—meaning the whole process of judging—as it is also the means we take to realize them.

We have seen that self-development and acquisition of virtue in Aristotle and Thomas were motivated by the pursuit of happiness (*eudaimonia*). For Aristotle, this means a life that is self-sufficient enough to afford time for contemplation of permanent things; for Thomas, it consists in vision of God as the end of one’s life. For Smith, the *telos* of human activity could also be seen in the achievement of happiness, albeit understood in more mundane terms than attainment of moral perfection or some state of grace (Hill 2001, 11).

Based on the most popular understanding of Smith’s work, one would be inclined to suppose that it is in the acquisition of wealth that Smith sees human happiness. Though not entirely incorrect, this intuition would be incomplete. Although Smith gives numerous indications that self-interest is men’s rational motive for their material acquisitiveness, their more fundamental desire is to acquire sympathy from other men for their economic situation

and for their riches. Men seek wealth and position primarily in order to obtain approval, the positive sympathy of their fellow men: ‘Nay, it is chiefly from this regard to the sentiments of mankind, that we pursue riches and avoid poverty’ (TMS, iii, 16). As Grampp has said of Smith's economic man: ‘[h]owever strong his acquisitive instincts are, his desire for approbation is stronger’ (1948, 317). Thus, it is not wealth as such that drives us to ‘the emulation which runs through all the different ranks of men.’ It is rather the desire ‘to be observed, to be attended to, to be taken notice of with sympathy, complacency, and approbation’. Man is not content with merely material satisfaction. He aspires to a recognition of his value which can come to him only from the attention of others (Lamb 1974, 678). Smith sees through the materialism to the moral consideration behind it, that of the desire to belong to a group that is worthy of belonging to.

The ‘desire of bettering our condition’ exemplifies the characteristic restlessness of all desire because, as the well-known passage in the WN continues, this desire comes ‘with us from the womb and never leaves us till we go into the grave,’ and ‘there is scarce perhaps a single instant in which any man is so perfectly and completely satisfied with his situation as to be without any wish of alteration or improvement of any kind’ (WN II, iii, 28). It is this particular expression of restless desire that creates the ‘blessing of opulence’, as he terms it in his *Lectures* (1982, 185), because ‘an augmentation of fortune is the means by which the greater part of men propose and wish to better their condition’ (WN II, ii, 28).

Yet, the restlessness characterizing human condition also has to do with one’s anxiety for being praiseworthy. As Berry notes (2012, 485-486), it is in line with modern psychology that we are ‘anxious’, that is, susceptible to Lockean ‘unease’. For Aristotle, all humans properly aim at *eudaimonia*, which is a ‘perfect and self-sufficient end’ (NE 1097b15–20). Those who attain that end are living life as it should be led. It is a complete life and, as such, one without ‘desire’. But for the ‘modernist’ Smith, humans as creatures of desire are always on the move, restless without finding complete or permanent rest anywhere (Berry 2012, 480).

The eventual result is that we do not rely on actual praise or blame but seek to act in such a way that an impartial spectator would approve of our conduct (TMS III, ii, 6). Inasmuch as men are always concerned to be praiseworthy, this anxiety will remain. Smith (1982, 497) indeed professed that ‘man is an anxious animal.’ While it is true, Smith does declare that the ‘natural and usual’ state of mind is ‘tranquillity’, wherein lies happiness (TMS III, I, 75). He further claims that the ‘prudent man’ will have ‘no anxiety to change so comfortable a situation’ (TMS VI, I, 13), in no way can this individual be presumed to be in a ‘desire-less’ state.

When read in this light, *WN* appears to urge betterment rather than mere material acquisition. It speaks not to greedy or avaricious man but to (potentially) just and prudent man. It aims not at accumulation for its own sake, but at civilization, material and moral progress, and the decreasing of anxiety in man. Smith views material and moral progress as interconnected; one cannot exist without the other. In *WN* he made it clear that the wealth lay in the well-being of the people. This covered not only their material prosperity but also their moral welfare (Berry 2010, 13). That we are held in consideration is both ‘the most agreeable hope’ and ‘the most ardent desire of human nature’ (TMS I, iii, 16). This ‘ardent desire’ for esteem constitutes the key explanation of that incentive to better our condition. It is the drive for self-betterment that provides the motive for capital accumulation.

Whereas economists in the neoclassical tradition analyse choice under a full-information assumption (Levy 1995, 305), Smith’s theory of the impartial spectator indicates that any person’s preferences, values, and purposes are neither given nor stable. They are unknown fully at any one moment, but not ultimately unknowable. The epistemic limitation is a result of man’s limited self-experience. The eventual (partial) knowability is a result of careful and ongoing observation of one’s own moral life as if one were a third-party spectator. A critical mark of Smith’s theory is that man needs to learn to see his interests in their true perspective. Finding out what those interests are, however, is a process which economists are unable to capture if they assume given preferences. As Fleischacker (2004, 61, 63) notes, ‘[w]hen we ask after the ‘nature’ of human beings we are looking for what human beings ‘really’ want, beneath the surface trappings (...) Human nature always includes what people aspire to, for Smith; it is never reduced [as in the economist’s version of utilitarianism] to the desires they merely happen to have.’

Toward the end of rescuing that which is valuable in ancient and early modern economics for the sake of a truer modern economics in the chapter that follows, I present a synthetic image of an ‘aspiring agent’. It involves shared aspects of the moral anthropology of Aristotle, Aquinas, and Smith.

Chapter 5. The Aspiring Economic Agent

The view of the human agent that emerges from the above discussion of Aristotle, Thomas Aquinas, and Adam Smith is a picture of a human being directing himself in relation to his own good by practical reason (*phronesis* in the Greek, and the virtue of *prudential* for Aquinas). When a person acts on the basis of practical wisdom it means his choices are embedded in an ongoing process of self-formation. His decisions shape his judgment of his capabilities and interests, which then inform his future decisions. He does not possess complete and accurate knowledge of himself and others, but he is constantly in a process of (self-)development. His wants and needs are not certain and given determinants of his actions, as they are in the mainstream economic assumptions. It is rather his actions which from his unique experience influence his judgment.

Through the exercise of practical reason, he develops and improves his vision of the good and the good life. This image of human agent highlights a number of important aspects of human action and reasoning which could be summarised as the acquisition of values. It is in this context that the concept of an aspiring agent is defined in the following sections. An aspiring agent is someone (i.e., a human person) who through her choices discovers and learns what is meaningful and worth pursuing, and over time adjusts her preferences in line with these values. Without enough experience or at least reflection about what the goal entails, we cannot provide an explicit description of how to arrive at it. We often choose from many imagined futures and can rarely expect with certainty that what we choose will work out. Through open-ended involvement in the various practices of our life we learn to exercise a practical judgment, a judgment that becomes operative in practice but which we cannot always give a fully specified or systematic account of. We nonetheless act on it. It enables us to express what is valuable for us and to form ourselves in accordance with what we find valuable.

The concept of aspiration renders the idea of rational economic choice an inadequate representation of human choice in conditions of an unknown or not fully specified goal. Whereas rational choice theory paints a picture of a goal-oriented agent, guided by known and stable references, the account of an aspiring agent recognizes that decision making is not always based on a well-defined goal. The aspiring agent seeks to construct a good life for himself and those who depend on him guided by a vision of the good which he is open to revisit based on new experience.

Decision making based on practical wisdom rejects the notion that all alternatives can be compared on a single measure of utility. Rather, this process entails an exploration of the

multiplicity of incommensurable values embedded within the available options. In what follows I sketch a more detailed profile of aspiring agent and show how this concept can help understand the uniqueness of a human economic agent.

5.1. Definition of the Aspiring Agent

As showed above, standard (instrumental) economic approach narrows the focus of analysis of choice, since it requires that the desires, needs, interests, and expectations which serve as the input in the economic agent's reasoning have been fully formed prior to the agent's making the decision. In this approach, it is not possible to conceive of choice itself as valuable in itself as a learning experience or self-discovery. Whatever has determined one's preferences (understood as total subjective evaluations) is assumed to have been fully considered and understood so that the choice could be made. The important aspect of human agency that involves the dynamic discovery and creation of preferences (or ends) over time is not inquired into in standard economic theory. I suggest, however, that it deserves attention from economists, especially those interested in preference change. It could be usefully studied under the framework of aspiration.

The concept of aspiration as a form of human agency has been developed recently by the philosopher Agnes Callard (2018). Callard defines aspiration as 'nothing other than the process of gradually working oneself all the way into [the] thought' of what it is going to be like to acquire the value one is pursuing (2018, 34). In other words, it is a process of value acquisition and at the same time a form of practical rationality. It is a process through which one strives to become in some way a better person. This process, as she notes, is well captured in Iris Murdoch's description of humanity in general: 'man is the creature who makes pictures of himself and then comes to resemble the picture' (2018, 49). Callard argues that we can rationally bring ourselves to see things differently: 'Coming to value something tends to represent a deep change in how one sees and feels and thinks. Acquiring a new value often alters the structure of one's priorities' (2018, 5).

I suggest that Callard's concept of aspiration seen as a process of value acquisition captures what lies at the heart of preference formation, even as preference formation is broader than value formation. At the same time, it encapsulates the essence of what human agency meant in the pre-modern virtue-ethics tradition. Drawing parallels between the image of man shared by the above discussed pre-modern thinkers and Callard's framework of aspiration, I propose three defining characteristics of an aspiring economic agent:

- 1) The aspiring agent is a proleptic and expressive reasoner, rather than merely an instrumental reasoner. His acts seem to stem from what he values, not just from what he knows. Proleptic and expressive reasoning determine his choice and behaviour, along with instrumental reasoning.
- 2) Complexity and uncertainty of choice, as well as conflicts and incomplete grasp of values all provide learning opportunities for the aspiring agent.
- 3) The aspiring agent should be regarded as a unique individual with a measure of autonomy, responsible for the formation and development of his needs and wants.

5.1.1. Practical reason: proleptic and expressive

For Callard aspiration exemplifies a distinctive form of practical rationality. In her view, an aspiring agent should be seen as guided by what she calls proleptic reason. Etymologically, proleptic describes something which exists before its proper or historical time. Proleptic reason is a reason which is not fully comprehended by the agent but rather undergoes a process of formation as the person acquires more experience and knowledge of her situation. It is analogous to Aristotle's dynamic process of value acquisition. What makes Aristotle's trainee-in-virtue a proleptic reasoner is that he does not act to match a preconceived goal that he fully understands. Rather, through his action he learns more about the value he pursues. He does so in order to be able to appreciate it fully at the end of his pursuit. Proleptic reason seems better suited to explain one's value change than instrumental reason. For instrumental reason takes values as given.

In addition to Callard's framework I suggest an aspirant is also motivated by expressive reason. This accounts for the fact that people act in ways that encourage a sense of self-worth. While proleptic reason describes the form of agency discussed in Aristotle's theory of habituation, expressive reason corresponds with Smith's concept of praiseworthiness. I shall explain both claims in turn.

An aspiring agent is a proleptic reasoner in the sense that 'he does not fully have in view the end for the sake of which he is acting' (Callard 2018, 15). The end he pursues is one whose value he is also seeking to learn. In that he fails the basic prerequisite of instrumental choice: acting for the sake of some envisioned end. His thought about what he is doing cannot be completely clear for two reasons. First, he has insufficient contact with the value to understand how or why it is to be achieved, and, second, because other values may be

distorting his vision.

Thus, proleptic reasons have a certain provisionality about them which reflects the provisionality of the agent's own knowledge of the goal that he pursues. This shows his inchoate, anticipatory, and indirect grasp of some good he is trying to know better. In these cases, one does not demand that the end result of one's agency match a preconceived schema. For, one hopes eventually to get more out of what one is doing than one can yet conceive of.

Aspiration could be thus described as a movement from one valuational position to another. The aspirant sees that he does not have the values that he would like to have, and therefore seeks to move himself toward a better valuational condition. He senses that there is more out there to value than he currently values, and strives to come to see what he cannot yet get fully into view. Aspiring parents know that they do not really know what it is like to be a parent before engaging in the activities involved in parenting. An aspiring doctor knows that he cannot fully know what it is like to be a doctor until he has spent some time performing his duties as a doctor. Aspiration therefore has a distinctive rational form that is not the rationality of deliberation, calculation, preference, or decision. The aspirant is in a position to appreciate the reasons that motivate his action—though not yet fully.

Aspiration, when seen as exercise of proleptic reason is not captured by a single moment of intention or decision. It changes and solidifies over time. It accounts for the fact that people act not only from, but also, at other times, for the sake of acquiring knowledge of value. Like Aristotle's trainee, an aspirant's disposition is being formed in his learning of the value he is pursuing. He can use his dim or inchoate awareness of value to come into contact with the value that, he can then see, has guided the process all along.

Although Callard focuses her concept of aspiration on proleptic reasoning alone, it seems that more needs to be said to explain how one comes to find some values more appealing than others. Also in need of clarification is what forms one's initial interest in some good that one cannot fully grasp before engaging oneself with it. An answer to this question can be found in what Hargreaves Heap calls expressive rationality, namely, 'the reflective capacity of human beings who self-consciously deliberate about their preferences, groping for what is worthwhile'.⁴² It accounts for one's concern for self-respect and self-worth. It is much like Smith's recognition of the human need of praiseworthiness and capacity to consider how one's actions would seem to an impartial spectator.

As Hargreaves Heap understands it, action motivated by expressive concerns, unlike in the instrumental model, does not form a means to a given end. Instead, it is implicated in

⁴² Similar notions of non-instrumental rationality can also be found in Akerlof (1984) and in Stewart (1992 and 1995).

the choice of ends.⁴³ It is part of deciding on who to be, what ends to have. It would thus be better described as activity which is its own end, and gives expression to the agent's self-identity in the action itself. Expressive reason is

the reason that guides action when primarily taken for its symbolic properties: that is, the action says some, usually, evaluative thing about the person taking it. Thus the person appears 'honourable', 'caring', 'moral', 'just', 'fair', etc. through such expressive acts. (Hargreaves Heap 1999, 143).

The expressively rational individual often acts from the position of uncertainty with respect to what values he wants to entertain. The concept itself accounts for the fact that action is not simply about satisfying given ends. For, it is also implicated in the creation of the kinds of ends that instrumental rationality takes as given.

Choice involves something more than mere preference satisfaction. The pursuit of self-respect is open-ended in a way that defies the means-to-given-ends framework. It is about the choice of ends in conditions of genuine uncertainty, i.e. conditions marked by apparently unresolvable conflicts between different ways of being. One could have done x but did y instead. And that could say something about that particular person. A mother of two young children who, say, would like to be active in the labour market faces a choice between sending the children to preschool when they are very young, which would enable her to go back to work sooner, and deciding to spend more time with them at home, which would allow her to fulfil herself as a mother, but could limit her possibilities for professional development. Either choice says something about her in relation to her priorities.

Similarly, expressive rationality account of some examples of consumer behaviour focuses on the manner in which the purchase of particular good or service can be seen as the agent's way to communicate the expressive meaning of his consumption to others. His pattern of consumption, then, says something about the kind of person he is, the kind of things he values and finds worthy. (This provides another layer of meaning for the term 'conspicuous consumption'.)

In short, expressive rationality expresses one's ideas about what aims one finds

⁴³ One could wonder whether expressive rationality could not be treated as a variant of instrumental rationality with the specification of an objective of self-respect. Hargreaves Heap's answer is negative, because it is not possible to fit all features of human choice behaviour into the means-ends framework: 'It does not make sense to talk about action as maximising self-respect because there is not the information available which would allow the calculation of optimal actions in pursuit of this objective.'

valuable and one's capacity to decide what is worthy and what is not. Insofar as coming to appreciate a value cannot happen in an instant, especially if one experiences competing values, proleptic rationality describes the process involved in learning to see the value that one wants to acquire. Instrumental rationality based on given preferences cannot account for this important element of human choice.

Taken together as elements of aspiration, proleptic and expressive rationality show the latter to be a process in which an agent (i.e. a person) changes himself into 'someone who can appreciate the value of the activity or relationship or state of affairs that constitutes the end of the pursuit' (2017, 44). An important aspect of aspiration is therefore its transformative value: aspiration changes what one cares about. Transformative pursuits aim at values, the appreciation of which is connected to the performance of the activity (e.g. the value of entrepreneurship connected to running a business) or involvement in the relationship (e.g. the value of parental love connected to becoming a parent) in question⁴⁴.

Indeed, this is because the pursuits themselves form a kind of value-education. It gradually changes the agent into the kind of person who can appreciate the value he chose to pursue. Callard (2018, 44) illustrates it clearly when she says that:

In the course of becoming a teacher or a friend or a reader of ancient Greek, one learns to appreciate the values that are distinctive of teaching or friendship or reading ancient Greek. But one does not fully appreciate them until one is at, or close to, the end of the process of transformation. For it is the end-state (teaching, parenting, translating) that offers up the actual engagement with the value on which any full appreciation of it must be conditioned.

5.1.2. Aspiration as a learning process

As the definition of proleptic reasoning already indicates, aspiration could be seen as a learning process wherein the agent learns about the good he pursues. From the perspective of expressive reason, aspiration could also be seen as a process of self-discovery in which one learns what one finds worth pursuing. In non-proleptic instrumental reasoning, the agent looks

⁴⁴ On the phenomenological level this same insight is captured in the increasingly popular debate started by Paul's (2015a) famous article which she later developed in a widely debated book on so called 'transformative experience' (Paul 2015b). The analogy with the aspiring agent in my view is limited to the similar criticisms of instrumental reasoning which are accounted for in the logic of aspiration. The crucial difference between Paul's conception of practical reasoning and the one exercised by the aspiring agent is the latter's reliance on the virtue ethics tradition which is not considered by Paul.

to the world for a match for her antecedently fixed goal. Callard gives an example of a person who goes to the store in order to get milk for her cereal. 'If the store is closed, she will be disappointed—and she knows this in advance' (2018, 121). In other words, she knows in advance what will constitute the success of her action, and what failure. She can say exactly why she is doing what she is doing.

An aspiring agent as proleptic reasoner, by contrast, does not demand that the end result match a schema he currently claims in his possession. For he would not know what to ask for. He hopes eventually to get more out of what he is doing than he can yet conceive of. Proleptic rationality illustrates the possibility that we might engage in practical reasoning in order to improve the very conception of the good that drives our reasoning. It is in this sense that agency of the aspirant can be seen as a form of value learning.

In the context of the shopping analogy, a proleptic reasoner would be better described as someone who enters the store with a vague idea of what she wants (women shopping for clothes or shoes is a common example, hence the change of pronoun) and is open to learning about the possibilities that can potentially meet this unspecified goal she has. She decides to purchase something that satisfies a need which could not be defined in detail prior to her entering the store. But it is also something that she understands to be in line with her well-developed and still developing self-image. Shopping and other consumption decisions form perhaps the most trivial example of learning through aspiration. Less trivial ones could be imagined in decisions through which an agent involves himself in some activity that he understands as potentially transformative of his more fundamental values, e.g. trying to give up smoking, or engaging in a risky venture.

These less trivial cases of learning what is worthy of pursuing are often marked by an internal conflict of values which can be resolved by aspiration. Callard shows that a conflict at the level of value 'fractures the agent's evaluative perspective: in order to get the desirability of one of the things he wants fully in view, he must step out of the point of view from which the other appears desirable' (2018, 61). Seeing one option as valuable gets in the way of seeing the other as valuable so that the deliberative question, 'Which of these two things should I do?' cannot be raised. One cannot set aside one's way of looking at the world and somehow examine things from the outside of one's self, for reflection does not afford a person the opportunity to 'step back' from her values. For example, a smoker who finds smoking particularly helpful in tackling his stressful job cannot assume a perspective of a non-smoker who values his health and the health of his loved ones and is therefore not tempted to smoke at all. However, if this health-oriented perspective becomes appealing to the smoker, he finds himself in a conflict of values.

As an intrinsically conflicted agent he experiences a desire to smoke that his newly discovered valuing of health brings him to see as inappropriate. If he resolves to give up smoking, he could be better described as an aspirant who wants to become a non-smoker. That is, rather than someone who based on a deliberation exercised from a neutral position came to a conclusion that non-smoking is preferable to smoking. The aspiring nonsmoker is marked by the fact that he wants not only to behave differently, but also to come to see things differently. He wants to cease being tempted, which is something he does not yet experience. Since he cannot fully appreciate the position of a non-smoker while he is still tempted to smoke, his conflict cannot be described in terms of vacillation between a desire to smoke and a desire not to smoke. In seeking to resolve the intrinsic conflict, he seeks to inhabit fully and undistractedly the point of view of a non-smoker. In other words, he is learning to appreciate the attitude toward which he orients himself as an aspirant. He proceeds away from his rejected desire to smoke toward a different, anticipated future commitment not to smoke.

The intrinsic conflict which leads to aspiration should not be confused with the similar situation of conflict that can be found in situations of *akrasia*, or weak will. *Akrasia* is defined as failing to act on one's better judgment. It is possible because the agent in question is able to deliberate about his available options and decide what he should do 'all things considered' (Davidson 1980, 22). A dieter, for example, is able to reflect on pros and cons of refraining from a dessert. She can be considered rational if she acts on this better judgment of hers. Her ability to arrive at an 'all things considered' judgment captures the agent's sense that she knows enough. Unlike an aspiring agent she does not feel the need to learn why the value she is pursuing (good health) is worth the effort.

The proleptic rationality characteristic of aspiration, in turn, is directed at the reasons that the agent takes himself to be in the process of acquiring or learning. It is in this sense that the aspirant can be seen as in motion (like Aristotle's or Aquinas's agent who was shown to be *en route* to some completion). His conflict is one that can be depicted only over a stretch of time. For, nothing can move in no time. Hence the proleptic rationality that characterizes aspiration is confined to 'diachronic intrinsic conflict' (2018, 96).

Thus, a proleptic reasoner is someone who is on the road to a new way of thinking. He cannot fully adopt the deliberative perspective of one who appreciates the value to which he is still in an aspirational relation. Aspirants 'inhabit conflicting perspectives not only at a moment, but as part of a larger project of self-change in which they are engaged' (and which is also understood as self-development) (2018, 97). Aspirational conflict is grounded in the fact that one is in the process of moving between two perspectives on value. Over time, this learning constitutes a change in who the person is, which is to say, a change in what she

values.

5.1.3. Autonomy and responsibility

Acquiring new values is a process, and one that requires some degree of effort. Thus, aspiration can be seen as a form of agency. It gives an account of the work that an agent does in making himself into a person with new values. Callard calls this work ‘self-creation.’ In this perspective aspiration is that form of agency in which one acts upon oneself to create a self with substantively new values:

On an aspirational account, self-creation is agent-driven learning in the domain of value. The aspirant brings herself to a different view as to what matters in life, and comes to appreciate what she once did not (2018,108).

One does this by allowing oneself to be guided by the imagined self one is bringing into being. This is what makes the process one of aspiration as opposed to, for example, self-control or self-determination where one’s present values control what one should be in the future.

Insofar as the aspiring agent is not merely open to changes that might come but is active in guiding himself with reference to a target he pursues (however dimly he grasps it), his becoming something little by little not only happens to him, but it is something *he does*. One cannot simply ‘choose’ to become a nuclear physicist or a pianist, regardless of one’s innate talent. And even one’s talent and skills are insufficient in themselves for successful pursuit of such aspirations. Those careers require some sacrifice and dedication over many years. They are in this sense a product of work rather than choice.

This work consists in personal commitments by way of which one engages in a kind of reflexive self-construction and formation of one’s identity. In that it requires not only control of impulses (as is often the case with the akratic), but also undertaking of specific projects, the redirection of inclinations, the reformation of habits, the abandonment of old and adoption of new projects, as circumstances require, and, overall, the harmonization of all one’s deep commitments for which there is no recipe or blueprint. Giving up an addiction is one example of such a commitment, but more profound ones can be thought of, scholarship or any profession, marriage, parenthood, priesthood, becoming a crime boss. The extended time-frame of aspiration allows the capturing of such commitments in a way that is not possible in

the static traditional approach.

The concept of aspiration allows us, therefore, to account for one's wants as work in progress. It does not presuppose the existence of a given or 'true' self, but allows for a possibility that you are not yet 'who you really are' or 'who you are supposed to be'. This can be reasonably interpreted to include the idea that one is always in a process of self-evaluation or self-scrutiny as Sen (2009) calls it. It in turn has important implications for the nature of an aspirant's autonomy.

Autonomy is sometimes understood to lie in one's ability to make choices that align with one's true self against some external pressures for one to make choices in line with what someone else wants one to be—a heteronomous self. An obvious example is a teenager extending his guitar-playing hobby into a career against his parents' wish that he trains to become a lawyer. On Callard's account, an aspirant's autonomy is not of this kind because this account of autonomy presupposes the existence of 'one's true self'. An aspirant's autonomy consists in his ability to make himself into what he really is through discovery of values. He is becoming someone who holds values which he can endorse upon coming to comprehend them.

Insofar as the aspiring agent is not simply someone who acquires a new value unconsciously, but actively works to become a person with new values, he is also responsible for orchestrating the value-acquisition. He is responsible for getting himself into a different value-condition from the one he is in. An aspirant's value-transition is his own work, which means that he is a certain kind of cause: a cause of himself. His work consists in his struggling to sustain interest in some activity, relationship or career or other experience that will later become second nature, as well as in his repeated attempts to 'get it right.' These attempts must be performed without the benefit of knowing exactly what rightness consists in. Most generally, it lays in the fact that he always wants and strives to be farther along than he is.

In saying this, we cannot deny that the aspirant receives help. Major value-acquisitions reflect the influence of one's environment, especially the people in it. Parents, teachers and friends have transformative effects on the people they parent, teach and befriend. But they cannot, at least typically, have these effects on someone unless the person actively participates in the process. They assist rather than substituting for the activity of the agent himself. The work of appreciating the value of, say, wisdom, is not work that someone can do for (or to) you. Virtue itself is not teachable through external means. Although the habits on which it is based are.

Where external assistance is needed, as Aristotle's theory of habituation shows, is in

helping the agent grasp or specify his goal. The pursuit of this goal, and learning to appreciate it is, however, the responsibility of the agent himself. In order for this pursuit to qualify as a phenomenon or manifestation of agency, it must be in some way the agent's achievement.

Agency, in turn, involves responsibility. An act is an event for which it makes sense to ask who is responsible. In the case of aspiration, one is positively responsible, praiseworthy, for the good valuational condition one attains through aspiration. One is negatively responsible, blameworthy, for the culpable failure to aspire to a better condition. One is responsible for one's value-achievements to the degree that one's aspirational work figures into the explanation of one's arriving at the endpoint of holding certain values.

There are aspects of the logic of aspiration which have been extensively developed in other discourses: autonomy, freedom, self-determination. The goal of the thesis is to put a separate lens on it to show economists to what extent these features are important in their discipline. In order to show how the concept of aspiration can help better understand preference formation and change I offer alternative interpretations of the accounts of preference instability discussed above. I describe them as conflicts of preferences which can be resolved by aspiration.

5.2. Alternative Interpretation of Preference Change

The analysis in Chapter 2 of the two informal accounts of preference change that challenge the instrumental framework of rational choice theory exposed a difficulty that the standard framework faces in determining the agent's true self and her true preferences. This chapter uses the notion and taxonomy of an aspiring agent to provide alternative interpretation of these examples. I suggest that looking at both the multiple selves and the sour grapes frameworks from a perspective of an aspiring agent who strives to improve his choices (and thus improve himself) offers an alternative perspective on what is entailed in preference change. Such a perspective is not to be found in the instrumental approach to choice.

The focus of my interpretation is on the fact that in both cases the agent's interests are not well understood in the first place. I argue that the two models may be seen as instances of preference learning that takes advantage of expressive and proleptic reason. As such they can be interpreted as cases of increased self-experience and reflexive self-development rather than utility maximisation. In Schelling's model the formative experience originates in conflicts among various interests and preference orderings within the person. In Elster's framework it is triggered by the change in the content of one's feasible set. In both cases the person

experiences a state of tension between what one can and what one might want to do. His preferences should not be seen as fully determined but rather as being formed.

5.2.1. Multiple selves – not self-command but self-creation

The discussion of the multiple selves framework above showed that many important purposive behaviours are ignored if an individual's preferences and values are treated as fixed and given, and not only given but given unambiguously. The example of a reluctant smoker's behaviour is just one among many which show how preferences do in fact change, even over short periods of time. Since this instability cannot be captured by rational choice theory, it is based on unchanging preferences, Schelling and others offered an alternative theory of changing behaviour. There the agent is viewed as consisting of multiple selves taking turns in commanding his behaviour, based on the respective values of the individual selves.

In the case of a reluctant smoker, one of his selves is committed to giving up his addiction for most of the week, and the other cannot resist temptation to reach for a cigarette at a Saturday party at his friends'. In another famous example used by Schelling, a woman is divided into two selves, one of whom wants to deliver her baby without painkillers. Yet this self anticipates her later self's inability to withstand the pain when it comes. The problem with the multiple selves approach, as indicated earlier, is that it identifies the agent with one of his temporal 'selves'. It therefore fails to give a faithful account of the agent as person. I suggest that the concept of an aspiring agent can provide a more satisfactory account of person. I also offer an explanation of preference change as the person's self-formation.

It is not unreasonable to imagine the woman or the reluctant smoker from Schelling's examples as aspirants who have a certain vision of their future selves which they are attracted to but which they do not fully comprehend. For, never before did they have (or perhaps they cannot recall) the experience of child-bearing or not smoking for an extended period of time. In this perspective their choice to pursue that image of themselves can be described as a process of self-discovery and self-learning. Or it could be a movement from one valuational position to another.

Rather than conflict between multiple fully-formed conflicting selves with complete preferences, we can see the situations analysed in these examples as cases of proleptic and expressive reasoners who are in the process of forming their preferences (and thus themselves) in relation to a certain vision of who they want to become. On the aspirational model it can be true that they are not yet their 'real' or 'true' selves, for their wants are

themselves work in progress. As discussed above, if a person aspires to become a nonsmoker, he is not pretending to already be one. Rather, he wants to come to see the world in the way in which a nonsmoker does, because he thinks that is the right way to see things. The question which of the two selves reveals preferences of the real person appears as the wrong question to ask, for neither is the real self. The person in these examples is best defined as an aspirant who is endeavouring to become (but is not yet) the imagined self.

By imagining himself engaging in the action in question, an aspirant gains access to reasons based on appearance. These reasons ‘stem from a perspective on the action that is unavailable from the point of view of the project of deliberating as to what it is best to do overall’ (Callard 2018, 88). An ‘all things considered’ judgment is not a possibility here because the agent does not have enough knowledge of the prospects that he is comparing. The phrase ‘all things considered’ connotes a finality (or at least termination of enquiry) that captures the agent’s sense that he knows it all or at least enough. But in the case of proleptic reasoning the agent relies on his imagination of a prospect that strikes him as appealing. He does not fully comprehend the prospect until he gets to experience it. Weighing of different selves is therefore not helpful in finding out which preference reflects the true self.

Callard’s distinction between extrinsic and intrinsic intrapersonal conflicts can help explain this. She argues that deciding for the one option over the other is a method of resolving extrinsic conflicts. For the idea of a decision already invokes options being presented, as it were, side by side. Someone can make a decision between two things only if he can ask himself which of the two he should do. Intrinsic conflicts are not resolvable by deliberation (or weighing of options) because one cannot step back far enough from them to even get the deliberative question in view. Deciding between conflicting desires by standing back from them and somehow plumping for the one over the other is not a real possibility for an aspiring agent.⁴⁵ Instead of ‘stepping back to a neutral point of view, aspirants gradually step forward into a state of lessened conflict’ (Callard 2018, 80). Aspiration itself is how he resolves his conflict.

Aspiration is a diachronic process by which an agent effects change on his own axiological point of view. Aspirants aim to direct their own attention to values in such a way as to appreciate more fully one value or set of values, and to become immune or insensitive to those values which intrinsically conflict with the first set. An aspirant is someone who works to improve her desires, her feelings, her evaluations, and, more generally, her own capacity for responding to reasons. On this interpretation, the choice of whether to smoke a cigarette

⁴⁵ Or ‘mak[ing] up one’s mind’ as Frankfurt (1988, 172) suggests.

on Wednesday could affect the likely action of the agent on Saturday. Resisting the temptation to smoke now can help break the agent's habit and reduce the urge later (Whitman 2004).

We can regard the long-term self as having control over the short-term self in a sense of the long-term self, serving as an aspiration, an imagined self that the current self aspires to. In this framework, the person emerges through an internal *process of interaction* among the various aspects of herself, including desires that are in conflict.⁴⁶ The claim here is that the person is to be found in the process of resolution of this conflict by aspiration. In the language of neoclassical preferences, the individual is found in the 'reflective' capacity that tries to make sense of and integrate the competing preference orderings.

Sen proposed that one's reflective capacity could be usefully accounted for in a framework of meta-preferences or higher order preferences.⁴⁷ Schelling suggests that conflicts of preferences can be resolved by pre-commitment. Both approaches might be useful in some contexts, but they cannot explain aspiration. For, both assume that the person in question can be identified either with some existing meta-preference or one of the several preferences that are in conflict. The notion of given preferences is taken here to a different level. The aspiration framework is different in that it identifies the person with a process of interaction between preferences. The resulting model of human choice gains greater depth than the traditional instrumental approach. In this model, individuals seek to discover and refine their own preferences over time, through a process of experimentation and self-correction.

What makes the aspirational cases special, is that they are cases in which the agent's sense of where he is going is gaining shape as the agent learns more about what he wants to value. An active sense of self-management is involved here, but not in the sense of some strategic interactions of well-formed sub-personal agents as they are modelled in the instrumental framework.⁴⁸ It is rather in the sense of discovering and revising one's preferences in view of one's changing values. This is an important aspect of human agency unaccounted for in the standard approach. It is important because in the end it is not theoretical but practical reasoning that provides the answer to the question about what good one should pursue. This practical reasoning is based on our values and what we consider the

⁴⁶ In this case reasoning in terms of multiple selves differs from the approach that I criticise here in that it respects connectedness and continuity of different selves as axiologically significant. Treating multiple selves as separate entities may be seen as a step toward the conclusion that personal identity is no longer relevant in matters studied in this framework. It does not always need to be seen this way. Bénabou & Tirole (2002) are an example of economists who have used the multiple selves approach to defend a conception of the individual as a single being.

⁴⁷ Economics more recently has started to explore this account of preferences, especially in accounting for preferences people have for status and esteem (Akerlof & Kranton 2000).

⁴⁸ See for example Strotz (1955).

right thing to do, rather than a scrupulous weighing of all possible consequences which we have only incomplete knowledge of.

The multiple selves framework seems more useful to capture behaviour known as incontinence or akrasia. Akratic intrinsic conflicts, unlike most intrinsic conflicts, have a short duration. The akratic's intrinsic conflict 'begins when the agent arrives at a considered judgment conflicting with her original inclination, and ends as soon as she acts' (Callard 2018, 82). As mentioned above, akratics differ from aspirants in that akratics do not feel the need to acquire a better grasp of the value toward which they are oriented. They are able to form all-things-considered judgments. In this sense they fit the static image of the person captured in the multiple selves framework. The aspirant, by contrast, is in motion. Her conflict is one that can be depicted only over a stretch of time. Hence the proleptic rationality that characterizes aspiration is confined to diachronic intrinsic conflict. Much like Aristotle's and Aquinas's agent who is perceived as *en route* toward some completion, a proleptic reasoner is someone who is "on the road to a new way of thinking" (Callard 2018, 96).

5.2.2. Sour grapes: adaptive preference change as value learning

Elster's sour grapes analysis has been followed up by a number of critical extensions, qualifications, and clarifications. In this section I use the concept of aspiration to suggest another possible interpretation of what and how changes in the story of the fox, insofar as the fox is understood as a personification of a human agent. In this alternative version of the story, the fox is seen as an aspiring agent and a proleptic reasoner.

As already indicated above, adaptation to a change in the fox's feasible set could be seen as neither sour grapes nor character planning but as character building (or formation) that results from the fox's experience and change of values. Such a change could not be interpreted as sour grapes since the agent actively participates in the acquisition of his new values. Neither could it be rendered as character planning in Elster's sense, for it does not constitute a strategic upgrading of existing options in line with a given higher-order preference but rather entails an adjustment of one's values.

It would not be implausible to suggest that Fox's (let's just make that his name) initial desire for the grapes was based on imperfect knowledge or understanding of the actual experience of acquiring them, which is why his decision to try to reach them can be seen as aspiration. Fox's goal to get the grapes is underdefined in the sense that he does not fully understand what it takes for him to reach them. (He might not even have had that kind of

grapes before, and thus the attraction is to the unknown good that is nevertheless assumed to be good.) The experience of the failure in reaching the grapes teaches him something about himself. With some aspirational projects, our failure to succeed at them will simply mean that we should turn our efforts in a different direction. We will have sacrificed many things for what could turn out to be nothing, or worse than nothing. But this is just to say that the choice itself does not ensure success: for precisely what one fears is that one will not see one's way to finding meaning in pursuing some yet underdefined goal. Fox might understand that the grapes are not really something he should try to acquire. Alternatively, he may realise that more effort is required on his part to reach them. Is he willing to give it? Is he the kind of person who strains, perhaps in vain, for grapes? This is why the inaccessibility of the grapes can be treated as a learning experience. If it is, then Fox's walking away need not mark an irrational change of preferences—a souring of the grapes—but rather a rational decision to acquire new values.

Elster allows for a possibility of a rational change of preference in such case if it results from character planning. I argue, however, that character planning in Elster's sense is not compatible with aspiration. As we have seen, character planning involves strategic shaping of one's preferences in line with one's higher order preference-ranking, so that they coincide with existing possibilities. Aspiration, in turn, means becoming someone who develops a new perspective on these possibilities, not only a new hierarchy of preferences.

Character planning, then, much like Schelling's notion of self-command, boils down to preference self-management so that they satisfy some interest or value that one already has. 'Character planning' itself implies some adjustment in one's preferences that is targeted at a known end by manipulation of the existing possibilities. In the Fox's case this goal or higher order preference is his dissonance reduction. An aspirant, in turn, instead of manipulating his beliefs, decides to change as a person by means of finding new meaning in the options that he has. He is not making a decision from the values he already has. He gradually acquires the 'full version of the attenuated values he currently grasps' (Callard 2018, 16). Such a process requires the agent to be open to form new goals and new reasons to pursue them thus enabling him to become a new person. This alternative interpretation allows for the possibility of genuine learning which is not obvious in Elster's character planning.

Therefore, I propose to describe the process as character building through which one becomes a different person with new values and preferences that one did not have before. Character formation unlike character planning entails a character-transformation. In order to be transformative it cannot lead merely to dissonance-reduction. Character formation may lead to value-discovery. Zimmerman (2003) offers a similar account of character formation

which he calls ‘unconstrained character building’ as opposed to ‘constrained character building or planning’ (he uses the terms ‘planning’ and ‘building’ interchangeably, as do I). His notion of unconstrained character building, however, seems unnecessarily limited in that it refers only to cases of self-abnegation in oppressive situations: a slave, battered housewife, sweatshop worker. Unlike self-abnegation in which one imposes values on himself because they help manage an unwelcome circumstance, character formation is an active process that can take place irrespective of external circumstances. It is a rational process in the sense of proleptic reasoning. One is working to apprehend the value one seeks to acquire.

Even though it could be argued that character planning employs some reflective reasoning, that reasoning is constrained within the limits of one’s given and stable goal (metapreference, value criterion). If the higher order preference ranking is given, this can easily be modelled in the standard rational approach. Character planning is thus rational in the instrumental sense. It consists in the adjustment of reasons for an existing ranking of preferences. It then fits into the rational choice theory framework in which preferences do not change simply in reference to feasibility.

Character planning may be irrational from the perspective of aspiration, since it closes oneself within the values one already has. What one likes and desires is somehow thought to be contained within oneself, and on this basis one tries to achieve what one wants. Character building, on the other hand, relies on being open to change in one’s value criteria in a way that one does not fully control. It requires some notion of mental work, the work of aspiration: being receptive, letting things happen to oneself, letting oneself be changed, and having the capacity to integrate the experiences in a reasonably successful way. The new experience allows the aspirant to see things differently. It is not simple preference learning based on acquiring more information about the object of pursuit. It is learning about one’s reasons to acquire it. There seems to be no space for such openness in Elster’s account.

My alternative interpretations of multiple selves framework and sour grapes situations show that what appears as preference instability and might suggest irrational action, can in fact be explained by a rational process of aspiration. The subtitle of Elster’s book would then be unwarranted: ‘Studies in the Subversion of Rationality’.

Insofar as preference formation of an aspiring agent is a gradual process, which may involve intrinsic conflicts, resolution of these conflicts through aspiration is a crucial aspect of human agency. It is a process where the very identity of the person—the personality—is defined and formed. The perspective of an aspiring agent shows not only preferences alone, but also their source and the manner of determination define the agent and her values.

The strive for self-betterment or self-fulfilment appears to be a crucial factor of human

agency: the *reason* for action in many circumstances, the formal cause of much human activity. However, it has not received serious consideration in the growing literature on preference formation and instability. The standard picture only recognises the fact of preferences inconsistency, the end result of preference change, but does not account for the process of change itself. It fails to account for the possibility that an agent's preference change may suggest important changes in her motivating reasons. Standard instrumental theory cannot tell us how to be rational when the decision's primary import is a fundamental change in one's core values. Therefore, no substantive theory of rational preferences seems sufficient to back up the weighing of selves or determine the rationality of adaptation.

One caveat needs a mentioning. But should we really consider the concept of aspiration as an essential capacity of the person when it seems non-existent in people who (couch potato, social pressure, proud to be bad)? On the one hand, on the face of it, these examples seem to contradict the aspiration thesis. But aspiration could still be seen as innate if we consider that even the worst habits such as addiction or resignation to one's inferior status can be modulated by the eye of the impartial spectator and confronted with his judgment (except perhaps pathological cases of psychiatric disorder or irreformable criminals). On the other hand, the sympathy operator could be considered as an essential model of the individual because it relies on the idea that one has an innate ability to put oneself in another's shoes. But this ability itself depends a lot on one's environment and relations with others.

5.3. Aspiration vs. Welfare: The Good of Aspiring

Because of their emphasis on determining the precise nature and behaviour of preferences, the extended models of agency offered in multiple-selves and sour grapes frameworks seem to remain at least partially committed to the standard approach. Although they are correct to doubt the stability of preferences, and even as they provide insightful illustrations of preference change, they stop short of deeper inquiry into the nature of that change. More importantly, they fail to account for the way that the process of change affects personal welfare.

Instead, they rely on preferences to do all the explanatory work of their models. They do not allow that human behaviour often reveals an incomplete picture of what the person values. It often is merely a step in the process of formation of yet to be revealed preferences. Schelling's and Elster's models loosen up the axioms of given and stable preferences, and allow preferences to change or adapt, but neither of them considers the *meaning* of the

process of change as a process of the agent's self-formation. Elster's account of character planning has been shown to be at best insufficient in this regard. They do not capture aspiration as element of human agency which is valuable in itself. Contrary to this oversight, the virtue-ethics tradition puts emphasis on the good of self-formation. Mill's recognition that the cultivation and exercise of one's own capacity for autonomous choice is in itself 'one of the principal ingredients of human happiness' seems to correspond with this view of human agency.⁴⁹

What do we gain from looking at the above models in the light of aspiration? The analysis reveals that when we look at preference change or formation from a perspective that goes beyond the strictures of instrumental reason, it appears that part of what it means to be rational involves also deciding about or defining one's conception of the good and acting on it. Forming a rational response to one's environment, rejecting preferences one cannot satisfy (the sour grapes), or developing preferences that are in line with one's principles (intrapersonal conflict), depends on one's idea of the good. It depends on one's values, which one learns and endorses or rejects over time while making choices. An important aspect of human agency is to be found in one's pursuits of underdefined goals. Moreover, persons define themselves and their good (at least partially) through the exercise of this form of agency. Normatively speaking, it seems, then, that it is not just the outcomes of one's choices that matter for the agent, but the acts of choosing, and the participation in the pursuits themselves *as chosen pursuits*.

This indicates, in turn, that human welfare is to be found not only in satisfaction of preferences, but also in preference formation. In non-economic vocabulary this is simply one's ability and openness to cultivate a self who can call his values and goals into question and change them.

Part of this intuition seems to be captured in Hirschman's (1984) analysis of people's pursuits of goals the outcomes of which cannot be guaranteed. It is not uncommon that people invest their resources and efforts in activities the eventual success of which is uncertain at

⁴⁹ Mill (2001,55) deals with this question extensively in chapter III of *On Liberty*: "The human faculties of perception, judgment, discriminative feeling, mental activity, and even moral preference, are exercised only in making a choice. He who does anything because it is the custom makes no choice. He gains no practice either in discerning or in desiring what is best. The mental and moral, like the muscular powers, are improved only by being used. The faculties are called into no exercise by doing a thing merely because others do it, no more than by believing a thing only because others believe it. If the grounds of an opinion are not conclusive to the person's own reason, his reason cannot be strengthened, but is likely to be weakened, by his adopting it: and if the inducements to an act are not such as are consentaneous to his own feelings and character (where affection, or the rights of others, are not concerned) it is so much done towards rendering his feelings and character inert and torpid, instead of active and energetic. (...) It is possible that he might be guided in some good path, and kept out of harm's way, without any of these things. But what will be his comparative worth as a human being?"

time of investment. Hirschman (1984, 91) lists in this broad category activities such as pursuit of truth, beauty, justice, liberty, friendship, love, salvation, etc. (many of which are values). On the face of it, these items are all dictated by one's wish to pursue a certain goal. What is distinctive about them is the fact that the agents who engage in them oftentimes derive positive utility from the very fact of undertaking a given action. Even the items that did not begin as values become value-like.

As Hirschman notes, a means-end or cost-benefit calculus is impossible where an agent finds value in the very activity of striving for a goal, irrespective of whether it is successful in the end. Hirschman himself suggests what he calls 'a half-rational explanation', namely, a fusion of striving and attaining that is involved in undertaking of activities with an uncertain outcome: 'He who strives after truth (or beauty) frequently experiences the conviction (...) that he has found (or achieved) it. He who participates in a movement for liberty or justice frequently has the experience of already bringing these ideals within reach' (Hirschman 1985, 13).⁵⁰

Hirschman's analysis is similar to my take on agency. Both show that one can find an intrinsic value in one's pursuits of uncertain outcomes. Both indicate that the standard rational choice theory cannot account for this special kind of 'utility'. What my approach adds to this is a way to unpack the notion of this utility of a special kind and to show why it matters. My contribution also extends Hirschman's analysis to account for pursuits which are of value to the agent not only despite the uncertainty of success but also despite imperfect understanding of the goal itself. Such actions are not driven by a preconceived notion of how one wants something to be, as seems to be the case in character planning. They are more receptive and open; they let things happen. They are rational in a broader sense in which the instrumental mode is balanced by other ways of relating to the world (Sandven 1999b).

The discussion of the choices faced by a self-conflicted and adapting agent respectively show that what is missing in the instrumental means-to-given-ends vocabulary of rationality is the open-ended nature of action. It is reasonable in the first place to see the relationship between desires and the objects of the desires as a two-sided, interactive one. This is so that one develops desires and wants that are reasonably adapted to what one has a reasonable possibility of achieving, to what is in harmony with one's talents, to what one functions well in relation to, and to what one is happy with. This picture is based on a 'doing' rather than 'achieving' model of human agency. The journey is also the destination. It accounts for dynamic action in a manner that is lost on the instrumental account, which does

⁵⁰ See also the distinction made by Winston (1982, 193-97).

not capture instances of incomplete preferences or openness to try something new.⁵¹ It stresses the continuing evolution of preference in response to experience and life events (Hanemann 2007).

In this perspective, preferences, what one likes and does not like, wants and does not want, are thus something that one establishes and develops in interaction with the relations that one establishes with other persons and objects. One's idea of the good, of what one desires, wants, and likes is not formed independently of the confrontation of the desires with reality. Persons are involved with and interact with what they have preferences about. They are not mere spectators who do not participate but only observe an objective state of the world from the outside and then make their choices.

This perspective seems helpful in addressing the problem of true preferences indicated above. It shows that the agent understood as an aspiring person is in a dynamic relation to 'his true self'. Aspiring does not make it the case that he *is* his true self. Aspiring describes the process through which he works to *become* his true self. Thus, an aspiring agent is not a clear-headed maximiser of clearly-defined preferences. He is a more complicated and less certain actor, attempting to sort out what is worth doing and what sort of person to continue becoming.

An economist's response to this might be as follows. While it touches on something which is very human, it is not an issue that need detain economists. Most of the time, when people are acting in the economy instrumental rationality is enough to explain it. Expressive and proleptic rationality should be left to philosophers. Economists will make do with a restricted picture of the economic agent, assuming that people make decisions to satisfy their well-defined goals. Supporters of standard economic theory who could argue that simplifications used in economic account of human choice, especially in rational choice theory, are useful in many explanatory contexts and there is no need to obscure them with concepts that would not contribute to the existing theories. The strength of the rational choice framework lies in its unambiguous mode of explanation and easy-to-interpret causal models. I have tried to show, however, that these simplifications do not account for important factors which motivate choice. These factors, in turn, are indispensable to explain origin of preferences and make it much more important for economists to look more closely at the chooser as an active, deliberating creature and attempt to see the world from his actual

⁵¹ Cf. the contrast between a temple-building style of life, where meaning consists in the finished work, and a Sisyphus-style of life, where meaning comes from the doing, is pointed out by Wiggins (1988) and Taylor (1983). See also Camus (1942), where he imagines that we must say that Sisyphus is happy in his never-ending work.

perspective. Even in decisions which to an outside observer may seem as primarily driven by reasons based on means-end calculation, people are often motivated by reasons which fall outside of this framework of instrumental reasoning. Consumption which is traditionally regarded by economists as a textbook case of means-ends thinking could instead be seen as an act of discovering one's needs, one's values, and even self-formation. Similarly, many entrepreneurial decisions which involve high levels of risk and creativity are difficult to explain *ex ante* merely as the entrepreneur's motive of profit maximisation. If economists want to better understand how people form their preferences, outside of the most trivial contexts, they would benefit from a better understanding of the 'non-economic springs of economic action' (Hirschman 2013) such as one's aspiration to become a better person.

The unfinished or indeterminate character of our ideals and value structure should be accounted for more explicitly in economic theory, at least insofar as the theory informs policy making. It has already been indicated above that changing preferences pose a challenge to welfare economists who came to understand that people's welfare should not be identified with satisfaction of their revealed preferences, if they are unstable and endogenous. The account of an aspiring agent developed above contributes an additional insight to the welfarist programme in pointing out the unacknowledged value of preference formation—self-formation—as an important element of welfare. It can be argued that an accomplishment of one's personal aspirations constitutes a form of a basic good. Increase in this good, i.e. realizing of one's personal potentials, could reasonably be construed as at the same time an increase in the societal common good.

On the other hand, in that process of self-formation, individuals may systematically fail to act in their own best interests, even as they perceive those interests themselves. In view of changing preferences, welfare economists have been posing the question of how to determine good-defining preferences the satisfaction of which would lead to improvement of one's welfare. A related question for policy makers is: Who is best suited to carry out this task? More specifically: to what (if any) extent it is the responsibility of the state? To what extent should it be left for the agent himself to decide? These questions evoke the old dilemma of liberalism which Bernard Russel once identified in the areas of education and upbringing. It has to do with the difficulty of finding such basis for education and upbringing which would not make the teacher or parent choose a particular set of values or ideals. It seems that the concept of the aspiring agent stands in the centre of this dilemma insofar as the source of the good he or she wants to pursue cannot be well-defined.

Wilhelm von Humboldt is quoted by Mill (2001, 54) to have thought that "the end of man (...), is the highest and most harmonious development of his powers to a complete and

consistent whole”; that, therefore, the object “towards which every human being must ceaselessly direct his efforts, and on which especially those who design to influence their fellow-men must ever keep their eyes, is the individuality of power and development”; that for this there are two requisites, “freedom, and variety of situations”; and that from the union of these arise “individual vigour and manifold diversity”, which combine themselves in “originality.”

The account of the aspiring agent shows that the good for the sake of which we make certain choices and refrain from others is often not defined well enough by the choosing agent. We learn, correct and improve the idea of the good we want to pursue through our personal experiences. It is hard to imagine that the state institutions and agencies could be made sufficiently familiar with these experiences and our reflection on them as to be able to correctly identify our true needs and preferences. This is why heavy-handed intervention of the state into determining what is good for us cannot be supported. On the other hand, insofar as individuals need help in discovering the value that is worthy of pursuit, an extreme position which says that we should let people decide for themselves what is good is equally indefensible. Mill (2001, 55) captures this accurately when he notes that “human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing.”

The way the interactions between the individual agents and state-level institutions are conceived of depends on the accepted conception of human rationality. The old questions of state intervention have been put into a new perspective by the debate around ‘nudging’. The theory of nudges is also a good example of how policy making may be affected by which theory of economic agency is accepted. It might seem that nudges are an ingenious solution to the old dilemma of liberalism. The policy of nudging promises a middle-way solution to the problem of state intervention in individual decision making. It is a policy which aims at improving people’s welfare without constraining their choice.

It is this new ‘third way’ instruments which will serve as an illustration of possible application of the concept of the aspiring agent in policy. The account of an aspiring agent suggests two ways of approaching the controversial issue of nudging. Insofar as nudges may interfere with the agent’s process of preference discovery, they may be seen as undesirable obstacles of the agent’s self-development: hence, welfare-reducing. Insofar as they create opportunities for the agent to aspire to a good or value he might have not otherwise discovered, or appreciated enough, nudges can be seen as potentially welfare-enhancing. Using the homo economicus as a normative ideal leads to policy interventions which do not

account for a broader notion of one's welfare. It may potentially result in an increasing manipulation of people's choosing by the state, thus increasing the role of the state in people's lives. When the category of the aspiring agent is allowed to fill out the thin anthropology of homo economicus, the need for allowing people more autonomy for self-formation becomes apparent. The role of the state ceases to be one of orchestrating people's decision making and becomes one of creating facilities for people's self-formation.

The following chapter will present a more detailed and practical illustration of the otherwise abstract problem.

Chapter 6. Economic Policy Illustration: Nudges

One of the major purposes of economic theory is to serve as a toolbox for normative analysis by policy makers. Involved here are both predicting behaviour and assessing welfare.

Traditionally, economic policy advice has been based on the assumption that economic agents act rationally in the axiomatic sense (defined in Chapter 2), within whatever constraints they face. They have been assumed to have stable, context-independent, and internally consistent preferences, as revealed in their economic decisions. The satisfaction of those preferences has been taken to be the appropriate criterion for evaluating alternative policies. Standard welfare economics has, further, been grounded on the postulate of identifying welfare (or well-being) with the satisfaction of persons' preferences.

This approach has been challenged by the growing evidence regarding instability and context-dependence of preferences.⁵² The belief that satisfaction of preferences always leads to increases in well-being, or even happiness, has been questioned.⁵³ By demonstrating that individuals are not fully rational in the instrumental sense assumed by mainstream economics, recent findings of behavioural economics cast serious doubts on the relevance of this normative criterion for judging welfare. If an individual's choices are neither fully rational nor completely coherent, then they cannot serve as a reliable foundation for a standard of welfare.⁵⁴

In response to this challenge, behavioural economists have inspired policy makers with a new approach to welfare improvements which harnesses their findings about various deviations of preferences from the standard model. The new approach suggests that policies can be designed to minimise the influence of various distorting factors on people's choices, so that they make better decisions that also increase their welfare. This approach has become known as 'soft paternalism', or more affectionately as 'nudging', based on a famous book *Nudge: Improving Decisions about Health, Wealth, and Happiness* by its most prominent representatives, Richard Thaler and Cass Sunstein (2008). The book has provided an

⁵² As DesJardins writes: 'Why should this be the goal of public policy when we recognize the obvious and acknowledge that many individual preferences are silly, foolish, vulgar, dangerous, immoral, criminal, and the like? Why should we think that satisfying the preferences of a racist, criminal, fool, or sadist is a good thing (...) What is so good about satisfying preferences?' (DesJardins 2006, 65-66).

⁵³ As DesJardins (2006, 66) writes: 'Whether what I want is a good thing depends on what it is that I want'. Modern happiness research and Positive Psychology gives us an insight of what make people happy.

⁵⁴ 'If preferences are a product of available information, existing consumption patterns, social pressures, and governmental rules, it seems odd to suggest that individual freedom lies exclusively or by definition in preference satisfaction, or that current preferences should, on grounds of autonomy, be treated as the basis settling political issues', Sunstein (1993, 204).

important argument for third-party intervention in the preferences of citizens under certain circumstances. This timely thesis attracted attention of policymakers internationally.⁵⁵ While in general preference manipulations by the state via its institutions appear highly dubious, the idea that people could, without coercion, be ‘nudged’ into certain behaviours has become quite popular. Yet, while becoming ‘one of the hottest ideas in current policy debates’ (Hausman and Welch 2010, 123), it has also given rise to a lot of criticism and controversy.⁵⁶

In what follows I discuss the policy tool of nudging as an example of how important it is for policy makers to understand human choice in a broader framework than that of the abstract homo economicus.

The authors of nudging claim to be offering policy ideas that are ‘informed by a more accurate conception of choice, one that reflects a better understanding of human behaviour and its wellsprings’ (Sunstein 2000, 13). They ground their fresh understanding of human behaviour in recent studies in psychology. They apply their results to the assumptions they make about human decisions and choice behaviour. These assumptions are claimed to be more accurate, since they are backed by scientific results of psychological experiments. What they hope to gain through this approach is improvement of ‘law’s ability to move society toward desired outcomes’ (Sunstein 2000, 38). Nudges seem to be effective where education is not enough but in a long run they are supposed to have an educational value by habituating people into certain behaviour patterns (e.g. saving).

I look at the idea of nudging from the perspective of the aspiring agent. Insofar as nudging aims at improvement of persons’ choices it may seem to be an idea that an aspiring agent would welcome as helpful in pursuing his goal of self-betterment. I argue, however, that in order for nudging to serve the good of an aspiring agent, it cannot rely merely on the portrayal of a false concept of the economic agent as its normative ideal. I show that normative assessment of nudge economics should not be viewed as a straightforward matter. I then qualify the enthusiasm for this type of government intervention which its advocates find in richer model of human psychology. What they have ‘discovered’ should not increase our enthusiasm for government intervention. In many cases, as I will soon show, this enthusiasm seems premature and needs to be checked by more critical reflection.

I suggest (and explain further in Chapter 7) how nudging could be less controversial

⁵⁵ In 2015 President Obama issued an executive order that encouraged federal agencies to employ behavioural insights to design more effective government policy and programmes. Within the UK government the Behavioural Insights Team was created.

⁵⁶ Libertarian paternalism has found a number of prominent defenders (e.g. Camerer 2006; Trout 2005; Loewenstein and Haisley 2008; Amir and Lobel 2008; Desai 2011; Sunstein 2013) and an equal if not greater number of critics (Sugden 2008; Rizzo and Whitman 2009a, 2009b; Binder 2014).

while more effective if its normative objectives are based on a fuller conception of agency. A more filled-in agent which, aside from additions from psychological and cognitive aspects of choice, includes moral motivations standing behind decision making. I argue that it might be helpful to see the potential nudgee (the one who is nudged) in part as an aspiring agent, who does not fully know his or her ends, rather than as a hypothetical creature with complete (or at least sufficient) knowledge of her interests. The alternative view overlooks and potentially undermines an important element of well-being: the value inherent in the formation of one's own preferences.

The view of the person as an aspiring agent provides three categories of evaluation which can help determine whether a given nudge is permissible or justifiable. Nudges cannot be justified if they stand in the way of a person's formation of preferences by preventing at least one of the following: learning from experience, taking responsibility for choices, and development of virtue. The final is imperative. Lacking a development of virtue in the person, nudges will become permanent guardrails preventing departures from a pre-determined path, rather than temporary training wheels that teach one how to ride. Nudges that fail any of the three criteria could be seen as welfare-reducing. Conversely, insofar as they create opportunities for the agent to aspire to a good or value he or she might have not otherwise discovered, nudges can be seen as potentially welfare-enhancing.

Criticisms of nudging come from many different directions (e.g. Bovens 2009; Hausman and Welch 2010; Goodwin 2012; Niker 2017). My own criticism is to some extent related to those expressed by others, but I then put the problem into a new light. My approach offers lines of criticism which when viewed carefully can in fact rescue the project of nudging from many objections which are currently raised against it. For, the agent that I use to criticize nudging is also a better candidate to be open to nudging, rightly understood and carried out.

Section 1 of this chapter provides an overview of nudging. Section 2 challenges the case made for nudging by the authors and the leading advocates of this innovative policy tool. It shows that the key defining aspects of nudging which are invoked in their defense could have two different interpretations, each dependent on the implicit economic anthropology assumed by nudge advocates. One relies on the abstract, static homo economicus; whereas the other is based in the concept of an aspiring moral agent. I argue that nudges can help people pursue their interests if they are informed by a more careful understanding of human agency, rather than relying on the idealised image of man borrowed from neoclassical economics.

6.1. What Are Nudges?

In recent decades, behavioural economics took a normative turn and has become more policy-oriented, making an increasing contribution to a new subfield of economics: behavioural welfare economics and policy⁵⁷. Policy of nudging is one of its most fruitful applications, which is clearly demonstrated in the sustained growth of the literature on the subject during the last ten or fifteen years.⁵⁸

As already indicated, the traditional concept of welfare as preference satisfaction, grounded in the belief that the individual knows best what is in his interest and reveals it in his choices, lost some credibility with the growing evidence that people often make irrational choices. Some people procrastinate, others are unrealistically optimistic, others run serious health risks, and others are discourteous or even violent. If persons' choices fail to reveal coherent or context-independent preferences, then it is no longer possible to identify their choices with their 'true' preferences, which reveal their 'true' interests (Thaler and Sunstein 2003, 1164). As behavioural economics has shown, individuals often do not come to decision problems with pre-existing preferences. They rather form them only when confronted with particular problems, and they are sensitive to details of 'framing'. Thus, the authors of the nudge strategy claim that the decision-making situation can be designed to improve choices, so that they are closer to what the individual would choose in a situation free from obstacles. They have concluded that the findings of behavioural economics justify policies which 'nudge' individuals towards those choices that are in their best interests.⁵⁹

Nudging involves the use of various behaviour-modification techniques and policy instruments. The term was first introduced by Thaler and Sunstein (2008). Though rhetorically powerful, the concept itself has not been clearly defined. That has given rise to much confusion. Thaler and Sunstein define nudges mainly by example. As an analogy filled out by anecdotes and references to studies that show this or that, nudge is not a conceptually strong notion. The most unified account of the concept they propose describes nudge as 'any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their (...) incentives' (2008, 6). In a later formulation, Sunstein defines nudges as 'initiatives that maintain freedom of choice while also *steering* people's decisions in the *right* direction (*as judged by people themselves*)' (2014a, 17; italics added). That is vague in all the right ways so as to remain open for policy

⁵⁷ These authors have started to build the foundations for a discipline the purpose of which is to extend welfare analysis to cases where agents make inconsistent choices (Bernheim, 2008).

⁵⁸ Rebonato (2012) provides a good overview of some relevant literature.

⁵⁹ e.g. Camerer et al., 2003; Sunstein and Thaler, 2003 Thaler and Sunstein, 2008.

uses.

Thaler and Sunstein (2008) demonstrate how policymakers (*qua* choice architects) can make major improvements to the lives of others by designing ‘user-friendly environments’. Examples of nudge policies range from simple techniques, such as serving drinks in smaller glasses, designating sections of supermarket trolleys for fruit and vegetables (Marteau et al. 2011, 263), or redrawing lines on roads to prevent speeding, to requirements that household energy bills contain comparative consumption information – the idea being that this encourages reductions in energy usage when one sees that she is, say, in the 92nd percentile of users (Osborne 2008). The two most cited examples in *Nudge* are: foods-placement in a cafeteria to optimize healthy choices, and default opt-ins for private pensions. In cafeterias a clever ordering of food—with the less healthy choices placed further away—could make it less likely that individuals choose unhealthy food, say, in a sudden moment of weak will. For private pensions, an employer can foster her employees’ savings decisions by making enrolment in pension schemes the default, with an easily available ‘escape clause’ for those who prefer otherwise.

Some nudges work because they inform people, other nudges work because they make certain choice easier, still other nudges work because of the power of inertia and/or procrastination (Sunstein 2015a, 208). A reminder is a nudge, so is a warning. A GPS nudges, a default rule nudges.⁶⁰ By presenting information in particular ways, the state can nudge people towards being more aware of and sensitive to the salient aspects of a situation (as in the case of the ‘traffic-light system’ for nutritional content). Such policies might reduce the subject’s exposure to misinformation or offer helpful suggestions of ways of achieving their goals.

To count as a *mere* nudge as opposed to coercion, the intervention must be easy to avoid and avoiding it must not incur any serious costs on the subject of a nudge. It must also leave the set of available options ‘essentially unchanged’. It is crucial for the concept that the possibility for a person’s own decisions remains. Thaler and Sunstein stress that choice architecture should be primarily intended to facilitate an individual’s pursuit of her own goals (2008, 5-6). Therefore, a subsidy is not a nudge, a tax is not a nudge, a fine or a jail sentence

⁶⁰ The most obvious nudges consist of default rules, which establish what happens if people do nothing at all. Others include simplification (of applications for job training or financial aid); disclosure of factual information (calorie labels); warnings, graphic or otherwise (on cigarette packages); reminders (of bills that are about to come due); increases in ease and convenience (through intuitive graphical user interface design); uses of social norms (energy use disclosures compared to one’s neighbours’); nonmonetary rewards (such as public recognition); active choosing (as in the question: Which retirement plan do you want? Or, Do you want to become an organ donor?); and pre-commitment strategies (through which people agree, in advance, to a particular course of conduct, such as a smoking cessation program).

is not a nudge.

Thaler and Sunstein's concept of nudging is but one account of how to use behavioural economics in policy, as well as, how to understand those uses normatively. Theirs is also the first and best known policy approach of its kind, and the most widely discussed in current debates. It is therefore their account that I take as the case-in-point illustrating why philosophical complexities of human choice may deserve more attention in applied economics.

6.2. Circling Back to Homo Economicus

Following the liberal tradition according to which the citizens themselves know best what is good or bad for them and even a democratically legitimised government does not have the right to judge their convictions (Kirchgässner 2017), economists often reject paternalism. This holds also for soft paternalism. Behind this liberal concept there is the vision of mature and enlightened individuals. They are able to act in accordance with their interests. They make their own decisions in accordance with these interests and the available alternatives. They do all this autonomously.

The notion of a mature and enlightened individual is also reflected in the neoclassical economic assumption of given preferences. Economists' somewhat truncated understanding of autonomy as 'consumer sovereignty' simply posits to respect an agent's given, complete and consistent preferences as the only source and ultimate measuring rod of normative judgments.⁶¹ By harnessing the modified vision of a mistake-prone individual, nudge advocates seem to be undermining this traditional liberal vision. Advocates of libertarian paternalism claim that once preferences are recognized to be context-dependent and tainted with 'behavioural anomalies', the notion of individual sovereignty appears to be not well-defined (Brennan and Lomasky 1983). It is not uncommon for people to mis-predict what is good for them, which raises serious problems for the idea of consumer sovereignty.⁶² This recognition has become the foundation of Thaler's and Sunstein's innovative policy of nudging as they say:

... we emphasize the possibility that in some cases individuals make inferior choices, choices that they would change if they had complete information, unlimited cognitive

⁶¹ See, e.g., the principle of 'preference autonomy' suggested by Harsanyi (1982, 55): 'In deciding what is good and what is bad for a given individual, the ultimate criterion can only be his own (...) preferences.'

⁶² What if what I want is miswanting? (Kahneman 2011)

abilities, and no lack of willpower (Thaler and Sunstein 2003, 175).

While people no doubt often do make choices they later regret, the belief that they would change those choices in an ideal world of full information, perfect knowledge, and unfailing character is not very well founded. In fact, it looks a lot like the assumptions made about choice in the standard neoclassical approach to economic agency. On the one hand, Thaler and Sunstein recognise that on the behavioural level people are not like the theoretical homo economicus, but on the other hand they seem nonetheless to propose this ideal as a normative benchmark. They assume that choices made in these idealised conditions are superior to choices made in a less ideal environment by less than ideal humans.

The apparent reliance on ‘homo economicus’ is even more visible in one of Sunstein’s articles (2012). He quotes Rebonato’s sceptical characterisation of libertarian paternalism:

Libertarian paternalism is the set of interventions aimed at overcoming the unavoidable biases and decisional inadequacies of an individual by exploiting them in such a way as to influence her decisions (in an easily reversible manner) towards choices that she herself would make if she had at her disposal unlimited time and information, and the analytic abilities of a rational decision maker (more precisely, of Homo Economicus). (Rebonato 2012, quoted in Sunstein 2012, 27)

Interestingly, Sunstein does not argue with this description of his theory. But this makes the theory all the more debatable because it implies that a particular—idealised—model of human decision is granted an epistemic priority that is far from being self-evident.⁶³ Libertarian paternalists take behavioural economics seriously in their description of human behaviour, but they otherwise ignore it by normatively adhering to the overly demanding rationality principles endorsed in traditional economics.

In what follows I argue that the normative reliance on the ‘homo economicus’ view of choice is a problematic element in Thaler and Sunstein’s theory of nudging. It downplays the role of human agency in discovering and learning one’s true ends over time. It also fails to appreciate the value of that process for humans. It overlooks the not uncommon possibility that the chooser’s personal experience of his choice situations might over time enable him to develop insights into what is worthwhile and what is not. In short, this behavioural approach

⁶³The ‘constrained individuality’ view of nudge economics essentially abandons behavioural reasoning for the standard view of rationality and individuality (Davis 2015).

can have an adverse effect on a person's well-being. That is, if it aims to affect the person's good merely by influencing causes beyond his control without engaging his conscious deliberation and action.

Contrasting the implicit normative foundation of nudging defined by an idealised, static picture of *homo economicus*, with the image of person as an aspiring agent, challenges three important notions underlying the theory of nudging: mistaken or inferior choice; good or welfare-increasing choice; and that a nudge can increase or, at worst, be neutral for one's welfare. These notions acquire different meaning when seen from the position of an aspiring agent.

6.2.1. Preventing mistakes

Many of Thaler and Sunstein's examples of nudges suggest that the problem they want to address consists in bad choices resulting in decreased well-being of the chooser. The way they define mistaken decisions or errors of choice is problematic, and this section explains why. They catalogue many factors that can lead to mistakes in human judgment and decision-making. Their famous cafeteria example targets the mistake of eating too much unhealthy food. The other often quoted example of the American '401(k)' retirement saving plan addresses the fault of not saving enough money for retirement. Other mistaken choices they list can result from unrealistic optimism (e.g. in decisions related to marriage or starting a business), status quo bias, loss aversion, and other cognitive or volitional factors of choice which are hard to control.

More generally, bad decisions are understood here as decisions people 'would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control' (Thaler and Sunstein 2008). Some commentators (e.g. Lecouteux 2015, 29) have pointed out that this implies that error is defined here relative to the latent preferences that the individual would have revealed if not subject to reasoning imperfections. While it could be disputed whether the reference of some latent preferences is needed here in order to characterise a decision as mistaken, the remark itself points to a different problem with Thaler and Sunstein's approach. They do not conceive of mistakes as deviations from some objective notion of the good. Instead, they understand them as decisions that people themselves regret upon reflection. Nudging is therefore supposed to help people make choices they will not regret. They are supposed to correct or prevent people's mistakes and thus help them to achieve their 'true' or more 'authentic' ends.

One problem with this approach is that the phenomena observed by behavioural

economists are described as deviations, bias, or anomalies which can and should be counteracted. The experiments in psychology and social sciences on which these ideas are based are interpreted from the perspective of standard rationality theory. They are presented as revealing mistakes people would not make if they were like homo economicus. They accordingly show the psychological heuristics and biases people exhibit as incidental and correctable rather than as fundamental to their nature, or even as essential to how people make choices (Davis 2015).

But if we take the human tendency to value the present stronger than the future (hyperbolic discounting), for example, it is far from obvious that it is an error in decision-making, rather a feature of human nature. Nudges that are supposed to correct for this ‘mistake’ are based on a somewhat arbitrary decision to value long-term preferences higher than short-term preferences. Since it can be predicted that a person might regret his choice in the future, the ‘mistaken’ choice needs to be prevented.

It is, however, not always the case that increasing someone’s wellbeing in the future at the expense of today is the right thing to do. As Sunstein himself admits, ‘there can be a thin line between a self-control problem and a legitimate focus on short-term pleasure’ and continuing, that ‘no choice architect should engage in a program of nudging that disregards the importance of short-term pleasures, or pleasures in general, which are of course crucial parts of good lives’ (2016, 47).⁶⁴

Equally arbitrary would sometimes be to reverse the order of priority and privilege the short-run rather than the long-run preferences as individuals’ true ones. Privileging the impulsive over the planning self seems just as arbitrary as, for instance, privileging long-term over short-term preferences, a strategy that is common in multiple-selves models, and rightly criticized (Sugden 2006).⁶⁵ Suppose that people eat high calorie food, or drink a great deal of alcohol, or fail to exercise, because they value today and tomorrow, and not so much next year or the next decade. If a nudge succeeds in getting persons to focus on their long-term interests, it might increase aggregate (intrapersonal) welfare over time. But if a person is seen as continuous over time, and not a series of selves, efforts to counteract present biases are, by hypothesis, undermining the ends of the chooser at the time of choice.

That the mistake is correctly identified and understood seems crucial for the legitimacy and effectiveness of nudges. They are, after all, intended to help people achieve ends which are good for them (their ‘true’ ends) by preventing mistaken choice. The limitation of Thaler and Sunstein’s approach in this regard is also that it does not consider the

⁶⁴ On the importance of purpose and pleasure, see Dolan (2014).

⁶⁵ See Read (2006) on the intricacies of these kinds of implications of multiple-selves models.

possibility that what might appear as poor or mistaken decisions are not necessarily or merely results of cognitive or psychological biases. While—as empirical experimentation demonstrates—in many cases choices are indeed the result of cognitive biases and heuristics⁶⁶, decision-making processes (for there are many) are influenced by other factors which are neglected in the framework postulated by nudge advocates. As the analysis of the aspiring agent shows, poor choices could, instead, be a reflection of one’s incomplete understanding of what one should value or which of one’s values should (now) be pursued.

Thirdly, whether a choice should be considered a mistake does depend on the context. However, the context is not only related to the decision framework. It is not per se mistaken to choose an ‘objectively’ less healthy option—cake over fruit. The impact on health varies for different people, depends on their general condition of health, life-style, age, how often they have already or are likely in the future to choose cake, etc. Whether a certain decision is in the best interest of some individual depends on properties of the individual herself. The merit of one specific decision cannot be evaluated without knowledge of these properties (Schnellenbach 2012).

From their understanding of mistaken decisions Thaler and Sunstein infer that nudges are a good thing given that they allow individuals to overcome their various biases and blunders (Thaler and Sunstein 2008, 19-41) that affect their everyday behaviour, and to help them resist temptations (Thaler and Sunstein 2008, 43–56). This approach neglects, however, the possibility (detailed above) that persons form their values and ideas of the good in part through learning about their biases and weaknesses. Their real or apparent mistakes might, in fact, help them understand what is good for them by providing what I have called a formative experience. The authors are endeavouring to correct errors in an individual’s reasoning while respecting her judgements about her own well-being. In doing so, *Nudge* does not seem to appreciate that sometimes people only arrive at their judgments of well-being and learn to appreciate them in their efforts to overcome the difficulties, temptations, juggle priorities, etc. A reluctant long-term smoker who makes repeated resolutions to give up his addiction is a case in point.

In asking what an agent would choose (what his preferences over relevant outcomes would be) if he experienced no problems of cognition or self-control, nudge advocates imply that there is a true choice one would make if one had appropriate conditions for choice. A nudge is in place because it ostensibly creates the appropriate conditions. But the account of an aspiring agent shows that as long as the person has not fully formed the idea of the good

⁶⁶ Requiring a person to choose from thousands of mutual funds for their pension schemes has been empirically shown to decrease choice rates as opposed to choice situations with only few mutual funds (Iyengar et al. 2004).

she wants to pursue, it is hard to say what decision framework can help her achieve the end. It does not seem possible to design a choice framework to help realise an unknown good as end. Therefore, constructing a ‘proper’ decision framework cannot be seen as a sufficient factor determining the proper choice for that person. It will not be sufficient for the person to make a choice she will not regret.

The tendency for customers in a cafeteria to favour the earlier items in the line is just one of the many factors in their behaviour that needs to be taken into account. That cake is chosen over fruit is not necessarily, or not exclusively, as a result of how food is organised in the store. While the choices of the cafeteria customers are indeed context dependent in ways that have psychological explanations (more prominently displayed items are more likely to engage attention, etc.) it still does not seem relevant to customers’ interests or goals broadly understood.

In the exercise of their decisions people are motivated by what they think is good for them. If that notion is in flux, because, for example, the person is only learning to appreciate the value of eating healthily, this value cannot be assumed to be a known (i.e., understood) ideal that guides his correct choice. It could be precisely because the end which guides the choice is not fully developed that the choice is seen in retrospect as a mistake. If one regrets the choice upon reflection, it does not necessarily mean that it could have been prevented by a modified arrangement.

When it makes sense to learn by oneself and when we should learn from the experiences of others very much depends on the concrete situation. Learning by trial and error is not always better (Schnellenbach 2012) than learning from the wisdom of others or, indeed their trials and errors. On the other hand, Glaeser (2006) shows the argument that nudges reduce the incentives to learn is ambivalent. One needs to investigate concrete cases. The existence of behavioural anomalies demonstrates that individual learning is not sufficient to avoid these.

Preventing people’s mistakes can be relatively uncontroversial in cases in which it is obvious or generally agreed what is good for people (e.g. crossing the street properly). In the much more common cases in which it is not obvious what a person’s good consists in, nudges become problematic since there is a danger that they arbitrarily impose ideas of the good or make discovery of the good for oneself impossible.

6.2.2. Nudging for the good

Nudging is intended to improve welfare of the persons concerned. In order to succeed, choice

architects need to know which choice reflects the person's well-being. Is it better for the person if she spends or if she saves some part of her income (and what part in each case)? Is it better for her if she buys insurance or if she does not? Is it better for her if she keeps her diet or enjoys a family meal, indulging in food she would otherwise refuse herself? It is hard to see how these questions could be answered merely by simulating what the person in question would choose if she had been free from temptation and reasoning imperfections.

This is however how Sunstein and Thaler propose to resolve such problems when they say that people are not acting in their own best interests if their decisions are ones 'they would change if they had complete information, unlimited cognitive abilities, and no lack of willpower' (2003, 1162). Sugden notes that it is difficult to see how such an idealised criterion can be treated as empirical. For, it is not obvious how we can determine what complete information, unlimited cognition, or complete willpower entail without making normative judgements. This counterfactual reasoning advocated in *Nudge* has been rightly criticised as (yet another) 'view from nowhere'.⁶⁷ The argument relies on a purely hypothetical basis: If individuals were fully rational and were choosing according to their informed preferences, they would do *X*. Surely, one may wonder what is it to be 'fully rational', and to have 'informed preferences'? Similarly, how can one know (or even make sense of) what choice someone endowed with 'unlimited cognitive abilities' will make?

The 'view from nowhere' is shared by behavioural and welfare economists (Hédoïn 2015). It is a perspective typical of 'homo economicus', which has no history, no future, no commitments, and no context-specific considerations. This is not the perspective that the aspiring agent assumes in his decisions.

One way to solve the problem of determining what is good for the person is proposed in behavioural welfare economics: find out what people truly want.⁶⁸ That provides theoretical basis for nudging. But it is not always obvious what 'to truly want' means. If we take the above example of the reluctant smoker, it is not immediately obvious whether his true wants are those revealed at the moment he resolves not to smoke or when he succumbs to temptation and accepts a cigarette from a friend. The chosen position is defended as desirable from the perspective of the chooser himself. But this is simply a reiteration of the counterfactual position. The true choice, according to nudgers, is one that the chooser will not regret.

⁶⁷ Cf. Sugden (2008), Hédoïn (2015).

⁶⁸ A similarly sounding concept, one of 'true preferences' was suggested by John Harsanyi (1982, 55). These are preferences a person would have if she 'had all the relevant information, always reasoned with the greatest possible care, and were in a state of mind most conducive to rational choice'.

The conceptual basis of nudging seems to share the more profound problem entailed in liberalism. That is the assumption that one can be allowed to pursue one's idea of the good within an ethically neutral institutional framework, which itself respects each individual's judgment and does not impose any value criteria on them. In a libertarian spirit (here meaning just hyper-liberal and morally libertarian, rather than anti-government) Sunstein and Thaler recommend an unassuming policy tool. It is not to contain a doctrine about what constitutes people's welfare.⁶⁹ It merely suggests ways to improve people's own decisions regarding what is good for them.

In this, they want to be able to say that their proposals will steer each individual in the direction that he would have chosen for himself under ideal circumstances. In their counterfactual reasoning they seem to rely on the idea that anything increasing the extent to which people make choices that they regret less than they otherwise would is respectful of their internal sense of what is important.

However, as Anderson (2010) has noted, it is one thing to determine what people's preferences would be if they had true beliefs and were free of rational foibles, and it is a different thing to determine what is good for people. *Nudge's* approach seems to suggest that one's decision situation can be modified so as to help him achieve an outcome that he finds agreeable, without the need to define that outcome in advance. The decision framework itself is meant to be neutral ethically. It is as if one's ends could exist or be formed in isolation from one's decision situation.

Binder and Lades (2015, 11), however, argue that a substantive theory about *which choices actually make individuals better off*, is ultimately needed. Similarly, Hédoin (2015, 75) notes with regard both to proponents of nudging and other forms of 'soft paternalism' that 'they cannot avoid grounding their normative extension of behavioural economics on an explicit moral doctrine (i.e. a systematic philosophical account of what is good or right), which here should take the form of a substantive theory of the good'. Hausman also shows that in order to determine what is good for people, 'some grounds independent of their distorted preferences' are needed. If policy makers can determine what is truly good for individuals, then they can devise policies that will lead people to make better choices.

Intuition or common sense may help here. But the point is that some theory of the good is ultimately required. Scientific results may also give indications regarding what is the

⁶⁹ Sunstein and Thaler avoid defining welfare claiming that they 'are not attempting to say anything controversial about welfare, or to take sides in reasonable disputes about how to understand that term' (Sunstein and Thaler 2003, 1163). The abstract character of this concept has been criticised by Sobel (2016, 51) who argues that the 'the notion of a fully informed self is a chimera', because the great variety of possibilities, choices, and lives an agent might lead are not available to a single consciousness.

best thing to do (*e.g.* the fact that the regular consumption of certain kinds of fatty food increases the risk of fatality due to cardiovascular problems). But this only indicates that positive and normative considerations are deeply entangled. A choice architect has the responsibility for organising the context in which people make decisions. But unless what is good can be objectively determined or agreed upon, the goal cannot be improvement of one's welfare. It is hard to imagine how policy advice could be formulated without a concept of well-being or welfare. What is deliberately undertaken must have a goal. Nudges are no exception. At the bottom of this choice-steering, there must be some fundamental value which cannot be entirely eliminated in the process of habituation of citizens on this liberal model. Relevant literature in the field still lacks a concept of well-being which could be consistently applied in a "behavioral" setting, *i.e.*, in a setting where individual preferences change and tend to be inconsistent.

In avoiding taking a position on what is good for persons, nudging offers a procedural approach to welfare. What is not captured in this ostensibly neutral approach is the possibility that one's formation of ends and ideas of the good itself is an important element of choice and important factor of welfare. The idea of nudging does not account for this possibility. It locates wellbeing in given end states, outcomes of choices which one would make, if one were a fully-knowledgeable, rational agent. The nature of this outcome is seen as mere utility, or as reducible to utility.

According to Sunstein, what people ultimately want is the promotion of their own well-being, and it does not really matter how that comes about. Choice is evaluated here only by outcomes. What matters for the choice architect is that a desirable outcome is reached. The choice framework is designed to steer people to make the correct choice. Arguing, however, as they do, that no actual choice takes place in an idealised, context-free world, Thaler and Sunstein propose that people's choice situations are to be designed in ways that minimise the negative influence on decisions of the anomalies of cognition and weak will.

By definition, they are not able to design a normatively neutral framework. They instead propose that nudges should steer people toward choices which help them achieve ends that they themselves approve of. The aspiring agent, however, does not act to maximise some well-defined utility. He acts, rather, to realise some vaguely specified value, often in order to 'to become the man he wants to become' (Buchanan 1999, 259).

Action is not always or merely outcome-oriented. It could (also) be an expression of one's values. The end outcome is not the only criterion defining the right action. For the aspiring agent, the true, correct, superior preferences are ones she is busy learning to value, rather than ones she would have valued if she were not biased.

6.2.3. Skipping levels

The third important assumption which seems implicit in nudging holds that human limitations are best corrected by a third party through a properly designed nudge. This is apparently because people in their fallibility and susceptibility to bias cannot be trusted to correct themselves. The choice architect responsible for the design of choice environment is seen as an impartially benevolent spectator, who helps persons achieve their own ends without influencing them normatively. If therefore, she nudges a person into ways of acting which she finds in the person's best interest, and it turns out to be in that person's best interest, we should be inclined to welcome nudging as a true improvement to decision making. There are at least two reasons making this optimism unjustified.

First, as mentioned in the previous section, it is rarely obvious what one's true interests might be. People not only are uncertain whether their actions serve their goals, but in some important circumstances are also uncertain of the goal itself. When it is not clear what the goal is, the choice architect needs to make a value judgment that is not necessarily in line with what the person's judgment if performed in an ideal environment. This creates a danger that the choice architect will be relying on her own values and beliefs instead of the chooser's. Since a nudge cannot be neutral among available alternatives, in practice it steers people toward choices that the architect expects are going to be seen as beneficial by the chooser himself. Since there is no 'neutral' frame, any way of presenting a situation of choice will necessarily make an option more salient than another one (by defining, say, a default option, a first option in a list of alternatives, a set of possible options, etc.).

But even if nudges are designed so as to further ends which choosers approve of, there is a second, more profound problem with this strategy for welfare-enhancing. In line with behaviourism nudging takes people to be (mostly) passive, plastic, malleable entities. Humans respond to external modifications in their environment, but do not influence it themselves. Given that the capacity to form one's ends is an element in a person's well-being, when a person's value formation is substituted by a third party's valuation (which is arguably what may happen in the process of nudging), he can be justified in feeling deprived of his inherent capacity to determine his own ends based on his own values.

My analysis of the aspiring agent shows that changes in preferences, values, or desires can also be initiated by the agent, rather than merely passively experienced. It is thus unlike the anthropology implicit in the multiple selves and sour grapes examples. Neither of which

account for active participation of the person in formation of her identity. As rational agents, in Korsgaard's words, we 'are faced with the task of making something of [ourselves]' (2009, xii). With the passive, atemporal homo economicus as its implicit normative standard, nudging neither seems interested nor capable of helping people develop capacities or abilities for this kind of transformative choice (Wartenberg 1990).

Nudges can alter the behaviour of individuals to coincide with those who accept certain norms, but it cannot provide *the reasons* necessary to alter people's behaviour in the long run. Thus, on their own, successful nudges merely lead people to act as if they had a well-established notion of the good they pursue, without attempting to engage with the individual at that deeper level where he might be, actually only, trying to establish what end to pursue; whereas the aspirant's goal is nothing other than coming to see the value for herself.

Insofar as nudges are designed so as to select the 'prudent' or 'best' alternative without having to invest the otherwise necessary careful deliberation they do not engage the chooser's key capacities and potential. One could go as far as to claim that nudging could stand in the way of a person's taking efforts entailed in aspiration. This is especially true if it prevents the person from acquiring the kind of experience in which he gains knowledge about what is and is not good for him.⁷⁰ Imagine a video game in which a player was allowed to win a game *by* skipping a few—perhaps the most difficult—levels. The general concept of nudging does not seem to appreciate this heuristic aspect of personal identity.

The idea behind nudging is that the decision framework is to be so designed that one's deficiencies and weaknesses would not have to be overcome consciously in order to produce a desirable outcome. This approach relies on a vision of individuals as passive entities *that* are either unable to build their own normative assessments or who do not particularly care about forming them. In Shiffrin's suggestive terminology, paternalistic policies attempt to substitute the policy-maker's judgment of what is good for the agent for that of the agent. In such interventions the choice architect usurps the practical reasoning process.

This view stands in contrast to what Aristotle, Aquinas and Adam Smith thought of human welfare. Each of them showed in their own way that the good of humans includes and is realised by their active contribution to the shaping of their own preferences—by their 'own intelligent creativity and self-determination' (Aquinas, I-II prol.). People are not just 'passive

⁷⁰ Goodwin (2012) argues that alongside these traditional restrictions on freedom, the process of self-realisation and overcoming internal obstacles to action (e.g. addictions, phobias, aversion and prejudices) may also be important; I would say, as part of positive freedom.

loci of experience', whose welfare can be satisfied without their active participation in the process. Something can suit or fit a person and hence be good for her only if, in some sense, she could come to care about it on her own.

From the perspective of an aspiring agent whose values and understanding develop over time, nudging may be problematic insofar as it aims to influence the pursuit or formation of the agent's commitments. It could also be problematic to the extent to which it allows her to learn and to decide what to value based on personal experience. This argument does not apply to all forms of nudging, for there are too many to list. But it might nonetheless serve as an important factor in deciding uncertain cases in which there is a danger of compromising a person's capacity to decide for himself and form himself.⁷¹

A similar argument has been raised by White (2013, 127) who claims that the greatest threat posed by nudging is what he calls 'value substitution'. This idea refers to the fact that since regulators do not have access to information about people's true interests, they must impose their own idea about them when designing paternalistic rules and policies. White locates this threat in the area of personal autonomy which he defines as 'the right of individuals to govern their own lives, setting their own agendas, goals, and interests, and being free to make choices as they choose (provided those choices don't interfere with anyone else doing the same).' What is special about this definition is that it stresses the importance of self-government and self-determination which can, but need not, be interpreted as capacities of an ideal, fully rational, fully formed and fully informed agent.

White is not, of course, the first one to associate self-governance and self-determination with autonomy. In his analysis of possible risks posed to autonomy by the policy of nudging, however, he implies something that is rarely pointed out in other similar accounts, namely that autonomy as self-governance and self-determination should matter not only in terms of choice itself, but also in terms of motivations that govern those choices in the first place. For White, therefore, interference of nudges in this process of forming one's values may be harmful to one's autonomy.

The above discussion of the key assumptions which are implicit in the notion of nudging shows that insofar as the design of nudges relies on the conceptual framework in which the abstract, atemporal homo economicus operates, it is not necessarily fit for furthering well-being. That is all the more true if one is referring to the well-being of the person as aspiring agent. In her, welfare can be found in the capacity for self-formation, which homo

⁷¹ For similar arguments see Darwall (2006) and Shiffrin (2000).

economicus does not have. What follows from the above discussion is that the advocates of nudging fail to consider human agency and wellbeing from a broad enough perspective. They fail to account for a realistic picture of human agency. In trying to influence people's preferences rather than their reasons for preferring, they seem to neglect that real persons, unlike the rational homo economicus, are often not able to choose what they truly want, at least not in an(y given) instant. In their decisions, they exercise proleptic and expressive, not merely instrumental reason. What they truly want often depends on values which they form over time, not on a choice framework itself. And those values are valued for reasons unrelated to achieving some specific goal.

The purpose of the foregoing was to show that nudging, at least in Thaler and Sunstein's version, informed by the counterfactual reasoning of homo economicus, relies on a misjudged normative ideal. Its objective is to prevent people from making mistakes they will regret. The perspective of the aspiring agent suggests that the tool of nudging could be better used to facilitate people's formation of their goals and themselves. Could nudging be informed by the aspiring agent to help achieve his ends? This is the subject of the next chapter.

Chapter 7. Nudging for the Aspiring Agent

The problems with nudging discussed in the previous section do not necessarily disqualify nudges as a potentially useful policy instrument. Sunstein and Thaler are right to note that processes of individual preference formation are highly context-dependent, i.e., preferences are heavily influenced by the given choice architecture. That people at times benefit from some nudging seems fairly uncontroversial, especially where nudges seem to work quite well (road markings to help visitors who are used to right-hand traffic to cross the street safely, or cash machines that return the card before delivering the cash). In some cases, this so called ‘choice architecture’ (Thaler and Sunstein 2009, 89ff) is unavoidable.⁷² Normative evaluation of nudging should not therefore take as its reference point a situation free from paternalism, which some critics of nudging might consider ideal. ‘No nudges’ does not seem to be the right response: it is neither feasible nor desirable. And there is a common-sense aspect of nudging that anyone who has ever been a teacher or a parent understands and does well to take advantage of.

I do not claim that nudge advocates offer baseless concerns about various threats to our wellbeing that arise from our biases and misuses or abuses of liberty. However, I do argue that a more complicated story can be told about the issues in question. It is not the case that merely because nudges are relevant in certain situations, we can then justify their use in any or all contexts. It is thus quite important not to throw all nudges into one bag as a uniform category. A nudge is one tool of instruction, and it can aid in (self-perceived) heuristic learning, as it can also hinder the same process.

The question that arises from the debate on the permissibility of nudging is not whether we should abandon the use of nudges altogether. The question we should ask is rather about specifying the purpose nudging is meant to serve. Then we should establish whether that purpose forms a realistic goal of contributing to the well-being of real people, rather than, say, that of an artificial construct called homo economics. As the previous chapter showed, it is important that, by helping to increase one element of welfare, nudging does not block other ways in which people achieve well-being.

In what follows I propose a positive account of morally defensible nudging understood

⁷² While there are indeed choice situations where a planner cannot avoid framing the options and influencing decision-makers, the claim that this is always the case is unfounded. Cf. Sugden (2008) who questioned the inevitability of preference manipulation, proposing as an alternative to leave it to the market how to design and present choice situations. Although shaping still raises a flag because of the possibility of one agent controlling another, it seems that in many cases some form of paternalism will be inevitable (Glaeser 2006).

as *induced self-formation* of the person. By this I mean the use of nudges which increase people's welfare by providing them with means to improve their conscious decision making and not by manipulating their choices and inhibiting their reasoning.⁷³ I suggest three applications of nudging which show how a given nudge can help further the goals of an aspiring agent by taking account of the features characteristic of human agency discussed in previous chapters. The proposed applications support the use of those nudges which allow the person scope for self-formation (moral, social, physical, spiritual, etc.), and that do not deny her the status of agent—namely, the one who chooses *for herself*. They rely on the potential of nudging to facilitate development of virtues, to enable learning of ends (and reasons for them), and to encourage the choosing agent to assume responsibility for her actions. A nudge which combines all three capacities stands a chance of being genuinely transformative of people's behaviour, with their own active and conscious participation.

This is in opposition to nudges being used merely as one-shot devices to make people behave so as to bring about desirable policy goals. True, nudge-style techniques make no claims about the potential for the conversion of behaviour into virtuous dispositions. Hence, any expectation of such might (rightly) be taken as a case of demanding more than they are designed to deliver. Still, it seems that there is scope for nudges to play a role in facilitating a virtuous life. While this would not represent the typical idea of a nudge as defended by libertarian paternalists, it nevertheless provides a worthwhile adaptation of their original design. If successful, this type of nudging would not be just freedom-preserving but could potentially be freedom-extending.

The range and diversity of policies recommended under the umbrella of nudging is very wide indeed (and the list is still growing), which forms part of the difficulty in addressing the question of the permissibility of nudging. I have already suggested above that nudges differ from one another in many aspects. The right balance between permissible and controversial nudging is therefore domain-specific. I will point out relevant differences where necessary. My account of permissible applications of nudging, however, is meant to be general enough to serve as a benchmark for most types of nudges. Sections 7.1 to 7.3 specify the three possible goals for which nudging could assist the aspiring agent. For illustration of the largely theoretical framework sketched in sections 7.1.-7.3, the last section of this chapter discusses a case of entrepreneurship as a potential area for the beneficial use of permissible, i.e. aspiring and agent-friendly nudges.

⁷³ A somewhat narrower but similar approach can be found in Binder and Lades (2014), who demand that any plausible paternalism should help individuals learn to become better decision-makers, and not undermine individual autonomy.

7.1. Virtue-Developing Nudges

In Chapter 5 the aspiring agent was defined as a proleptic and expressive reasoner. This reflects that there is a part of personhood which engages in actions for the sake of acquiring a value one currently does not have. For example, appreciation of motherhood or self-owned business. Through the exercise of proleptic reason, the aspiring agent learns to appreciate the value he chose to pursue. He changes himself into someone who over time develops better understanding of the reasons for holding the value. As a trainee-in-virtue (each virtue based in a particular value), he is also prone to make mistakes until he more fully understands the meaning of the value, including its attendant virtue(s) and opposing vices. In this section I consider the possibility of using the tool of nudging in order to help the aspiring agent on the way to acquire and pursue virtue.

As mentioned earlier, on Aristotle's account of virtue ethics, virtue cultivation is characterized primarily by the development and exercise of practical reason. It follows from this that nudges designed to cultivate specific virtues could work by supporting the conditions necessary for the development of practical reason. Recall that practical reason is the virtue needed to manage the other virtues. It is needed to discern and judge which virtues are required by a given situation, and to what extent each is needed. One could expect that nudges which are able to support the development of practical reason can assist a person in identifying a circumstance in which virtue is required, and identifying how best to act on it (once sufficient reason is given).

In a thorough exploration of the principal ethical objections to nudging as a regulatory tool, Sunstein (2015d, 43) makes a point which is reminiscent of Aristotle's account of virtue-acquisition:

Choice-making is a muscle, and the ability to choose can be strengthened through exercise. If nudges would make the muscle atrophy, we would have an argument against them, because people's learning depends on muscle-strengthening. We could imagine an ethical objection that would contend that some nudges do not allow people to build up their own capacities, and might even undermine their incentive to do so. If so, people's welfare is reduced.

This objection is in conversation, probably intentionally, with Aristotle's theory of

habituation. By repeatedly acting like a virtuous person, one learns to be virtuous oneself. One learns gradually to develop the ability to resist various urges or temptations or other obstacles which prevent one from acquiring virtue, and thus becoming a virtuous person. Sunstein himself agrees that not all nudges can be imagined to have that muscle-strengthening quality, and that more detailed analysis of a given nudge is needed to assess whether it is the case. For him, individual choice is helped by those nudges which promote active choosing or demand prompted choice, as well as those which give information and serve as reminders. He is more sceptical about the capacity of defaults to strengthen the choice-making muscle. He carefully suggests that, with respect to defaults and other problematic nudges, detailed characteristics of a given setting need to be investigated before any assessment can be made. Nudging which exploits cognitive biases would seem to have little to do with virtue formation. It is all the more foreign to the process when it results in value-substitution.

As I explained above this happens when the choice architect usurps the practical reasoning process ‘on behalf of’ the agent. Its aim is to generate the ‘right behaviour’ apart from (or, at the least, regardless of) the correspondingly right motivation. For, nudges tend to be understood as interventions designed to modify our automatic behavioural responses—decisions made in an instant—promoting certain targeted ‘welfare-promoting’ behaviours. They do so irrespective of the motivation or intentionality of the agent on whom they act. Such a focus on behaviour modification appears to be at odds with the focus on the development of character, which is central to the project of cultivating virtue. But we do not become just (or temperate, or brave) *merely* by doing just actions. If we perform just actions simply because we have been nudged unthinkingly – meaning, psychologically manipulated – into doing them, then we have not become just persons. If nudges override our conscious reasoning and block the learning process, they cannot help us to learn to make better choices in the future (White 2013,102). Mere behaviour-modifying nudges may even hinder our ability to develop the virtue of, say, self-command, or autonomy (Bovens 2009). On the other hand, however, it seems that nudges should not be seen merely as (potentially) dangerous for virtue acquisition. Certain restrained uses of particular classes of nudges could instead, with some improvements, be designed to help people cultivate virtue.

This is the position which Fay Niker (2017) takes in her Aristotelian account of morally acceptable nudge-style interventions into citizens’ behaviour. She claims that such interventions need to be compatible with creating and sustaining the conditions conducive to developing virtues and leading virtuous lives. The welfare-promoting aim of nudges is, on this view, understood in terms of promoting one’s ability to live well by supporting the process of virtue-cultivation. While there are several reasons for thinking that these new

behaviour modification techniques are not, and in fact could never be, relevant to the project of cultivating virtues, Niker defends the view that it is possible for nudges to play a role in the policy-led development of virtues. She distinguishes two categories of nudges: ‘discernment-developing nudges’ which have virtue-cultivating properties and ‘automatic-behavioural nudges’, which lack this potential since they merely nudge people to certain types of behaviour.⁷⁴

The discernment-developing nudges are consistent with the Aristotelian account of virtue. They have the ability critically to habituate citizens into better perceiving or discerning the particulars of a situation. Altering how a person sees a situation may influence virtue-acquisition through (self-)conscious habit formation. Niker suggests that by presenting information in particular ways or simplifying objective complexities of decision-making, the state can guide people towards being more aware of and sensitive to the salient aspects of a situation. A need for such a nudge could arise when an individual is uninformed or unable to process information correctly. If choosers lack sufficient knowledge of facts, information-providing nudges might be a welcome corrective.

Niker admits that the virtues cultivated by discernment-developing nudges would fail to meet the conditions necessary for full virtue. But this does not undermine their virtue-cultivation potential, since it is part of being virtuous to be able to perceive, and affectively respond to, the circumstances necessary for specific virtues. Discernment-developing nudges seek to help to cultivate this ability. Niker supports her argument with the following observation made by Sherman:

...action presupposes the discrimination of a situation as requiring a response, reactive emotions that mark that response, and desires and beliefs about how and for the sake of what ends one should act. We misconstrue Aristotle’s notion of action producing character if we isolate the exterior moment of action from the interior cognitive and affective moments which characterize even the beginner’s ethical behaviour.

(Sherman 1989, 178)

Still, with all their potential, discernment-developing nudges might not be sufficient to engage a person’s practical reasoning so as to motivate her to act virtuously. Nudges which do

⁷⁴ C.f., a similar distinction made by Medina and Justo (2016) between automatic nudges and reflective nudges. Automatic nudges do not require conscious engagement and deliberation; they motivate a passive response, and are not easily recognizable by the agents. Reflective nudges require conscious engagement and involve some degree of deliberation; they motivate an active response. Both their means and ends are to some degree noticeable and recognizable by the nudgees.

not require some degree of conscious engagement in the decision making can alter behaviour of individuals. But without providing the reasons for action, they cannot transform people's behaviour in the long run, especially after their pressure is removed. They can at best induce moral behaviour in people, but they cannot make them virtuous without engaging their practical reasoning. Think of a smoker. A mere nudge might be insufficient for him to become a non-smoker. On their own automatic behaviour nudges seem incapable of establishing original (or significantly altering existing) moral or social norms or values (Elster 2007, 104–107). And even discernment-developing nudges will only play a limited role in this process over the course of a person's lifetime. But nudges which encourage reflection might be a good starting point for a long-lasting transformation of one's character.

It seems that nudges could be more effective on this score if aside from helping people discern a situation which requires virtue, they could also provide reasons to hone virtue. The idea behind nudging is that, over time, this learning process should result in the formation of good habits (eating healthy food, saving for retirement, donating body organs). Those, ideally, should not be reliant on institutional support of nudging. These are in principle noble goals, but pursuing them has little to do with formation of virtue, unless over time one also acquires *good reasons for exercising the desirable habits* – virtue being understood as a disposition to choose the right things, at the right time, for the right reason. From the perspective of an aspiring agent who strives to acquire a value or an end that she finds appealing, a nudge that helps her specify her ends by giving reasons for that which is worth pursuing, could be a welcome source of assistance. The fact that persons are highly sensitive to framing effects (i.e. features they would not consider to be relevant if they were rational the way homo economicus is), could be harnessed to help persons become virtuous. This would not merely help them to behave as if they were such but actually also become so.

One way to do it would be by orchestrating original contact with values that will eventually become objects of aspirational pursuit. Such nudges toward values would bring people closer to those values which they may otherwise not (easily or routinely) discover for themselves. This could be done, for example, by way of inspiring people with some role-models or success stories, in order to show that there are other possibilities, other values that have been held, which have helped other persons to do certain admirable things. Instead of manipulating choices through defaults and other designs of choice frameworks, which are not helpful in overcoming causes of unreflective motivation, these nudges could come with relevant information which would prompt an individual to reflect on it.

Providing relevant information (e.g. calories on food labels, warnings on health-deteriorating substances) could also in principle assist virtue cultivation if it were used to

encourage reflection. Sunstein (2015a) too claims that nudging people by providing relevant information should in general be welcome. He agrees that good choice requires knowledge, especially if a given decision needs to be made by an act of active choice (as opposed to tacitly choosing a default). Too my knowledge, however, Sunstein is not interested in facilitating reflection on one's choice as much as making sure that the chooser receives enough information to make the 'correct' choice. He notes—and with good reason—that it is not always necessary (or valuable) for the chooser to 'acquire a stock of knowledge' or 'become an expert' in decisions that in the end lead to default settings (e.g. default settings in cell phones, tablets, health insurance policies, rental car agreements). He argues that where it is not necessary or valuable, a nudge might actually help by saving the chooser unnecessary effort. If acquisition of some information is too costly (financially or otherwise), it might be more effective to put a nudge in place which would help secure the 'desirable outcome'. It would indeed be hard to expect that everyone should learn details of health care provisions before signing up for an insurance policy. The same argument can be used in other potentially nudgeable contexts, as long as there is an agreement about what the 'desirable outcome' of a nudge is.

Sunstein's argument here relies on the instrumental means-end approach to choice making: in order to make good decisions (avoid mistakes), we need relevant information. This take on information-providing nudging does not have much to do with the learning of virtue. If we look at choice from the perspective of the aspiring agent, we see that the idea of a desirable outcome is often undetermined and becomes clearer only over time as one gains experience and learns to exercise virtue. This is especially the case in domains related to volitional aspects of choice, where nudging is becoming increasingly popular. In such domains it would be hard to defend nudges on the basis of their capacity to save the chooser the effort of learning. It is through that very effort that the aspiring agent becomes virtuous. Selection of 'the right amount' of information for the decision maker might therefore mean different things depending on whether the goal is mere convenience in arriving at a pre-determined or 'best' outcome on the one hand, or if it is prompting of reflection. It is the latter goal that would guide the choice architect interested in cultivation of virtue in people.

While it may sometimes be desirable to help people overcome difficulties of choice, that in itself does not make nudges conducive to virtue. Nudges which reduce self-conscious involvement of the agent, and also enable him to forgo the possibility of self-imposed effort to face challenges, betray the Aristotelian understanding of habituation. They do so by implying that right choice is only measurable by its outcomes, and not by the sort of process which leads to its achievement. A successful nudge could mean that one can be steered to do a

virtuous thing routinely and with great regularity, without ever becoming virtuous. Timely payment of one's bills in response to reminders, or forswearing the pleasure of eating cake because one hurriedly selected the proximate and, by design, healthy option, are not necessarily components of virtuous action.

The use of nudges designed to help people consciously exercise virtuous habits could be quite beneficial. White (2013, 121) provides one important reason when he notes that people use the same decision-making processes in many areas of their lives, not just the ones that are monitored by libertarian paternalists. People's improvement of their decision-making in some areas of their lives (e.g. diet or retirement savings), through a conscious change of values, may have analogues in other decisions in their lives. Discernment-developing nudges, along with nudges that provide information or reasons, can facilitate this.

Overall it could be stated that the perspective of a virtue-learning aspiring agent leads to several important conclusions about nudging. There is a danger that by harnessing people's susceptibility to bias nudges may get in the way of one's becoming virtuous. It is hard to see how manipulative nudges could contribute to development of virtue. Even seemingly beneficial and harmless reduction of complexity of a given choice problem by means of nudging can be not so desirable if they deprive the person of the ability to reason about his choice. It seems however that some nudging could be used to facilitate virtue-acquisition by creating new opportunities; or by non-coercively presenting people with values they might have not otherwise appreciated; or by providing relevant information intended to provoke conscious deliberation. In general, it seems, that if nudges are to help the aspiring agent acquire virtue, they should help him exercise the choice muscle rather than putting it to rest.

7.2. Opportunity for Discovery of One's Ends

Encouraging critically-reflected-on decisions is important, but not only for cultivation of virtue. It also fosters people's own learning of what is good for them. We have seen that the proleptic reasoner only learns about his goal as he goes along. He is not in full control of this pursuit. He does not claim full ownership and control of his plan but is open to changing it upon learning something new about himself and the values he is willing to endorse through his action. We have seen in Chapter 5 that situations of conflicts of values and those of uncertainty can be crucial for the aspiring agent's learning about what is good for him. They also help him to discover which ends to pursue.

While the above section discussed the possibility for nudging to facilitate virtue-

acquisition and virtue-cultivation, this section analyses the potential for nudges to facilitate a person's learning about and pursuing his true interests (or change one's view about this in face of a new experience). Recall that for the aspiring agent these are not necessarily known in advance but can be discovered and revisited over time and acknowledged in retrospect as sources of fulfilment. Being an essential contributing factor to an individual's identity and sense of self, this process of self-discovery and preference learning is sometimes argued to be one of the most important things people care about in life.

Insofar as the authors of nudging require successful nudging to need to correspond to a person's own idea of the good rather than to some normative objective imposed externally by a choice architect, expecting a nudge to provide the chooser with an opportunity to develop better understanding of her interests should not be seen as an unwelcome deviation from the original theory of nudging.

Thaler and Sunstein (2006, 259) suggest that people are better at making decisions about subjects on which they are well-informed and experienced (such as choosing between ice cream flavours) than in cases where they are not (such as choosing between financial or savings plans). Personal experience is one of the most effective ways in which people learn about the likely consequences of their actions. Thus, people are liable to have false expectations about the consequences of actions that they take only rarely: 'Unfortunately, some of life's most important decisions do not come with many opportunities to practice. Most students choose a college only once. Outside of Hollywood, most of us choose a spouse, well, not more than two or three times' (2006, 74-75). Yet, the same applies to actions that are taken regularly, but whose worst consequences occur only with low probability or only after a long lapse of time: 'Someone can eat a high-fat diet for years without having any warning signs until the heart attack. When feedback does not work, we may benefit from a nudge' (p. 75).

In their (2008) chapter 'When Do We Need a Nudge?' they show that it is these 'rare, difficult choices are good candidates for nudges'. Nudges are supposed to work in a setting where people have not yet made up their mind (consider the notorious cafeteria case), that is, when – to use the economic jargon again – they lack complete preferences. Nudging the chooser in a circumstance with which he is not familiar, is supposed to make the decision process simpler. It is said to steer the chooser toward a choice that he will not regret.

There is at least one serious problem with this position. Helping a person make a good decision in a situation, in which he or she does not know what the good decision is, relies on the assumption that the choice architect has the superior knowledge of what the person should in fact choose. While what forms the 'correct' decision may be relatively

uncontroversial in the case of matters of health or finance, choice of the right university or the right spouse is usually less obvious (if there is really a 'right' one, rather than a lot of more or less suited options). What distinguishes specific life choices from generic health choices is that practical reason must be exercised in the former, whereas it need not in the latter, since it is pretty much the same for all. Specific life choices are domains in which the chooser herself can best assess the possibilities she has. It would be hard to imagine that someone could do it for her. Or if someone did, it would need to be an intimate friend or relative. Practical reasoning involves one's unique sensibilities: desires, emotions, intellect, all of which are affected by the experience in question. Such judgments often call for personal acquaintance with the object of choice. Even an expert in a given area will usually be underqualified to know which direction a nudge should steer a person. Thus, when the good of a given choice for the individual in question cannot be determined in advance, nudging him or her in a specific direction does not seem to be a worthwhile venture. Could nudging be used to help people learn about what is good for them?

With respect to learning about one's ends, default rules do not seem to be particularly helpful since their architect assumes that one specific option is better than other (e.g. being enrolled in a saving plan rather than not). Defaults do not offer the kind of choice architecture which promotes learning. Sunstein himself admits that defaulting people into certain outcomes does not add anything to their stock of knowledge. And, depending on a particular context, that may be a significant loss of learning opportunity, especially when the use of default or the motivation behind using it is not obvious to the chooser, or when the default option is inferior to a choice one would make if presented with a broader range of possibilities (e.g. wine on a default dinner menu might seem as the best value for money, but upon a closer inspection of the wine list one could discover wine which turns out to be much tastier – to his pallet – and equally satisfying for his pocket). If people learn to rely on default rules set by the government or other choice architects, they may learn to be dependent and inactive (Carlin et al. 2010; de Haan and Linde 2012), or just morally incurious. This creates the danger of a society wherein policy makers are able to trick and manipulate, or wherein policy gets disconnected from an ethical purpose. Similarly, as Schubert and Cordes (2013) argue, there is a danger that nudging people by exploiting (if inadvertently) their unconscious learning mechanisms can 'create preference learning trajectories where individuals are locked in to preferences one can doubt are in the individuals' actual interests'. In such cases nudges that do not reduce freedom of choice in the same period may nonetheless reduce freedom of choice in future periods. They could stand in the way of one's learning of new needs, and of refining preferences, both of which are inherent in self-development and personal self-

transformation.

Active choosing, on the other hand, may be quite helpful in learning about one's options by requiring that a person compares given possibilities. For example, instead of manipulating a customer's choice by placing healthy food in his proximity and hiding unhealthy junk food in less conspicuous locations, or offering him a healthy default menu, member of the serving staff in a university canteen could openly confront (or simply enquire) the customer by asking him whether he prefers more or less (or no) sugar, cream, salt, and other potentially unhealthy ingredients.

While the argument for active choosing as opposed to default rules might be understood in an instrumental sense: based on the assumption that learning refers to acquiring knowledge of the details of a given choice situation, it could also serve a broader context. For an aspiring agent, a choice or decision-making situation is a learning experience of a different kind. It is an exercise in practical reasoning, and it is an opportunity to grow in practical wisdom. It provides contexts in which he learns to deliberate about what is good for himself and acquires more specific knowledge of what things conduce to the good life. Hausman and Welch (2010, 134) endorse active choosing mechanisms on the basis that they can alert an individual that a decision needs to be made in situations where the individual would not have been aware of the possibility for making a decision. Additionally, it might help individuals to recognize these decisions in future contexts.

An objection could be raised here from paternalistically oriented advocates of nudging that active choosing does not guarantee that the choice maker will choose what he will actually value in hindsight. In other words, active choosing does not shield people from mistakes. The legal scholar, Jeffrey Rachlinski (2003, 1214), offers one possible response to this worry when he writes that 'the role of individual learning and adaptation (...) cannot be ignored in assessing the need for paternalism. Simple experience might, in some contexts, be a much better cure for cognitive missteps than adopting a paternalist intervention.' Similarly, Aristotle, Aquinas and Smith each show that experience might be even more effective for non-cognitive missteps, such as failure to grasp one's interests or needs. The above example of the fox who mistakenly perceived the pursuit of unreachable grapes to be in his interest is a case in point. Sometimes a mistake cannot be prevented. But in matters of practical reasoning, whether a given choice constitutes a mistake that one should regret is something that cannot, or cannot easily, be determined in advance—certainly not at the level of a group-based policy. Some things must be learned the hard way. Choosing a college or a career path is one example.

This brings us to another important aspect of practical reasoning: learning from

mistakes. In fact, Sunstein and Thaler (2005, 181) touch on this in their discussion of an argument against paternalism, which they ascribe to John Stuart Mill: people should be allowed to make their own mistakes. Nudges, then, are not serving their function well when they reduce the incentives for individuals to learn by trial and error. They should not be correcting people's mistakes before they can make them (Schnellenbach 2012).⁷⁵ Persons do not learn from unmade mistakes. They also will be less likely to invest time and effort into improving their decision-making if they do not even know they are doing it wrongly. One way we discover that we have made bad decisions is through the consequences of decisions.

Wright and Ginsburg (2012) show the importance of having a chance to learn from one's errors for development of one's personality. Just because they result from cognitive biases and heuristics does not mean that such mistakes are inevitable, although some may be harder to correct than others. If people are allowed to make mistakes, they can then take corrective action. They can recognize mistakes only if such mistakes actually stand in the way of their goals.

On the other hand, Glaeser (2006) shows that even if the individuals recognise their 'errors' and the negative consequences of 'wrong' behaviour, this does not imply that next time they will not again be subject to such an anomaly. Still, the fact that a process of discovery is never complete or error-free does not in itself form a sufficient reason for a systematic interference with the process that nudging certainly represents. It is not implausible to think of nudges which are not designed to prevent mistakes but help the decision maker to find a way out of a mistaken decision. Such would be one that the chooser himself sees as mistaken and not one that is projected to be incorrect by the choice architect. Alternatively, nudges could be designed so as to extend the chooser's opportunity set. Promotion of certain role models combined with relevant social programmes (e.g. encouraging entrepreneurship) could be an example of such an approach.

In summary of the foregoing it can be suggested that nudging can help further the goals of an aspiring agent when it provides him or her with an opportunity to learn more about himself and his true interests. It follows that policy should at least not restrict, but preferably foster, the ability of individuals to engage in the effective learning of new values and objectives. In order to do this, it seems that two general conditions have to be met: first, agents should in principle be left free to acquire new preferences as they see fit. Second, they should be endowed with the maximum possible set of opportunities that does not seriously infringe on the opportunities of others. In this context active choosing is one possible form of

⁷⁵ Carlin et al. (2013), makes similar argument. In contrast to Schnellenbach (2011, 2012), however, they have no fundamental objections to soft paternalism.

desirable nudging, when it is valuable, all things considered, for the chooser to acquire knowledge about available options, or to be allowed to make a risky (and perhaps mistaken) decision. Being allowed to make mistakes can help assume and understand one's responsibility for the mistake. This brings us to the third permissible application of aspiring agent-friendly nudges.

7.3 Responsibility for Choice

The third defining characteristic of an aspiring agent analysed in Chapter 5 is his autonomy in initiating and taking responsibility for the formation of his needs and wants. Cultivating a sense of personal responsibility is an essential element in the process of becoming independent practical reasoners. This section looks at the potential of nudging to serve as a tool in encouraging people to assume responsibility for their actions or making the lines of responsibility clearer. The claim that people are largely responsible for the formation of their characters is not incompatible with the admission that they need help to do so. Nudging could be problematic if it interferes in this process of character formation, yet welcome perhaps if it could facilitate this process.

We saw that in the virtue-ethics tradition one's capacity for active contribution to the shaping of one's preferences has an inherent value for the person. It provides an opportunity for the exercise of one's expressive reason and self-formation. What matters to the aspiring agent is precisely the view of himself as an agent, 'as one who chooses and lives a particular life' (Korsgaard 1989, 126)). There is, in fact, empirical evidence that people do value choice intrinsically regardless of the consequences their choices might achieve (e.g. Bartling et al. 2014). Voting is a common example. When exercised consciously and deliberately, the capacity for agency – one's exercise of practical reason – enables the person to take responsibility (and credit) for her actions, including taking on of the risk involved in these actions.

Responsibility and accountability are important not just in terms of morality and our obligations toward our fellow human beings, but also in relation to ourselves. It would be difficult to improve oneself and one's decision-making (both procedurally and substantively) without accepting one's conscious and responsible participation in it. The aspiring agent's active choices and ability to form his own preferences, taken together, are a way to constitute his identity (Schubert 2015). He wants to be active in overcoming his weaknesses and deficiencies. He seeks to deserve credit for the possible success, or to take responsibility for his mistakes or failures. This sentiment is stressed by Griffin (1986, 67), who notes that '[o]ne

element of agency is deciding for oneself. Even if I constantly made a mess of my life, even if you could do better if you took charge, I would not let you do it. Autonomy has a value of its own'.⁷⁶ Thus, we should be worried about the way non-transparent nudging discourages active choice, for it may infringe individual's capacity to constitute himself as an agent. By discouraging active choice, it may discourage people from engaging in the existential, if effortful, task of self-formation that is at the heart of what economists refer to as preference formation.

When allowed to choose one's own reasons, and therefore to see oneself as an agent, an individual responsive to reasons is also morally responsible for her actions (Fischer and Ravizza 1998).⁷⁷ Since nudge advocates tend to explain choices in terms of psychological and cognitive abilities and limitations, they do not seem to notice the connection between nudges and responsibility for choice. Thus, nudges become particularly hard to accept for the aspiring agent, if choice architects induce behavioural changes that persons are not aware of, and for which they cannot take responsibility.

Valdman (2010) suggests that nudging could be seen as the product of an act of voluntary partial 'outsourcing of self-government' to some external body. He offers a view that there are certain domains where it is plausible to presume that a large majority of citizens benefits from contextual support – from partially 'outsourcing their agency' – in order to minimize the risk of severe distress later in life. Thus, for instance, nudges might be considered to be less of a threat to agency overall when it comes to basic retirement savings or severe health risks. Similarly, Conly (2013) seems to think that most of us would gladly outsource choices associated with quitting smoking or examining nutritional content which we do not enjoy making. On this reasoning, the public body would 'ease us of the responsibility of doing what we would rather not do on our own'. Far from finding such restriction of autonomy objectionable, we might welcome such laws and policies as 'unburdening' us to do the things in life we would rather do. The examples offered by Valdman and Conly, respectively, illustrate an important contrast. For there seems to be a significant difference between outsourcing one's agency in order to avoid severe distress in the future and letting a third party decide for us for the sake of being unburdened of difficult choice-making. Conly's line of argument seems to value convenience in a way that is not as explicit in Valdman's analysis. Convenience may become dangerously 'excessive'

⁷⁶ Buchanan (1999, 257) says: 'Man must bear the responsibilities for his own choices because of his artefactual nature, because man makes his own history.'

⁷⁷ For an opposing view see, for example, Trout (2005) who seeks to minimize the role of personal responsibility in explaining an individual's successes and failures in life (due to cognitive deficiencies and the limits of free will). Instead he attributes people's circumstances to the effects of surrounding institutions and cultures.

(Korsgaard 2009). Korsgaard illustrates this phenomenon with an example of a fictitious student, Jeremy, who, being unable to shape his will, stumbles through life, blindly following accidental impulses. Schubert (2015, 17) warns that nudging may contribute to a situation of excessive convenience resembling Jeremy's predicament, where agents are systematically discouraged from 'shaping their will' and exercising active choice. An agent discouraged over time from the demanding task of choosing may eventually unlearn the skills that enable him to navigate his life.

Bovens (2009, 215) sounds a similar caution when he states that 'short-term success of nudge may be consistent with long-term failure'. He argues that nudging disrespects autonomy if it denies the individual the chance to make mistakes and be held responsible for the consequences. Thus, the long-term effect of nudging may be 'infantilisation' of an agent, which would manifest itself in his 'decreased responsibility in matters regarding his own welfare'. Binder and Lades (2015) confirm these worries in the experimental evidence they quote for unintended negative dynamic effects of nudges—discouragement of critical reflection of decision behaviour and learning (e.g. de Haan and Linde 2012).

From the perspective of the aspiring agent, outsourcing of self-government seems to make more sense when its aim is to prevent or minimise the risk of serious adversity (e.g. not being able to finance treatment of a serious illness) than as a tool of mere convenience. But in order to settle this we need some reference against which we could determine what kind (and/or scope) of responsibility could legitimately be 'outsourced' without compromising one's agency. That is not an easy task. Likewise, what would qualify as an excessive attempt to 'unburden' oneself is far from clear. Conly suggests that when nudges are at work,

[r]ather than regarding people as generally capable of making good choices, we outmaneuver them by appealing to their irrationality, just in more fruitful ways. We concede that people can't generally make good decisions when left to their own devices, and this runs against the basic premise of liberalism, which is that we are basically rational, prudent creatures who may thus, and should thus, direct themselves autonomously.

An important way for nudges to decrease one's responsibility for a given choice is to be found in not allowing the individuals an association between individual behaviour and both personal and social consequences. It thus prevents people from taking responsibility for their choices. It has been suggested that traditional coercive measures to change behaviour distance the individual origin of the behaviour from its results and consequences. For instance, bans on

such-and-such activity or product exclude individual responsibility and preclude exercising self-control (at least within the law). While the main goal of taxes is deterring a choice by economic incentives, the consequences are excluded from the decision-making context, which involves a surrogate harm of diminished money.

Similarly, nudges which are often not personalized (e.g. informational campaigns) usually reinforce individual cognitive biases (Datta and Mullainathan 2014). Thus, it appears that if nudges could be designed to allow the connection between behaviour and consequences, then this could indeed be a promising path to true behavioural change with both short- and long-term positive consequences. If the original lenders, for example, were held to a greater degree of responsibility for excessively risky loans (through a more careful investigation of the trail of securitization), and borrowers were similarly held responsible for their behaviour (through tighter restrictions on bankruptcy and stricter guidelines on refinancing), those on both sides of the mortgage transaction would have incentive to think more carefully about their decisions. Failure to hold people accountable for their decisions reduces their incentive to make careful ones in the first place, and then in the future.

Another possibility for nudging to facilitate responsible decision making could be helping individuals develop critical thinking by simply informing them about the possible psychological processes that are likely to interfere with their deliberation, such as framing effects, so that they are less influenced by factors they are not aware of. Seeking to inform individuals about these issues so as to induce more conscious and responsible action seems more appreciative of the vision of the human as person. In this way, policy makers could move their efforts from trying to determine people's true interests toward creation of conditions under which people can themselves develop reasons for which they act and reflect on which preferences to hold.

In arguing for the possibility to foster responsible decision-making by means of some carefully designed nudging, one should be cautious not to fall into an extreme position. One such holds the individual fully responsible for all aspects of his decision environment. Persons can act through themselves even while being dependent on others and on their circumstances in various ways. Someone or something may give me a reason to act that I, if left to myself, would have never discovered. It seems that this could be a reasonable task for aspiring-agent-friendly nudges. These could provide reasons for action without preventing a person from autonomous action, so long as she understands and accepts these reasons. This way they could be reconciled with the recognition that, despite some dependence on external factors, persons retain capacity to decide for themselves.

What needs to be stressed in this regard is that taking responsibility for one's actions

not only means being held accountable for the effects of one's mistakes or bad choices. It also enables one to take credit for achievements and good decisions. This, as we have seen, is particularly valued by expressively rational agents who—in line with Smith's observations—like to feel praiseworthy. It would be hard to take much credit for a success that has resulted from being nudged, once one knows that fact. On this view autonomy is important because it affirms people as authors of their lives. But, as Thomas Hill (1991, 99) writes,

this is not simply to say that one enjoys being the author and so values living derivatively as a necessary condition of such enjoyment; that puts the focus in the wrong place. Even if it is true that one enjoys living as author, one enjoys it partly because this expresses what one is and wants to be.

The three permissible reasons for nudging developed in the foregoing sections show three possible ways in which nudging could serve as a tool of induced self-formation. Assuming the reality of the aspiring agent does not deny that people make their choices in environments which are often independent of their actions; and that various deliberately designed choice frameworks can make certain potentially mistaken or bad decisions easier and perhaps more beneficial. Given, however, that what seems valuable and worth pursuing for a person is often far from obvious even to him or herself, as the account of the aspiring agent informs us, makes the question of direction of nudging problematic. It seems therefore more advisable and less controversial to leave more autonomy to the choosing agents and instead of nudging them toward a certain goal, creatively use nudging instruments to assist them in formation of their own goals. Nudging intended to help people acquire virtue and learn to become autonomous and responsible choosers can be quite effective in educating a society of mature and creative individuals and groups/associations. Nudging based on manipulation of choice could at best habituate people into decisions which are considered good by the choice architects, but for which people themselves cannot take full responsibility or ownership.

Of all different forms of nudging discussed by Sunstein and Thaler, the best possible nudge, as assessed from the perspective of the aspiring agent who wants to find out for himself what is valuable and worth pursuing, who wants to live his life in accordance with his commitments and take responsibility for his self-formation, seems to require active choosing (Sunstein 2014). Unlike defaults and other 'hidden' nudges, it stimulates the person to reflect on the possibilities at hand in light of her own ends.

Regardless of the domain in which they are used, the most controversial nudges from

the perspective of an aspiring agent appear to be those which impose certain goals or values on the agent, as well as those which take away or dilute personal responsibility for choices. The danger lies particularly in the kind of nudging which aims at determinate social outcomes at the expense of a person constituting herself in her specific environment and context.

Instead of helping persons by nudging them towards what we think they would prefer if they were rational, it might be better to help them develop better understanding of themselves and their idea of the good. Then, let them make their own choices. In practice, this could mean focusing more on the conditions under which people could learn to involve themselves more in decision-making rather than conditions under which they will not make mistakes. The important economic phenomenon of entrepreneurship discussed in the last section of this chapter provides an illustration of how nudging could help create such conditions.

7.4. Nudging Entrepreneurs

The introduction to this chapter suggested that nudges designed with the purpose of facilitating the three characteristics of agency developed in the account of the aspiring agent can be transformative of the person's character. They could potentially expand his freedom of choice, as opposed to nudging designed merely to preserve the existing choice options. Entrepreneurship as an important economic category. At the same time, one common sphere of action in which people find personal and professional fulfilment, it offers an interesting case for the use of aspiring-agent-friendly nudges. It shows that people might not always benefit from the type of security from mistakes and comfort of having their ends (co-)decided for them. That is what some of the 'traditional' nudges promise to provide. For, the risk of mistakes and proleptic reasoning about ends are inherent components of entrepreneurship.

Entrepreneurs seek to bring into existence new products and services for which there has not previously been a market (Venkataraman 1997). They undertake pioneering and innovative efforts and have a special skill of identifying new possibilities. They also have the ability to build new structures for pursuing those possibilities. They are characterised by the desire for risk-taking for its own sake, a willingness to leap into the unknown, a readiness to fail, a need for self-expression, a craving for challenges and creativity (Phelps 2006; 2007; 2013, see in particular chap. 11). Their entrepreneurial projects come into existence thanks to their strong normative commitments: 'belief in the goodness of the venture's purpose, an interest in pursuing activities that reflect the entrepreneur's search for the good life, a vehicle

through which the entrepreneur can ‘be the person that they want to be,’ etc.’ (Dunham 2010, 519). In words of Sarasvathy entrepreneurship is ‘economics with imagination’ and ‘the task of entrepreneurship is to move us from the world we have to live in to the world we want to live in’ (Sarasvathy 2002, 96). Entrepreneurs are then one possible embodiment of the concept of the aspiring agent.

Just as the aspiring agent does, they too operate in the context of uncertainty and openness to learning. A view of entrepreneurial decision making as neoclassical (instrumental) rationality has little to say about these dimensions of entrepreneurship. In the business ethics literature informed by the Aristotelian tradition, however, entrepreneurship is shown to be a value-driven, virtue-based, creative, and dynamic process rather than a calculative, self-interested, and self-benefit-maximizing choice (Dunham 2010, Moberg 2008, Roca 2007).

One could wonder if this is not too idealised of a picture of entrepreneurship. It would not perhaps be implausible to expect that some behaviourally-informed nudge advocates might be inclined to highlight different aspects of entrepreneurial activity: overestimation of abilities, overconfidence, the desire for self-direction and self-determination which may lead to unpredictable consequences (Astebro et al. 2014). If we accepted the position of nudge advocates who want to use nudges to debias people’s excessive optimism, nudges might need to be introduced to prevent entrepreneurs from initiating too-risky ventures. Their creative ideas might be viewed as the product of biases that are reported to be destructive to the individual, and that lead to behaviours that proliferate market choices or disorder decision-making environments for the rest of society.

On the Schumpeterian view, for example, the entrepreneur disrupts the prevailing equilibrium by introducing some new innovation that completely alters existing patterns of supply and demand. Overtly cautious nudge advocates might also say that entrepreneurs are constantly disrupting the stasis of the marketplace, adding and replacing products, services, businesses, jobs, etc., from the market, which simply contributes to the proliferation of choices available to us, thus making new and dynamic cognitive demands on us. Nudges designed to prevent such ‘unwelcome’ consequences would surely curtail the ability of entrepreneurial individuals to act upon their initiatives. This could be problematic for an otherwise entrepreneurial society and for individual entrepreneurs.

Phelps, for example, warns us that economies that have tried to dampen or direct those entrepreneurial impulses – particularly socialist, corporatist, and heavily restricted market economies that shield citizens from risk and change to promote security, equality, or tradition(s) – have tended, on the whole, to experience lower productivity, lower employment

rates, lower workforce participation rates, and generally lower levels of economic wellbeing (2013, chap. 7). Individuals within such economies tend to report lower levels of life satisfaction and job satisfaction, and they report less stimulation, responsibility, and freedom in their workplaces (2013, chap. 8).

There is in fact solid evidence that entrepreneurship contributes to overall happy and healthy societies, and that it is an important element in fostering the possibility of healthy and prosperous societies over the long term. This holds, even after acknowledging short-term discomfort and harms. It can be disruptive (for the company itself and for the economic environment at large), and it never guarantees success. But without it, economic development would lack its engine. The entrepreneur is the agent of innovation, job creation, and wealth production—all things that societies need in order to support economic wellbeing (Decker et al. 2014). Thus, we often hear politicians promising to create more ‘business friendly’ environment to attract foreign and domestic investment. These efforts usually address institutional and cultural adaptations that permit and promote the wildly unpredictable, frantically creative, and energetic goal-seeking of entrepreneurs, who complicate and disrupt the status quo of orderly and predictable decision-making environments. All these incentivising activities are based on the understanding that societies need entrepreneurship and entrepreneurial values in order to secure general wellbeing over the long term. Thus, entrepreneurship illustrates a complex case of a social effort which is ultimately highly valued on a macro-level as beneficial for society, despite obvious risks and difficulties. There is also rich literature which shows evidence for how valuable entrepreneurship can be on a micro scale, namely, how it benefits entrepreneurs themselves and their communities.

The major problem with this otherwise commended activity is the fact that a large part of newly incorporated companies do not succeed. Those which fail cannot therefore provide the benefits they promise to their owners, shareholders and to society. According to recent statistics based on a Harvard Business School Study, as many as 75% of venture-backed start-ups in America fail (Henry 2017). The failure rate of all U.S. companies after five years was over 50 percent, and over 70 percent after 10 years. Based on the analysis of the earlier paragraphs, this data could be met with two types of reaction. The psychologically-informed experts on choice architecture could be expected to claim that this is an extravagant waste of resources which could be put to better use (i.e. serve realistic and tangible goals rather than fulfilling unrealistic dreams). Another type of response could come from those who see entrepreneurship as a valuable opportunity to acquire virtue, learn from mistakes, and take responsibility for one’s decisions. They would not like to see such initiatives inhibited, since it would mean a serious loss of this important opportunity. That does not mean they would

welcome a 75% or greater failure rate just because trying one's hand at a business venture is a valuable learning opportunity. They would, rather, try to find ways to help those who do take up the challenge of starting a risky enterprise stay afloat. I propose that some forms of aspiring-agent-friendly nudging could be useful to this effect.

In an article defending the possibility of practical wisdom as a framework for understanding entrepreneurial endeavours, Dunham (2010) offers a good summary of the praiseworthiness of two well-known and successful companies, The Body Shop and Starbucks, on this score. She notes that it was only through her pursuit of values and virtues that reflected her vision of the good life that the owner of The Body Shop was able to create an entrepreneurial venture that went on to become one of the biggest entrepreneurial success stories of her era. Her distinctive values would ultimately serve as critical differentiators in a crowded marketplace. But they were initially pursued for normative reasons and well before their value as strategic assets could be discerned. Similarly, the owner of Starbucks, guided by values of respect and fairness, departed significantly from retail practices of the time by offering his shop employees extensive training and education, health benefits, and stock options. Once again, what became critical differentiators began as personal, normative values, values that in this case were rooted in his own experiences as a young boy watching his ill father lose his job and health insurance. He took on the expense of providing his employees with those elements he saw as critical to the good life, long before it became clear that these actions would pay off strategically.

What these two examples show is that while the standard entrepreneurship literature based on the rational choice model emphasizes the role of risk and return in the entrepreneur's identification, evaluation, and pursuit of new opportunities, the broader account which conceives the entrepreneur as a practical reasoner (or even broader, as an aspiring agent) suggests that entrepreneurs in fact look to a broader range of personal and ethical values when determining their course of action. Similar insights are confirmed in studies which analyse reasons for business failure (Henry 2017). Interestingly, many of the reasons identified in these studies are due various failures in leadership. Even reasons related to product and costs or competition are ultimately tied to leadership and the leader's ability to manage different levels of his business activity. At the same time, studies of factors of success in business also stress the value of good leadership such as commitment to stay the course and stick with a chosen path, willingness to adjust, but not constantly adjusting, patience and persistence due to the timing mismatch of expectations and reality, willingness to observe, listen and learn (Gompers 2007). These results are a good illustration of how running a successful business requires virtue, openness to learning and taking on of responsibility. Can the framework of the

aspiring agent-friendly nudging be helpful in encouraging this attitude and help more companies stay in business? I believe so. Here is a simple idea for a nudge which could help achieve that.

Each year business owners, not unlike other citizens, file annual tax statements. A simple nudge which could help identify risky cases of vulnerable companies and enable managing their problems at an early stage could be the following. Underneath the section in which the business owner states the amount of his profit/loss made in a given year he is asked the following question: ‘Given your performance last year, would you like to take advantage of a free-of-charge consultation/mentoring session in which you can discuss your major ideas, needs and challenges and work out possible ways to improve your track record in the next twelve months?’.

Entrepreneurship studies show that there is a consistent set of factors that lead startups either to succeed or to fail, and that mentorship plays a crucial role in pivotal moments of many businesses. Entrepreneurs with little or no success in the initial phase of their activity are often discouraged from continuing. Lack of persistence is one of the main reasons for startup failure. It is conceivable that an optional meeting with a good mentor who can help identify areas of possible improvement – both technical, related to the product itself, and personal, aimed at improving the entrepreneur’s leadership skills – result in working out realistic milestones for the next 6-12 months and possibly help the business stay on the course longer than would be otherwise possible.

If entrepreneurship is seen in the broader context of an open-ended activity of an aspiring agent who seeks to pursue the good he finds in running his business, then mere tax incentives and better funding options might not be enough to rescue vulnerable enterprises. Equally if not more useful might be institutional settings which promote the exercise of virtue and responsible risk taking, which most entrepreneurs find increasingly lacking as they are confronted with challenges and difficulties. A nudge like the one proposed above, which requires an active choice of optional mentoring help, could be helpful in cultivating the entrepreneurial virtues. It could also provide a worthwhile opportunity for learning. It could in fact help the entrepreneur see her difficulties as an opportunity for improvement, without absolving her for the ultimate responsibility for the decisions made in an attempt to save her project.

The above example points at an important difference between creating environments that have general characteristics and manipulating environments to create particular outcomes. The paternalism that (at least some of) the nudge advocates endorse aims to direct individuals toward certain actions in aid of a determinate view of their wellbeing. This is

much different than simply advocating general conditions in which individuals might flourish in the ways they see fit and according to their own decisions. No one would disagree with the critics who say that how we frame choices reveals our preferences, morals, priorities; and no one would disagree that framing choices in some general way is inescapable. But given that constant, it does not follow that we have no principled objections available to us to avoid endorsing an increasing number of policies that direct individuals to particular, pre-defined ends.

The external ‘scaffolding’ which Clark (1997) recommends in the form of institutions that support practical reasoning and that focus our attention on better comprehension of our ends, might be one that helps one along the way of an uncertain and perhaps risky undertaking. This could be setting-up a new business venture, starting a career in X, changing jobs, etc., rather than one which is designed to set us on a pre-defined track or prevent the making of mistakes.

Summarizing the last two chapters, it should be recognised that the policy of nudging emerges as a convenient, attractive, and innovative regulatory tool. It recognises and addresses important sources of many present-day social problems, and it endeavours to help ‘busy people trying to cope in a complex world in which they often do not have the time to think deeply about every choice they have to make’. Its overall intention to make people’s lives easier and help them achieve their true interests, can be welcome as a noble, if optimistic objective. Whether it can be achieved by way of reducing bias and preventing mistakes, however, raises more questions than it answers.

If a person is not going to remain permanently using training wheels, the goal of nudging should be to contribute to people’s self-formation in order to make more paternalistic instruments – including paternalistic nudges – unnecessary. Unless the end for which nudging is applied is well-defined and universally agreed upon (safety on roads), instead of starting with a choice architect who wants to direct individuals’ decision-making in a given destination, we should start with individual persons, each of whom is forming the idea of her own good and is responsible for decisions based on her perceived values.

The nature of nudges is such that they need to be designed and applied in the context of a particular case, with an understanding of all relevant institutional and behavioural particularities. That nudges can be adapted to a given situation is probably their greatest advantage. The difficulty remains in dissecting the important factors that matter in a given context. The short case study of entrepreneurship and the nudge it proposes shows that there are unexplored ways in which this might nonetheless be possible and useful.

Overall, it seems more desirable if state intervention in preferences were exercised

explicitly and transparently. This would make the intervention open both to public debate and to judicial scrutiny. These could be welcomed even by those who support liberal democratic principles (i.e. state paternalism). Freedom-extending nudges could help increase the practical wisdom of individual citizens, which to some extent could influence the wisdom of the state.

It might be objected: suggesting that policy makers shift their focus from identifying ‘true’ preferences to the creation of conditions where such preferences stand a better chance of being developed by the person herself, would nonetheless give rise to controversies not dissimilar from those that are already present in the nudge debate. These can be expected especially with regard to specific notions of what forms the good of given conditions, and also what makes them better than others. This kind of question, however, is unavoidable in matters of policy, and one can expect that more satisfactory answers are produced if this fact is openly admitted, rather than being avoided as is often the case with nudge policymaking.

Chapter 8. Conclusions

This thesis showed that modern economics, with its focus on the instrumental homo economicus as an adequate representation of the economic agent, is not fit to represent moral personhood in its account of human choice. This includes economic choice. It cannot explain many important choices which humans make outside of the context of means-end reasoning. It does not recognise that in a variety of nontrivial cases it may be quite rational for a person to be confused, uncertain, or inconsistent. Neoclassical economics understands rationality in a Humean way, as the strictly instrumental pursuit of a known goal. As moral agents, however, we also have the capacity for practical reasoning. It is no less important in informing and guiding our actions. Our reasoning is not merely instrumental means-end calculation. Often it is only *ex post facto* that we can tell whether we are happy with an outcome which we could not have fully envisaged beforehand.

Even behavioural economics with its heavy reliance on psychological research does not adequately represent human action informed by practical reason. Instead, it has acquired a tendency to overexplain human behaviour as a product of cognitive and volitional biases. This thesis has shown that this mode of explanation is particularly unfit to explain choice behaviour originating in aspiration. For it does not adequately explain our conscious participation in the forming of our own choices.

The account of the aspiring agent developed in this thesis is an attempt to fill out the thin anthropology of homo economicus with a more realistic picture of the person. It focuses especially on the moral nature of the human, which is understood here as the capacity to choose between good and evil, and to form one's goals under some aspects of the good. This means, as well as she understands it to be good. This image of the moral person is based on the classical conception of man that, in its important aspects, is shared by Aristotle, Thomas Aquinas, Adam Smith, and even John Stuart Mill.

The reality of the aspiring agent points to the fact that not all of human goals are such that it is possible for people to specify them adequately in advance of and independently of their involvement in various activities through which they try to pursue these goals. Often it is only in or through those activities that one acquires a better idea of what one's goals actually are, what they mean (also to oneself), and how rightly to think about them. In their pursuits of these underspecified goals people are sometimes guided by values they already have, and at other times they act for the sake of acquiring the knowledge of some value that they have discovered only in outline. This is a very basic part of knowledge which guides many of our actions, even if we do not always realise it. We learn it by doing it, by involving ourselves in

some activity or relationship, by gradually gaining more familiarity with it, the way we might learn to be an entrepreneur, or a parent, or even a customer.

The theoretical construct of the aspiring agent departs in notable ways from the traditional model of rational action. It shows that the image of rational homo economicus is insufficient for causal explanation of human choice, and therefore it is an unsatisfactory basis for assessment of human welfare. In contrast to rational choice, which places moral questions outside the perceived bounds of economic action, the reality of the aspiring agent makes morals fully constitutive of a person's decision making. In seeking to construct the good life, the aspiring agent is guided by a desire to act in conformance with virtue. He is also willing to learn from mistakes and to take responsibility for his decisions. The formation of preferences and learning of values emerges from this picture as an integral part of the decision-making process. Not only preferences alone, but also their source and the manner of determination, define the agent and her values.

The concept of an aspiring agent poses a considerable challenge to mainstream economic theory. Instrumental rationality underlying rational choice theory is all about pursuing a known end in a way that most optimally satisfies an agent's given set of preferences. This is essentially a synchronic framework of decision-making. It focuses on values and preferences that are fully formed as well as complete. Much of human action does not, however, come about as the carrying out of explicit intentions, arrived at through explicit weighing of alternatives. Much of human choice is directed at some vaguely understood goal, which becomes more understood, and more of one's own over time; or, as the case may be, it may cease to seem one's own and is then foregone.

The concept of the human person presented in this thesis is not a comprehensive counterweight or even a complement to the notion of homo economicus. It would be difficult to provide an all-encompassing concept of what is involved in human action. One could hesitate in saying whether and in what way such a concept would benefit economics. The aspiring agent offers a more limited and less ambitious – but thereby also a more valuable – attempt to challenge the use of homo economicus as an economic representative of the human person. What is particularly important in this context is that the moral agent that has been presented here is not a counterfactual agent identified with what the individual would be if she were perfectly rational. The aspiring agent is an aspect of the real flesh-and-blood person, aware not only of the fact that her behaviour is driven by many psychological processes, over which she has limited control, but also aware of her capacity to make choices despite these constraints. She is, moreover, willing to learn to exercise that capacity as these limitations allow. Unlike with nudge theorists, this approach is not calling for an idealised person who

might be imagined independently of the very real constraints of ‘natural man’ (Buchanan 1999).

If the theory of human agency spelled out in this thesis is accepted, important implications can be envisaged for neoclassical economic theory of preferences and choice, as well as for policy which rests on this theory. That the instrumentally-guided homo economicus is not an adequate picture of the human being should be openly admitted in economic theory, especially insofar as it informs policy. Economists should acknowledge the special status of human economic agents and make conceptually-sound, empirically-informed, and substantive assumptions about people’s motivational and cognitive characteristics. They should pay more attention to how people perceive choice situations and how they go about specifying or identifying their ends. In that way they will work with a more accurate image of economic choice, and also provide more relevant policy advice.

My focus on the ‘aspiring’ dimension of human choice should be of interest especially to those economists who are committed to identifying the best possible picture of human welfare, but also philosophers and other social scientists who do the same. It may also provide behavioural economists and other behavioural scientists with new ideas for experiments, in so far as they are interested in explaining human action. Understanding preferences as (not yet fully) unknown allows for a much broader range of considerations to be accounted for in explanation of human decision-making. One-dimensional homo economicus models that do not specify how the stipulated preferences are formed, where they come from, and how they change, reinforce the tendency of economists to disregard human motivation. They also stress that actual people do not necessarily feel and think like the players in their models. This tendency results in a weakening of the explanatory power and predictive force of these models.

While this thesis does not deny the usefulness of simplifying assumptions in economics, and while it does not ask anyone to abandon economic rationality altogether, it does show that in problems related to human agency it might be worth going beyond simplified economic rationality. While the homo economicus concept captures something that is true of humans, it is not the whole truth. Satisfaction of preferences is not the final end. Although it is possible usefully to keep the model of the economic agent, who maximizes his utility and has stable preferences, it is also necessary to recognize that this model of standard economics is incomplete. The aim of this thesis is not to suggest that economics should overhaul its technical procedures and devices, but that it should subordinate them to a broader point of view that would inform them and mark out the limits of their applicability.

The account of a moral economic agent that I have developed here sheds new light on

the way in which policy goals can be formed. While with the standard notion of moral agency it is well understood that morally right conduct can promote greater economic output and efficient exchange, the recognition of the capacity of an economic agent to form her own idea of the good and act on it (as an important moral dimension), allows for recognition of interests and values beyond the purely material ones. This, in turn, has implications for conceptualizing and assessing welfare. It then seems crucial that institutions create favourable conditions for choice in which an evolving aspiring agent can change his preferences as a result of a conscious decision, as a consequence of personal development. In an ideal case, one's change in the perception of one's values and preferences should not meet an institutional obstacle which would prevent her from pursuing her choice. The best environment for choice is not necessarily one which allows for maximisation of available alternatives, but one which allows for change and adaptation in response to one's experience as dictated by individual and social circumstances.

Signifying a departure from conventional normative economics, instead of trying to promote given concepts of well-being, better informed welfare economics could focus on the expansion of opportunity and the facilitation of individual activity. It could focus on conditions that support reflection on which preferences to hold. Wellbeing is only partially constituted by economic pursuits of marketable goods. Equally important for welfare is the possibility to pursue one's conception of the good in a way that allows for one's self-formation. In proportion to the development of his individuality, each person becomes more valuable to himself, and is thereby capable of being more valuable to others.

The discussion of nudges above has shown that the concept of the aspiring agent is not only an interesting theoretical construct but that it can also serve as a useful reference point for evaluation of some policy instruments. Nudging could certainly be seen as a pragmatic and efficient way to address important social problems. The thesis has demonstrated, however, that the effects of this approach on welfare and welfare policy would be short-lived, if the approach relies merely on the instrumentalizable dimensions of the person. In the perspective of the argument developed in this thesis, it becomes apparent that government activity designed to alter behaviour may turn out to be ineffective or controversial, when it relies on an oversimplified analysis of the complex picture of what human choice entails, and thus what it takes to influence or change it.

Policies that push for better choices promise to be more effective and less controversial, when informed by a vision of the human person who is conscious of her limitations and over time develops ways to overcome them. This in the end is more gratifying for the person than being pushed into a decision without receiving the chance to face the

challenge unaided. Nudges designed with awareness of that fact stand a better chance of influencing choice without limiting human potential for creativity and innovation.

It is critical not to make too much of this account of the human, of course. It is not as though each decision one makes as consumer, as investor, as citizen, or as member of a society should be backed or preceded by complicated soul-searching. But in many important cases, some of which have been identified in the foregoing chapters, thinking about choice in the framework of aspiration might be helpful.

What should be of particular interest to economists is the idea that people themselves can form their own preferences. Instead of assuming that some higher-level ranking is given, it can be more productive (and closer to reality) to consider that this ranking is formed by the agent who chooses the life that, as an agent, he wants to live. It is equally important to recognise that once-formed preferences do not consist in a closed and stable system of total comparative evaluations, and that people change their preferences for reasons that are outside of the realm of preferences. These reasons, in turn, cannot be easily outweighed by redesigning of a choice framework as nudge supporters would like to see it. In many important cases it is not the hypothetical *homo economicus* that provides the source of normativity, but rather our long-term projects whose value we learn over time.

It should nonetheless be admitted that this is a much more complex approach to the study of choice, one that demands more thorough expertise on the part of the analyst, and very much more analytic effort in determining the specific elements of the context under investigation. That is certainly a considerable disadvantage, especially in comparison with the more straightforward analysis dictated by rational choice theory. There are many choice situations which may not allow for a detailed determination of relevant characteristics necessary to explain (let alone predict) choice. On the other hand, when necessary, the more complex approach is worth exploring, not least because it provides a richer account of what is normatively important in human society. More specifically, it aids at achieving a vision of society that can show *how* individuals count, once that which actually forms their welfare is better understood.

The concept of the aspiring agent seems compatible with many other emerging ideas, and could contribute to their further development, as well as inspiring interesting lines of future inquiry. The building of its own free-standing home, however, requires further work. Insofar as it offers constructive overlap with some of the recently identified research problems, there are possibilities for it to acquire a more independent status.

The aspect of openness of human action to unspecified future outcomes of one's present choices has been taken up in various areas of non-orthodox economics, from

evolutionary theories of entrepreneurship to studies of markets as spontaneous order in which preferences and values are created. The understanding of the economic agent in the context of aspiration and deliberation about one's choices in a perspective of some future goal that is perceived as good by the person, has important implications for broadening the standard notion of rationality.

The concept of the aspiring agent seems to correspond with what Herbert Simon calls 'ecological rationality'. This term has recently become popular in evolutionary economics. It depicts a self-organizing agent who behaves in a homeostatic manner by constantly adapting itself to its environment. Environmental factors operate upon self-organizing agents, but they also act upon those environmental influences, and the agents continually reconstitute themselves as agents over time.

With their expanding set of research tools and concepts, evolutionary economists are also starting to provide their own contribution to the nudge debate, which has similarities to my approach. A particularly interesting direction that their research is now taking is the perception of the creative entrepreneurial discovery of novel emergent properties as one of the motors that drives the economic process. The danger that nudges pose in this context lies in the possibility that they could obstruct the experimentation and error correction which would otherwise lead to improvement over time.

I began by noticing that many problems of modern economics are to be traced back to its departure from the broader discipline of moral sciences. Psychology and neuroscience continue to exert strong influence on economics, and thereby help it defend its position as an independent and amoral discipline. Yet, recent attempts to reclaim economics as a moral science show that studying utility calculations alongside moral considerations as co-determinants of economic behaviour can seriously challenge the discipline's independence (Rona and Zsolnai 2017). The aspiring agent, with his moral motivations dictating his economic behaviour, could certainly support this line of criticism. Their thesis and mine together indicate that the relationship between economic theory and philosophical anthropology deserves a more detailed study. And the status of economics as an amoral science needs to be revisited. Enough proofs have shown that it is not a formal science. But we need to go a step further – in fact, by taking a step back – and admit that economics is a human science; or, better, a moral science as Adam Smith understood it.

Many further areas compatible with the framework of the aspiring agent could be identified. There are questions to be explored, both empirical and philosophical, relating to the possible use of the concept of the aspiring agent in law and education. The contribution of this thesis, however, lies primarily in recognizing and addressing the problem involved in the

fact that economics delegated the study of 'non-economic' features of human action to other disciplines. As a result it now offers no better normative standard than homo economicus for use in the policy it informs. The aspiring agent offers a remedy to the problem of an artificial and unnecessary narrowing of economic anthropology. While further development of the concept is needed, the steps taken in this thesis provide a significant and dynamic foundation upon which other necessary extensions can be built.

Bibliography

- Ainslie, G. (1992) *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*, Cambridge: Cambridge University Press.
- Akerlof, G.A., Dickens, W.T. (1982) 'The Economic Consequences of Cognitive Dissonance', *The American Economic Review*, 72(3), 307-319.
- Akerlof, G.A. (1983) 'Loyalty filters', *The American Economic Review*, 73(1), 54-63.
- Akerlof, G.A. (1984) *An Economic Theorist's Books of Tales*, Cambridge: Cambridge University Press.
- Akerlof, G., Kranton, R.E. (2000) 'Economics and Identity', *The Quarterly Journal of Economics*, 115(3), 715-753.
- Alvey, J.E. (1997) 'Postmodern 'readings' of Adam Smith's 'discourse'', *History of Economics Review*, 26, 155-166.
- Amir, O. and Lobel, O. (2008) 'Stumble, Predict, Nudge: How Behavioral Economics Informs Law and Policy', *Columbia Law Review*, 108(8), 2098-2137.
- Anderson, J. (2010) 'Review of Nudge: Improving Decisions about Health, Wealth, and Happiness, Thaler R.H. and Sunstein C.R.', *Economics and Philosophy*, 26, 369-406.
- Anspach, R., (1972) 'The Implications of the Theory of Moral Sentiments for Adam Smith's Economic Thought', *History of Political Economy*, 4(1), 176-206.
- Aquinas, Saint Thomas (1945) *Basic Writings of Saint Thomas Aquinas*, New York: Random House.
- Aquinas, Saint Thomas, (1947) [1270] *Summa Theologicae*, translated by the Fathers of the English Dominican Province, Westminster, MD: Christian Classics.
- Aquinas, Saint Thomas, (1964) *Commentary on the Nichomachean Ethics*, trans. Litzinger, Chicago: Henry Regnery Company.

Archer, M.S. (2003) *Structure, Agency and the Internal Conversation*, Cambridge: Cambridge University Press.

Ariely, D., Loewenstein, G., Prelec, D. (2006) 'Tom Sawyer and the Construction of Value', *Journal of Economic Behavior & Organization*, 60(1), 1-10.

Aristotle (1894) *Ethica Nicomachea*, Bywater, T. (ed.), Oxford and New York: Oxford University Press.

Aristotle (1976) *Nicomachean Ethics*, Harmondsworth: Penguin.

Aristotle (1980) *Nicomachean Ethics*, Oxford: Oxford University Press.

Aristotle (1988) *Politics*, Cambridge: Cambridge University Press.

Aristotle (2010) *Rhetoric*, New York: Cosimo Classics.

Armour, L. (1976) 'Smith, Morality, and the Bankers', *Review of Social Economy*, 34(3), 359-372.

Arrow, K.J. (1951) *Social Choice and Individual Values*, Yale University Press.

Ashraf, N., Camerer, C.F., Loewenstein, G. (2005) 'Adam Smith, Behavioral Economist', *Journal of Economic Perspectives*, 19(3), 131–145.

Åstebro, T., Herz, H., Nanda, R., & Weber, R. (2014) 'Seeking the roots of entrepreneurship: Insights from behavioral economics', *Journal of Economic Perspectives*, 28(3), 49–70.

Augustine, Saint (1955) *De civitate Dei*, Corpus christianorum, series latina, 48. Turnhout: Brepols.

Augustine, Saint (1984) *City of God*, Harmondsworth: Penguin.

Bacharach, M. (2006) *Beyond Individual Choice: Teams and Frames in Game Theory*, Princeton, NJ: Princeton University Press.

Bartling, B., Fehr, E., Herz, H. (2014) 'The intrinsic value of decision rights', *Econometrica*, 82(6), 2005–2039.

Becker, G.S. (1976) *The Economic Approach to Human Behavior*, Chicago: University of

Chicago Press.

Becker, G.S. (1993) *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, New York: Columbia University Press.

Benn, S.I. (1976) 'Freedom, Autonomy and the Concept of a Person', *Proceedings of the Aristotelian Society* (76), 109-130.

Benn, S.I. (1988) *A Theory of Freedom*, Cambridge, New York, Melbourne, Sydney: Cambridge University Press.

Berg, N. and Gigerenzer, G. (2010) 'As-If Behavioral Economics: Neoclassical Economics in Disguise?', *History of Economic Ideas* 18(1), 133-165.

Berg, N., Hoffrage, U. (2008) 'Rational Ignoring with Unbounded Cognitive Capacity', *Journal of Economic Psychology*, 29(6), 792-809.

Berg, N., Lein, D. (2005) 'Does Society Benefit from Investor Overconfidence in the Ability of Financial Market Experts?', *Journal of Economic Behavior & Organization*, 58(1), 95-116.

Berg, N. (2003) 'Normative Behavioral Economics', *The Journal of Socio-Economics*, 32(4), 411-427.

Berg, N., Biele, G. and Gigerenzer, G. (2010) 'Does Consistency Predict Accuracy of Beliefs? Economists Surveyed about PSA'. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1692320 (Accessed: 3 March 2017)

Berg, N., Eckel, C.C., Johnson, C.A. (2008) 'Inconsistency Pays? Time-Inconsistent Subjects and EU Violators Earn More', working paper. Available at: <https://www.semanticscholar.org/paper/Inconsistency-Pays-%3F-%3A-Time-inconsistent-subjects-Berg-Eckel/256ff66e33d4dd03a5aba8a0b0806af539be037c> (Accessed: 6 December 2016).

Berlin, I. (1969) *Four Essays on Liberty*, Oxford: Oxford University Press.

Bernheim, B.D., Rangel, A. (2007) 'Toward Choice-Theoretic Foundations for Behavioral Welfare Economics', *The American Economic Review*, 97(2), 464-470.

Bernheim (2008), Behavioral Welfare Economics, NBER Working Paper No. 14622. Available

at: <http://www.nber.org/papers/w14622> (Accessed: 17 May 2017).

Bernheim, B.D., Rangel, A. (2009) 'Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics', *The Quarterly Journal of Economics*, 124(1), 51-104.

Berofsky, B. (1995) *Liberation from self: A theory of personal autonomy*, New York: Cambridge University Press.

Berry, C.J. (2010) 'Adam Smith's Moral Economy', *The Kyoto Economic Review*, 79(1), 2-15.

Berry, C.J. (2012) 'Adam Smith's Science of Human Nature', *History of Political Economy* 44(3), 471-492.

Bénabou, R., Tirole, J. (2002) 'Self-Confidence and Personal Motivation', *Quarterly Journal of Economics* 117(3), 871-915.

Billet, L. (1976) 'The Just Economy: The Moral Basis of the Wealth of Nations', *Review of Social Economy*, 34(3), 295-316.

Binder, M., Lades, L.K. (2015) 'Autonomy-Enhancing Paternalism', *Kyklos*, 68(1), 3-27.

Binder, M. (2014) 'Should Evolutionary Economists Embrace Libertarian Paternalism?' *Journal of Evolutionary Economics*, 24(3), 515-539.

Bovens, L. (2009) 'The Ethics of Nudge', in Grüne-Yanoff, T. and Hansson, S. (eds.) *Preference Change: Approaches from Philosophy, Economics and Psychology*. Berlin and New York: Springer, 207-219.

Bowie, N.E. (1991) 'Challenging the Egoistic Paradigm', *Business Ethics Quarterly*, 1-21.

Bowles, S., Gintis, H. (2011) *A Cooperative Species: Human Reciprocity and its Evolution*, Princeton, NJ: Princeton University Press.

Bradley, R. (2009) 'Revising Incomplete Attitudes', *Synthese*, 171(2), 235-256.

Brennan, G., Lomasky, L. (1983) 'Institutional aspects of "merit goods" analysis', *FinanzArchiv/Public Finance Analysis*, 41(2), 183-206.

- Broadie, A. (2006) 'Sympathy and the impartial spectator', in Haakonssen, K. (ed.) *The Cambridge Companion to Adam Smith*. Cambridge: Cambridge University Press.
- Brock, D. (1988) 'Paternalism and Autonomy', *Ethics* 98, 550–565.
- Brown, V. (1991) 'Signifying Voices: Reading the "Adam Smith Problem"', *Economics & Philosophy*, 7(2), 187-220.
- Brown, V. (1994) *Adam Smith's Discourse: Canonicity, Commerce, and Conscience*, London: Routledge.
- Brown, V. (1997) 'Dialogism, the Gaze, and the Emergence of Economic Discourse', *New Literary History*, 28(4), 697-710.
- Buchanan, J.M. (1979) 'Natural and Artifactual Man', in Buchanan, *The Collected Works of James M. Buchanan*, vol.1. Indianapolis: Liberty Fund, 246-260.
- Buchanan, J.M. (1982) 'Order Defined in the Process of its Emergence', *Literature of Liberty*, 5(5), 7-58.
- Burnyeat, M.F. (1980) 'Aristotle on Learning to be Good' in Amélie Oksenberg Rorty (ed.) *Essays on Aristotle's Ethics*. Berkeley, Los Angeles, London: University of California Press, 69-92.
- Buss, S. (2012) 'Autonomous Action: Self-Determination in the Passive Mode', *Ethics*, 122(4), 647-691.
- Caldwell, B. (2001) 'Hodgson on Hayek: A Critique', *Cambridge Journal of Economics*, 25(4), 539-553.
- Callard, A. (2018) *Aspiration: The Agency of Becoming*, New York: Oxford University Press.
- Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., Rabin, M. (2003) 'Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism"', *University of Pennsylvania Law Review*, 151(3), 1211-1254.
- Camerer, C.F., Loewenstein, G., Prelec, D. (2005) 'Neuroeconomics: How Neuroscience Can Inform Economics', *Journal of Economic Literature*, 43, 9-64.

- Camerer, C.F., Fehr, E. (2006) 'When Does "Economic Man" Dominate Social Behavior?', *Science*, 311(5757), 47-52.
- Camerer, C.F. (2006) 'Wanting, Liking, and Learning: Neuroscience and Paternalism', *The University of Chicago Law Review*, 73(1), 87-110.
- Camerer, C.F., Loewenstein, G. (2003) 'Behavioral Economics: Past, Present, Future', in Camerer, C.F., Loewenstein, G., Rabin, M. (eds.) *Advances in behavioral economics. Roundtable series in behavioral economics*. Princeton, NJ: Princeton University Press, 1-61.
- Graham, J., Camerer, C.F., Loewenstein, G. (2005) 'Adam Smith, Behavioral Economist', *The Journal of Economic Perspectives*, 19(3), 131-145.
- Campbell, T. D. (1971) *Adam Smith's Science of Morals*, London: Routledge.
- Campbell, W.F. (2004) 'Adam Smith's Theory of Justice, Prudence and Beneficence', in Wood, J.C. (ed.) *Adam Smith: Critical Assessments*, London and New York: Routledge.
- Campbell, W.F. (1987) 'The Old Art of Political Economy', in Lowry, S.T. (ed.) *Pre-Classical Economic Thought*, Boston, Dordrecht, Lancaster: Kluwer Academic Publishers, 31-42.
- Camus, A. (1942) *The Myth of Sisyphus*, New York: Vintage International.
- Card, C. (2002) *The Atrocity Paradigm*, New York: Oxford
- Carlin, B.I., Gervais, S., Manso, G. (2010) 'Libertarian Paternalism, Information Sharing and Financial Decision-Making'. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1570158 (Accessed: 17 June 2017).
- Carlin, B.I., Gervais, S. and Manso, G. (2013) 'Libertarian Paternalism, Information Production, and Financial Decision Making', *The Review of Financial Studies*, 26(9), 2204-2228.
- Carrasco, M.A. (2012) 'Adam Smith: Self-Command, Practical Reason and Deontological Insights', *British Journal for the History of Philosophy*, 20(2), 391-414.
- Casson, M. (1982) *The Entrepreneur: An Economic Theory*, Lanham, MD: Rowman & Littlefield.

Choi, Y.B. (1990) 'Smith's View on Human Nature: A Problem in the Interpretation of The Wealth of Nations and The Theory of Moral Sentiments', *Review of Social Economy*, 48(3), 288-302.

Christman, J. (2009) *The Politics of Persons: Individual Autonomy and Socio-historical Selves*, Cambridge University Press.

Christman, John, Anderson, J. (2005) *Autonomy and the Challenges to Liberalism: New Essays*, New York: Cambridge University Press.

Churchland, P. (1996) *The Engine of Reason, the Seat of the Soul*, Cambridge, MA: MIT Press.

Clark, A. (1997) 'Economic reason: The interplay of individual learning and external structure' in Drobak, J. and Nye, J. (eds.) *The frontiers of the new institutional economics*, Cambridge: Academic Press, 269-290.

Coase, R.H. (1976) 'Adam Smith's View of Man', *The Journal of Law and Economics*, 19(3), 529-546.

Conly, S. (2013) *Against Autonomy. Justifying Coercive Paternalism*, Cambridge: Cambridge University Press.

Cowen, T. (1991) 'Self-Constraint and Self-Liberation', *Ethics* (101), 360-373.

Crawford, M.B. (2015) *The world beyond your head: On becoming an individual in an age of distraction*, New York: Farrar, Straus and Giroux.

Crosby, J.F. (1996) *The Selfhood of the Human Person*, Washington: The Catholic University of America Press.

Crossley, N. (2004) 'Fat is a sociological issue: Obesity rates in late modern, "body-conscious" societies', *Social Theory & Health*, 2(3), 222-253.

Danielson, P.A. (1998) *Modeling Rationality, Morality, and Evolution*, Oxford University Press.

Danner, P.L. (2002) *The Economic Person: Acting and Analyzing*, Lanham, Boulder, New York and Oxford: Rowman & Littlefield.

Darwall, S. (2006) 'The Value of Autonomy and Autonomy of the Will', *Ethics* 116 (2), 263-284.

Datta, S., Mullainathan, S. (2014) 'Behavioral Design: A New Approach to Development Policy', *Review of Income and Wealth*, 60(1), 7-35.

Davidson, D. (1980) 'How is weakness of the will possible?' in *Essays on actions & events*, Oxford: Clarendon Press.

Davis J.B. (2003) *The Theory of the Individual in Economics*, London and New York: Routledge.

Davis, J.B. (2006) 'The Normative Significance of the Individual in Economics' in Clary, B.J., Dolfsma, W., Figart, D.M. (eds) *Ethics and the Market: Insights from Social Economics*, London and New York: Routledge, 69-83.

Davis, J.B. (2010) 'The Individual in Recent Economics: Internalist and Externalist Conceptions' in Parisi, D.F., Solari, S. (eds.) *Humanism and religion in the history of economic thought: Selected papers from the 10th Aispe conference* (Collana di studi e ricerche dell'Associazione Italiana per la Storia del Pensiero Economico, 6), Milano: Franco Angeli.

Davis, J.B. (2011) *Individuals and Identity in Economics*, Cambridge: Cambridge University Press.

Davis, J.B. (2015) 'Bounded Rationality and Bounded Individuality' in Fiorito, L., Scheall, S., Suprinyak, C.E. (eds.) *A Research Annual (Research in the History of Economic Thought and Methodology*, Volume 33) Emerald Group Publishing Limited, 75-93.

Davis, J.B., (2004) 'The Theory of the Embedded Individual in Heterodox Economics' in Lewis, P. (ed.) *Transforming Economics: Perspectives on the Critical Realist Project*, London: Routledge, 132-151.

Decock, Wim (2013) *Theologians and Contract Law: The Moral Transformation of the Ius Commune*, Leiden: Martinus Nijhoff Publishers.

Decker, R., Haltiwanger, J., Jarmin, R., Miranda, J. (2014) 'The role of entrepreneurship in US job creation and economic dynamism', *Journal of Economic Perspectives*, 28(3), 3–24.

- De Haan, T., Linde, J. (2012) ‘“Good Nudge Lullaby’: Choice Architecture and Default Bias Reinforcement’, *The Economic Journal*, 128 (610), 1180-1206.
- De Roover, R. (1955) ‘Scholastic Economics: Survival and Lasting Influence from the Sixteenth Century to Adam Smith’, *The Quarterly Journal of Economics*, 161-190.
- Debreu, G. (1959) *Theory of value: An axiomatic analysis of economic equilibrium*, Yale University Press.
- Debreu, G. (1986) ‘Theoretic models: mathematical form and economic content’, *Econometrica: Journal of the Econometric Society*, 54(6), 1259-1270.
- De-Juan, O., Monsalve, F. (2006) ‘Morally ruled behaviour: The neglected contribution of Scholasticism’, *The European Journal of the History of Economic Thought*, 13(1), 99-112.
- Dekker, E., Klamer, A. (2016) ‘About Doing the Right Thing as an Academic Economist’ in DeMartino, G., McCloskey, D. (eds.) *The Oxford Handbook of Professional Economic Ethics*, New York: Oxford University Press.
- Dempsey, B.W. (1948) *Interest and usury*, London: D. Dobson.
- Dennett, D. (1991), *Consciousness Explained*, Boston, MA: Little, Brown and Co.
- Desai, A.C. (2011) ‘Libertarian paternalism, externalities, and the “spirit of liberty”’: How Thaler and Sunstein are nudging us toward an “overlapping consensus”’, *Law & Social Inquiry*, 36(1), 263-295.
- DesJardins, J.R. (2006) *Environmental Ethics. An Introduction to Environmental Philosophy*, Belmont: Thomson Wadsworth.
- Dewey, J. (2008) *The Later Works 1925-1953*, Vol. 3: 1927-1928, Carbondale: Southern Illinois University.
- Dietrich, F., List, C. (2013a) A Reason-Based Theory of Rational Choice, *Nous*, 47(1), 104-134.
- Dietrich, F. and List, C. (2013b) ‘Where do preferences come from?’, *International Journal of Game Theory*, 42(3), 613-637.

- Dolan, P. (2014) *Happiness by Design*, New York: Avery Pub. Group.
- Dobson, J. (1991) 'Reconciling financial economics and business ethics', *Business & Professional Ethics Journal*, 23-42.
- Dopfer, K. (2005) *The evolutionary foundations of economics*, New York: Cambridge University Press.
- Dunham, L.C. (2010) 'From Rational to Wise Action: Recasting Our Theories of Entrepreneurship', *Journal of Business Ethics* 92(4), 513-530.
- Dworkin, G. (1988) *The theory and practice of autonomy*, New York: Cambridge University Press.
- Elder-Vass, D. (2007) 'Reconciling Archer and Bourdieu in an Emergentist Theory of Action', *Sociological Theory*, 25(4), 325-346.
- Elmendorf, J.J. (1892) *Elements of Moral Theology*, New York: James Pott & Company.
- Elster, J. (1983) *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge: Cambridge University Press.
- Elster, J. (1985) 'Weakness of will and the free-rider problem', *Economics & Philosophy* 1(2), 231-265.
- Elster, J. (1987) *The multiple self*, Cambridge: Cambridge University Press.
- Elster, J. (1979) *Ulysses and the sirens: Studies in rationality and irrationality*, Cambridge: Cambridge University Press.
- Elster, J. (2007) *Explaining Social Behaviour: More Nuts and Bolts for the Social Sciences*, Cambridge: Cambridge University Press
- Etzioni, A. (1986) 'The Case for a Multiple-Utility Conception', *Economics & Philosophy*, 2(2), 159-184.
- Etzioni, A. (1988) *Moral dimension: Toward a new economics*, New York: The Free Press.
- Evans, C.S. (2002) *Preserving the person: A look at the human sciences*, Vancouver: Regent

College Publishing.

Evensky, J. (1987) 'The two voices of Adam Smith: moral philosopher and social critic, *History of Political Economy*, 19(3), 447-468.

Evensky, J. (2001) 'Adam Smith's lost legacy', *Southern Economic Journal*, 67(3), 497-517.

Evensky, J. (2005) 'Adam Smith's Theory of Moral Sentiments: On morals and why they matter to a liberal society of free people and free markets', *The Journal of Economic Perspectives*, 19(3), 109-130.

Evensky, J. (2007) *Adam Smith's moral philosophy: a historical and contemporary perspective on markets, law, ethics, and culture*, Cambridge: Cambridge University Press.

Fateh-Moghadam, B., Gutmann, T. (2014) 'Governing [through] autonomy. The moral and legal limits of "soft paternalism"', *Ethical Theory and Moral Practice*, 17(3), 383-397.

Fehr, E. and Camerer, C.F. (2007) 'Social neuroeconomics: the neural circuitry of social preferences', *Trends in cognitive sciences*, 11(10), 419-427.

Fehr, E., Fischbacher, U. (2002) 'Why social preferences matter—the impact of non-selfish motives on competition, cooperation and incentives', *The Economic Journal*, 112(478).

Finley, M.I. (1970) 'Aristotle and economic analysis', *Past & Present*, 47, 3-25.

Finley, M.I. (1999) *The Ancient Economy*, Los Angeles: University of California Press.

Finnis, J. (1990) 'Concluding Reflections', *Cleveland State Law Review*, 38, 231.

Finnis, J. (2011) *Natural law and natural rights*, Oxford: Oxford University Press.

Fischer, J.M., Ravizza, M. (1998) 'Morally responsible people without freedom' in Fisher, J.M., Ravizza, M. (eds.) *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Fischer, M., Lotz, S. (2014) 'Is Soft Paternalism Ethically Legitimate? - The Relevance of Psychological Processes for the Assessment of Nudge-Based Policies', Cologne Graduate School in Management, Economics and Social Sciences Working Paper. Available at: http://www.cgs.uni-koeln.de/fileadmin/wiso_fak/cgs/pdf/working_paper/cgswp_05-02.pdf

(Accessed: 8 June 2017).

Fleetwood, S. (2008) 'Structure, institution, agency, habit, and reflexive deliberation', *Journal of Institutional Economics*, 4(2), 183-203.

Fleischacker, S.F. (1999) *A Third Concept of Liberty*, Princeton, NJ: Princeton University Press.

Fleischacker, S. (2004) *On Adam Smith's Wealth of Nations: A Philosophical Companion*, Princeton, NJ: Princeton University Press.

Fleischacker, S. (2011) 'True to ourselves?', *The Adam Smith Review*, 6.

Force, P. (2006) 'First principles in translation: The axiom of self-interest from Adam Smith to Jean-Baptiste Say', *History of Political Economy*, 38(2), 319-338.

Frankfurt, H. (1971) 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, 68(1), 5-20.

Frankfurt, H., (1982) 'The importance of what we care about', *Synthese*, 53(2), 257-272.

Frankfurt, H. (1988) 'Identification and Wholeheartedness', in *The Importance of What We Care About*, Cambridge: Cambridge University Press, 159-176.

Frey, B. (1997) *Not just for money. A Personal Theory of Motivation*, Cheltenham: Edward Elgar.

Frey, B.S., Benz, M., Stutzer, A. (2004) 'Introducing procedural utility: Not only what, but also how matters', *Journal of Institutional and Theoretical Economics*, 160(3), 377-401.

Friedman, M. (1953) *Essays in positive economics*, Chicago: University of Chicago Press.

Fudulu, P. (2014) 'An Anti-Entropic Evolutionary Perspective on Human Ends Formation'. Available at: https://www.researchgate.net/publication/260835541_An_Anti-Entropic_Evolutionary_Perspective_on_Human_Ends_Formation (Accessed: 5 June 2017).

Garnett, R. (2017) 'Cooperation and Assistance in the Great Society: Adam Smith's Oikeiōsis Revisited', *The Adam Smith Review*, 10.

- Gilovich, T., Griffin, D., Kahneman, D. (2002) *Heuristics and biases: The psychology of intuitive judgment*, Cambridge: Cambridge University Press.
- Giovanola, B. (2009) 'Re-thinking the anthropological and ethical foundation of economics and business: Human richness and capabilities enhancement', *Journal of Business Ethics*, 88(3), 431-444.
- Glaeser, E.L. (2006) 'Paternalism and psychology', *Regulation*, 29(2), 32-38.
- Gold, N., Sugden, R. (2006) *Beyond Individual Choice: Teams and Frames in Game Theory*, Princeton, NJ: Princeton University Press.
- Gold, N., Sugden, R. (2007) *Theories of team agency*, Oxford: Oxford University Press.
- Gompers P. (2007) *Startup Genome Report 2006*. Available at: <https://startupgenome.com/report2006/> (Accessed: 30 June 2018).
- Goodin R (2000) 'Democratic deliberation within', *Philosophy & Public Affairs*, 29(1), 81–109.
- Goodwin, T. (2012) 'Why we should reject 'nudge'', *Politics*, 32(2), 85-92.
- Gordley, J. (1991) *The philosophical origins of modern contract doctrine*, Oxford: Clarendon Press.
- Gordon, B. (1975) *Economic Analysis before Adam Smith: Hesiod to Lessius*, London: Palgrave Macmillan.
- Grampp, W.D. (1948) 'Adam Smith and the economic man', *Journal of Political Economy*, 56(4), 315-336.
- Grice-Hutchinson, M., Moss, L.S. (1993) *Economic Thought in Spain: selected essays*, Cheltenham: Edward Elgar Publishing.
- Griffin, J. (1986) *Well-being: Its meaning, measurement, and moral importance*, Oxford: Clarendon Paperbacks.
- Griswold, C.L. (1999) *Adam Smith and the Virtues of Enlightenment*, Cambridge: Cambridge University Press.

- Grüne-Yanoff, T. (2012) 'Old wine in new casks: libertarian paternalism still violates liberal principles', *Social Choice and Welfare*, 38(4), 635-645.
- Gui, B., Sugden, R. (2005) *Economics and social interaction: Accounting for interpersonal relations*, Cambridge: Cambridge University Press.
- Haakonssen, K. (1989) *The science of a legislator: the natural jurisprudence of David Hume and Adam Smith*, Cambridge: Cambridge University Press.
- Haltzman, J., Noell, E.S. (2012) *Reckoning with Markets: The Role of Moral Reflection in Economics*, Oxford: Oxford University Press.
- Hammond, P.J. (1976) 'Changing tastes and coherent dynamic choice', *The Review of Economic Studies*, 43(1), 159-173.
- Hamouda, O.F., Price, B.B. (1997) 'The justice of the just price', *Journal of the History of Economic Thought*, 4(2), 191-216.
- Hampton, J.E. (1998) *The Authority of Reason*, Cambridge: Cambridge University Press.
- Hanemann, M. (2007) 'Comments on Sugden, "Cost-Benefit Analysis as Market Simulation" or Economics as if People Were Human Beings', paper presented to the Conference on Frontiers of Environmental Economics Washington DC, February 26-27.
- Hanley, R.P. (2009) *Adam Smith and the character of virtue*, Cambridge: Cambridge University Press.
- Hargreaves Heap, S.P. (1989) *Rationality in Economics*, Oxford: Basil Blackwell.
- Hargreaves Heap, S.P. (1999) 'Social Capital and the Economy,' in M. Setterfield (ed.) *Growth, Employment and Inflation. Essays in Honor of John Cornwall*. Basingstoke, New York: Macmillan/St Martin's.
- Hargreaves-Heap, S.P. (2013) 'What is the Meaning of Behavioural Economics?', *Cambridge Journal of Economics*, 37(5), 985-1000.
- Harsanyi, J.C. (1982) 'Morality and the Theory of Rational Behavior' in Sen, A., Williams, B. (eds.) *Utilitarianism and Beyond*. Cambridge: Cambridge University Press, 39-62

Hausman, D.M., Welch, B. (2010) 'Debate: To nudge or not to nudge', *Journal of Political Philosophy* 18(1), 123-136.

Hausman, D.M. (2011) *Preference, value, choice, and welfare*, New York: Cambridge University Press.

Hausman, D.M. (2016) 'On the econ within', *Journal of Economic Methodology*, 23(1), 26-32.

Hayek, F.A. (1945) 'The use of knowledge in society', *The American Economic Review*, 35(4), 519-530.

Hayek, F.A. (1961) 'The Non Sequitur of the 'Dependence Effect'', *Southern Economic Journal*, 27, 346-348.

Heath, E. (1995) 'The commerce of sympathy: Adam Smith on the emergence of morals', *Journal of the History of Philosophy*, 33(3), 447-466.

Hédoin, C. (2015) 'From Utilitarianism to Paternalism: When Behavioral Economics meets Moral Philosophy', *Revue de philosophie économique*, 16, 73-106.

Heilbroner, R.L. (1982) 'The socialization of the individual in Adam Smith', *History of Political Economy*, 14(3), 427-439.

Henry, P. (2007) *Why Some Startups Succeed (and Why Most Fail)*. Available at: <https://www.entrepreneur.com/article/288769> (Accessed: 30 June 2018).

Herzog, L. (2013) *Inventing the market: Smith, Hegel, and political theory*, Oxford: Oxford University Press.

Heyne, P. (2008) 'Are Economists Basically Immoral?' and Other Essays on Economics' in Heyne, P., Brennan, G., Waterman, A. (eds.) *Ethics and Religion*. Indianapolis, IN: Liberty Fund.

Hicks, J.R. (1986) *A revision of demand theory*, Oxford: Oxford University Press.

Hill Jr, T.E. (1991) *Autonomy and self-respect*, Cambridge: Cambridge University Press.

Hill, L. (2001) 'The hidden theology of Adam Smith', *European Journal of the History of Economic Thought*, 8(1), 1-29.

Hirschman, A.O. (1982) *Shifting involvements: Private interest and public action*, Princeton, NJ: Princeton University Press.

Hirschman, A.O. (1984) 'Against parsimony: three easy ways of complicating economic discourse', *Psychological Foundations*, 74(2), 89-96.

Hirschman, Albert O. (1985) 'Against parsimony: three easy ways of complicating economic discourse', *Economics and Philosophy*, 1, 7-21.

Hirschman, A.O. (2013) *The passions and the interests: Political arguments for capitalism before its triumph*, Princeton, NJ: Princeton University Press.

Hodgson, G.M. (2000) *From micro to macro: the concept of emergence and the role of institutions*, *Institutions and the Role of the State*, Cheltenham, UK, Northampton, MA, USA: Edward Elgar, 103-126.

Hodgson, G.M. (2004) 'Hayekian evolution reconsidered: a response to Caldwell', *Cambridge Journal of Economics*, 28(2), 291-300.

Hodgson, G.M. (2004) *The evolution of institutional economics*, London and New York: Routledge.

Hodgson, G.M. (2012) *From pleasure machines to moral communities: An evolutionary economics without homo economicus*, University of Chicago Press.

Hodgson, G.M. (2014) 'The evolution of morality and the end of economic man', *Journal of Evolutionary Economics*, 24(1), 83-106.

Hollander, S. (1977) 'Adam Smith and the self-interest axiom', *The Journal of Law and Economics*, 20(1), 133-152.

Hollis, M. (1977) *Models of man: Philosophical thoughts on social action*, Cambridge: Cambridge University Press.

Infante, G., Lecouteux, G., Sugden, R. (2016) 'Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics', *Journal of Economic Methodology*, 23(1), 1-25.

Irvine, Andrew (1998) 'How Braess' paradox solves Newcomb's problem' in Danielson, P.

(ed.) *Modeling Rationality, Morality, and Evolution*. New York Oxford: Oxford University Press, 161-185.

Iyengar, S.S., Huberman, G., Jiang, W. (2004) 'How much choice is too much? Contributions to 401(k) retirement plans' in Mitchell, O.S., Utkus, S.P. (eds.) *Pension design and structure: New lessons from behavioral finance*. Oxford: Oxford University Press, 84-87.

John, P., Smith, G., Stoker, G. (2009) 'Nudge nudge, think think: two strategies for changing civic behaviour', *The Political Quarterly*, 80(3), 361-370.

Johnson, E.J., Goldstein, D. (2003) 'Do defaults save lives?', *Science*, 302 (5649), 1338-1339.

Jolls, C., Sunstein, C.R., Thaler, R.H. (1998) 'A behavioral approach to law and economics', *Stanford Law Review*, 50, 1471.

Joyce, R. (2007) *The evolution of morality*, Cambridge, MA: MIT Press.

Justman, S. (1996) 'Regarding others', *New Literary History*, 27(1), 83-93.

Kagel, J. (1987) 'Economics according to the rats (and pigeons too): what have we learned and what we hope to learn' in Roth, A. (ed.) *Laboratory Experimentation in Economics: Six Points of View*, New York: Cambridge University Press.

Kahneman, D., Sugden, R. (2005) 'Experienced utility as a standard of policy evaluation', *Environmental and Resource Economics*, 32(1), 161-181.

Kahneman, D., Tversky, A. (1979) 'Prospect Theory: An Analysis of Decision under Risk', *Econometrica*, 47(2), 263.

Kahneman, D., Tversky, A. (1984) 'Choices, values, and frames', *American Psychologist*, 39(4), 341.

Kahneman, D., Tversky, A. (2000) *Choices, values, and frames*, New York: Cambridge University Press.

Kahneman, D. (2011) *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux.

Kahneman, D., Knetsch, J.L., Thaler, R. (1986) 'Fairness as a constraint on profit seeking: Entitlements in the market', *The American Economic Review*, 76(4), 728-741.

Kahneman, D., Knetsch, J.L., Thaler, R.H. (1986) 'Fairness and the assumptions of economics', *Journal of Business*, 59, 285-300.

Kahneman, D., Slovic, P., Tversky, A. (1982) *Judgments under uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.

Kavka, G.S. (1982) 'The paradox of future individuals', *Philosophy & Public Affairs*, 11, 93-112.

Khalil, E.L. (1997) 'Etzioni versus Becker: Do Moral Sentiments differ from Ordinary Tastes?', *De Economist*, 145(4), 491-520.

Kirchgässner, G. (2017) 'Soft paternalism, merit goods, and normative individualism', *European Journal of Law and Economics*, 43(1), 125-152.

Klein, D. B. (1992) 'Promise Keeping: A Model of Credit Information Sharing', *Economics and Politics*, 4(2), 117-136.

Klein, D.B. (2015) 'Virtues, Layers, and Lenses in Adam Smith's The Theory of Moral Sentiments', GMU Working Paper in Economics, No. 15-06. Available at SSRN: <https://ssrn.com/abstract=2554157> or <http://dx.doi.org/10.2139/ssrn.2554157> (Accessed: 10 January 2016).

Klein, D.B. (2016) 'Adam Smith's non-foundationalism', *Society*, 53(3), 278-286.

Knight, F. H. (1922) 'Ethics and the Economic Interpretation', *The Quarterly Journal of Economics*, 36(3), 454-481.

Knight, F.H. (1997) *The ethics of competition*, New Brunswick, NJ: Transaction Publishers.

Knight, F.H. (2012) *Risk, uncertainty and profit*, New York: Courier Corporation.

Korsgaard, C.M. (1989) 'Personal identity and the unity of agency: A Kantian response to Parfit', *Philosophy and Public Affairs*, 18(2), 101-132.

Korsgaard, C.M. (2009) *Self-constitution: Agency, identity, and integrity*, New York: Oxford University Press.

Koslowski, P. (2006) [SSRN] 'The Economy of Happiness'. Available

at: <https://ssrn.com/abstract=945328> or <http://dx.doi.org/10.2139/ssrn.945328> (Accessed: 5 May 2011).

Kőszegi, B. and Rabin, M. (2007) 'Mistakes in choice-based welfare analysis', *The American Economic Review*, 97(2), 477-481.

Kőszegi, B. and Rabin, M. (2008) 'Choices, situations, and happiness', *Journal of Public Economics*, 92(8), 1821-1832.

Krapiec, M. (1993) *Person and Natural Law*, M. Szymańska, trans., New York: Peter Lang.

Lamb, R.B. (1974) 'Adam Smith's system: sympathy not self-interest', *Journal of the History of Ideas*, 35, 671-682.

Langholm, O. (1979) *Price and value in the Aristotelian tradition: A study in scholastic economic sources*, Bergen: Universitetsforlaget.

Langholm, O. (1982) 'Economic freedom in scholastic thought', *History of Political Economy*, 14(2), 260-283.

Langholm, O. (1987) 'Scholastic Economics', in Lowry, T. (ed.) *Pre-Classical Economic Thought*, Dordrecht: Kluwer Academic Publishers, 115-135.

Langholm, O. (1992) *Economics in the medieval schools: wealth, exchange, value, money and usury according to the Paris theological tradition, 1200-1350*, Leiden, New York, Köln: EJ Brill.

Langholm, O. (1998) *The legacy of scholasticism in economic thought: antecedents of choice and power*, Cambridge: Cambridge University Press.

Langholm, O. (2003) *The merchant in the confessional: trade and price in the pre-reformation penitential handbooks*, Leiden and Boston: Brill.

Lapidus, A. (1994) 'Norm, virtue and information: the just price and individual behaviour in Thomas Aquinas' *Summa Theologiae*', *The European Journal of the History of Economic Thought*, 1(3), 435-473.

Lavergne, M. (2004) 'The opportunity criterion: consumer sovereignty without the assumption of coherent preferences', *The American Economic Review*, 94(4), 1014-1033.

- Lawson, T. (2006) *Economics and reality*, London and New York: Routledge.
- Lecouteux, G. (2015) *Reconciling normative and behavioural economics*, Doctoral dissertation, Ecole Polytechnique. Available at: <https://pastel.archives-ouvertes.fr/tel-01175744/document> (Accessed: 20 November 2016).
- Lemmons, R.M.H. (2011) *Ultimate Normative Foundations: The Case for Aquinas's Personalist Natural Law*, Lanham, MD: Lexington Books.
- Letwin, W. (1964) *The Origins of Scientific Economics*, New York: Doubleday.
- Levine, D.P. (1998) 'The self and its interests in classical political economy', *The European Journal of the History of Economic Thought*, 5(1), 36-59.
- Levy, D.M. (1995) 'The Partial Spectator in the Wealth of Nations: A Robust Utilitarianism', *European Journal of the History of Economic Thought* 2(2), 299-326.
- Lewis, P.A. (2011) 'Varieties of Emergence: Mind, Markets, and Novelty', *Studies in Emergent Order*, 4, 170-92.
- Lichtenstein, S., Slovic, P. (2006) *The construction of preference*, New York: Cambridge University Press.
- Lichtenstein, S., Slovic, P. (1971) 'Reversals of preference between bids and choices in gambling decisions', *Journal of Experimental Psychology*, 89(1), 46.
- Little, I.M. (1949) 'A Reformulation of the Theory of Consumer's Behaviour', *Oxford Economic Papers*, 1(1), 90-99.
- Loewenstein, G., Haisley, E. (2008) 'The economist as therapist: Methodological ramifications of 'light' paternalism' in Caplin, A., Schotter, A. (eds.) *The Foundations of Positive and Normative Economics: A Handbook*. Oxford: Oxford University Press.
- Lopes, H. (2014) 'Individuals, persons and agency theory—contrasted views on social interactions at work', Working paper. Available at: https://repositorio.iscte-iul.pt/bitstream/10071/7756/1/DINAMIA_WP_2014-04.pdf (Accessed: 13 March 2017).
- Lowry, S.T. (1974) 'Aristotle's' Natural Limit and the Economics of Price Regulation', *Greek, Roman and Byzantine Studies*, 15(1), 57.

- Lowry, S.T. (1987) *Pre-classical economic thought: from the Greeks to the Scottish enlightenment*, Dordrecht: Kluwer Academic Publishers.
- Lowry, S.T. (2003) 'Ancient and medieval economics' in Samuels, W.J., Biddle, J.E., Davis, J.B. (eds.) *A Companion to the History of Economic Thought*, Oxford: Blackwell, 9-27.
- Macfie, A.L. (1959) 'Adam Smith's Moral Sentiments as foundation for his Wealth of Nations', *Oxford Economic Papers*, 11(3), 209-228.
- Macfie, A.L. (2013) *The individual in society: Papers on Adam Smith*, London and New York: Routledge.
- Mackie, J. (1990) *Ethics: Inventing right and wrong*, London: Penguin.
- Magnuson, E. (2008) 'Rejecting the American dream: Men creating alternative life goals', *Journal of Contemporary Ethnography*, 37(3), 255-290.
- Mäki, Uskali (1992) 'On the method of isolation in economics' in Dilworth, C. (ed.) *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26, 319-354.
- Mäki, Uskali (1994) 'Reorienting the assumptions issue', in Backhouse, R. (ed.) *New Directions in Economic Methodology*. London: Routledge, 236-256.
- Mäki, Uskali (2000) 'Kinds of assumptions and their truth: Shaking an untwisted F-twist', *Kyklos*, 53(3), 303-322.
- Mäki, Uskali (2009) 'Realistic realism about unrealistic models' in Kincaid, H., Ross, D. (eds.) *The Oxford Handbook of Philosophy of Economics*, New York: Oxford University Press, 68-98.
- Mäki, Uskali (2012) 'The truth of false idealizations in modeling' in Humphreys, P., Imbert, C. (eds.) *Models, Simulations and Representations*. New York: Routledge, 216-233.
- Margolis, H. (1984) *Selfishness, altruism, and rationality*, University of Chicago Press.
- Markman, A.B., Zhang, S., Moreau, C.P. (2000) 'Representation and the construction of preferences' in Dietrich, E., Markman, A.B. (eds.) *Cognitive Dynamics*, Mahwah, NJ: Lawrence Erlbaum Associates.

- Marshall, A. (1890) *Principles of Economics*, London: Macmillan.
- Marteau, T.M., Ogilvie, D., Roland, M., Suhrcke, M., Kelly, M. (2011) 'Judging Nudging: Can Nudging Improve Population Health?', *BMJ*, 342: d228.
- McCloskey, D. (2006) 'The Faithful and Hopeful Economic Agent'. Available at: <https://www.deirdremccloskey.com/docs/faithful.doc> (Accessed: 18 November 2016).
- McCloskey, D. (2009) 'Adam Smith, the Last of the Former Virtue Ethicists' in Young, J.T. (ed.), *Elgar Companion to Adam Smith*, Cheltenham: Edward Elgar Publishing.
- McKenzie, R.B. (2009) *Predictably rational?: in search of defenses for rational behavior in economics*, New York: Springer.
- McPherson, Michael S. (1983) 'Interpretive Aspects of Economic Inquiry' in Haan, N., Bellah, R.N., Rabinov, P., Sullivan, W.M. (eds.) *Social Sciences as Moral Inquiry*. New York: Columbia University Press.
- Medina, J., Justo, P.H. (2016) 'Not all nudges are automatic: freedom of choice and informative nudges', paper presented to the European Consortium for Political Research Joint Session of Workshops – 2016 Behavioural Change and Public Policy Holger Strassheim and Silke Becke Chairs, Pisa, Italy, 24-28 April. Available at: <https://ecpr.eu/Filestore/PaperProposal/9bce4cb9-f3ba-4069-9b9acb755fc0df23.pdf> (Accessed: 8 June 2017).
- Meek, R.L. (1955) *The Scottish contribution to Marxist sociology*, London: Lawrence & Wishart.
- Meikle, S. (1996) 'Aristotle on business', *The Classical Quarterly*, 46(1), 138-151.
- Meikle, S. (1997) *Aristotle's economic thought*, Oxford: Oxford University Press.
- Meikle, S. (2000) 'Quality and quantity in economics: The metaphysical construction of the economic realm', *New Literary History*, 31(2), 247-268.
- Melzer, A. M. (2001) 'Anti-anti-Foundationalism: Is a Theory of Moral Sentiments Possible?', *Perspectives on Political Science*, 30(3), 151-155.
- Mill, J.S. (2001) [1859] *On Liberty*, Kitchener: Batoche Books.

- Mills, C. (2013) 'Why nudges matter: A reply to Goodwin', *Politics*, 33(1), 28-36.
- Mills, C. (2015) 'The heteronomy of choice architecture', *Review of Philosophy and Psychology*, 6(3), 495-509.
- Mitchell, G. (2004) 'Libertarian paternalism is an oxymoron', *Northwestern University Law Review*, 99, 1245.
- Moberg, D.J. (2008) 'Mentoring and practical wisdom: Are mentors wiser or just more politically skilled?', *Journal of Business Ethics*, 83, 835-843.
- Monaghan, L.F. (2008) *Men and the war on obesity: A sociological study*, New York: Routledge.
- Montes, L. (2009) 'Adam Smith's Stoicism and Self-Command'. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.148.2363> (Accessed: 12 July 2016).
- Montes, L. (2003) *Adam Smith in Context: a critical reassessment of some central components of his thought*, New York: Palgrave Macmillan.
- Montes, L. (2008) 'Adam Smith as an eclectic Stoic', *Adam Smith Review*, 4, 30-56.
- Morrow, G.R. (1923) *The ethical and economic theories of Adam Smith*, New York: AM Kelley.
- Morrow, G.R. (1927) 'Adam Smith: moralist and philosopher', *Journal of Political Economy*, 35(3), 321-342.
- Mueller, J.D. (2010) *Redeeming economics*, Wilmington: ISI Books.
- Muñoz de Juana, R. (2001) 'Scholastic Morality and the birth of Economics: the thought of Martin de Azpilcueta', *Journal of Markets and Morality*, 4(1), 14-42.
- Murphy, J.B. (2002) 'Equality in exchange', *American Journal of Jurisprudence*, 47, 85-121.
- Myers, M.L. (1900) *The soul of modern economic man*, Chicago: University of Chicago Press Economics Books.
- Myrdal, G. (1969) *The Political Element in the Development of Economic Theory*, New York:

Simon and Schuster.

Nelkin, D.K. (2007) 'Do we have a coherent set of intuitions about moral responsibility?', *Midwest Studies in Philosophy*, 31(1), 243-259.

Neumann, J. von, Morgenstern, O. (1944) *Theory of games and economic behavior*, Princeton, NJ: Princeton University Press.

Niker, F. (2017) 'Policy-Led 'Ecological' Virtue-Cultivation: can we 'nudge' citizens towards developing virtues?' in Harrison, T., Walker, D. (eds.), *Theory and Practice of Virtue Education*, Taylor & Francis.

Nozick, R. (1994) *The nature of rationality*, Princeton, NJ: Princeton University Press.

Oakley, A. (1994) *Classical economic man: human agency and methodology in the political economy of Adam Smith and JS Mill*, Brookfield, Vermont: Edward Elgar Publishing.

Osborne, G. (2008) "Nudge, nudge, win, win", *The Guardian*, 14 July.

Otteson, J.R. (2002) *Adam Smith's marketplace of life*, Cambridge: Cambridge University Press.

Pareto, V. (1971) *Manual of political economy*, Basingstoke: Palgrave Macmillan.

Parfit, D. (1982) 'Personal identity and rationality', *Synthese*, 53(2), 227-241.

Pasche, M. (2014) Soft Paternalism and Nudging - Critique of the Behavioral Foundations, MPRA Paper No. 61140. Available at: <https://mpra.ub.uni-muenchen.de/61140/> (Accessed: 21 May 2017).

Patrone, D. (2005) *Leaving Morality Where It Is: Contingency and the Particularistic Approach to Morality*, New York: Lexington Books.

Paul, L.A. (2015a) 'What You Can't Expect When You're Expecting', *Res Philosophica* 92(2), 149-70.

Paul, L.A. (2015b) *Transformative Experience*, New York: Oxford University Press.

Payne, J.W., Bettman, J.R., Johnson, E.J. (1992) 'Behavioral decision research: A constructive

processing perspective', *Annual Review of Psychology*, 43(1), 87-131.

Peter, F., Schmid, H.B. (2007) *Rationality and commitment*, Oxford: Oxford University Press.

Pettit, P. (1991) 'Decision theory and folk psychology' in Bacharach, M., Hurley, S. (eds.) *Foundations of Decision Theory: Issues and Advances*, Oxford: Blackwells, 147-75.

Phelps, E. (2006) 'Economic culture and economic performance: What light is shed on the Continent's problem?', paper presented to "Perspectives on the Performance of the Continent's Economies" Conference of CESifo and Center on Capitalism and Society, Venice, 21-22 July 2006. Available at: <http://www.columbia.edu/~esp2/Center3rdAnnualConf2006Sept27.pdf> (Accessed: 12 June 2017)

Phelps, E. (2007) 'Entrepreneurial culture', *The Wall Street Journal*, February 12. Available at: <https://www.wsj.com/articles/SB117124449952605357> (Accessed: 12 June 2017).

Phelps, E. (2009) 'Economic justice and the spirit of dynamism', *First Things*, October 1. Available at: <https://www.firstthings.com/article/2009/10/economic-justice-and-the-spirit-of-innovation> (Accessed: 12 June 2017).

Phelps, E. (2013) *Mass flourishing*, Princeton, NJ: Princeton University Press.

Plott, C.R. (1996) 'Rational individual behavior in markets and social choice processes: the discovered preference hypothesis' in *The Rational Foundations of Economic Behaviour*. London: McMillan, 225-250.

Pollak, R.A. (1970) 'Habit formation and dynamic demand functions', *Journal of Political Economy*, 78(4), 745-763.

Pollak, R.A. (1976) 'Habit formation and long-run utility functions', *Journal of Economic Theory*, 13(2) 272-297.

Posner, R.A. (1997) 'Are We One Self or Multiple Selves?: implications for law and public policy', *Legal Theory*, 3(1), 23-35.

Posner, R.A. (1998) 'Rational choice, behavioral economics, and the law', *Stanford Law Review*, 50, 1551-1575.

Potts, J. (2000) *The new evolutionary microeconomics*, Cheltenham: Edward Elgar Publishing.

- Qizilbash, M. (2012) 'Informed desire and the ambitions of libertarian paternalism', *Social Choice and Welfare*, 38(4), 647-658.
- Quante, M. (2011) 'In defence of personal autonomy', *Journal of Medical Ethics*, 37(10), 597-600.
- Rachlinski, J.J. (2003) 'The Uncertain Psychological Case for Paternalism', *Northwestern University Law Review*, 97(3), 1165-1226.
- Railton, P. (1986) 'Moral realism', *The Philosophical Review*, 95(2), 163-207.
- Raphael, D.D., Macfie, A.L. (1984) 'Introduction' in Raphael, D.D., Macfie, A.L. (eds.) *The Theory of Moral Sentiments*, Indianapolis: Liberty Fund, 1-52.
- Rawls, J., (1980) 'Kantian constructivism in moral theory', *The Journal of Philosophy*, 77(9), 515-572.
- Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.
- Rawls, John. (1999) *A Theory of Justice*, Revised Edition. Cambridge, MA: Harvard University Press.
- Raz, J. (1986) *The Morality of Freedom*, New York: Oxford University Press.
- Read, D. (2006) 'Which side are you on? The ethics of self-command', *Journal of Economic Psychology*, 27, 681-693.
- Rebonato, R. (2012) *Taking liberties: A Critical Examination of Libertarian Paternalism*, Basingstoke and New York: Palgrave Macmillan.
- Reid, D. (2008) 'Thomas Aquinas and Viscount Stair: the influence of scholastic moral theology on Stair's account of restitution and recompense', *The Journal of Legal History*, 29(2), 189-214.
- Rich, P. (2014) 'Comparing the axiomatic and ecological approaches to rationality: fundamental agreement theorems in SCOP', *Synthese*, 195, 1-19.
- Rieskamp, J., Busemeyer, J.R., Mellers, B.A. (2006) 'Extending the bounds of rationality: evidence and theories of preferential choice', *Journal of Economic Literature*, 44, 631-661.

- Rizzo, M.J. and Whitman, D.G. (2009a) 'Little brother is watching you: New paternalism on the slippery slopes', *Arizona Law Review*, 51, 685-739.
- Rizzo, M. J. and Whitman, D. G. (2009b) 'The knowledge problem of the new paternalism', *Brigham Young University Law Review*, 4, 905-968.
- Roca, E. (2007) 'Introducing practical wisdom in business schools', *Journal of Business Ethics*, 82, 607-620.
- Roll, E. (1938) *History of Economic Thought*, London: Faber&Faber.
- Roover, R. (1957) 'Joseph A. Schumpeter and scholastic economics', *Kyklos*, 10(2), 115-146.
- Rona,P, Zsolnai,L.(2017) *Economics as a Moral Science*, Switzerland: Springer International Publishing.
- Rorty, A.O. (1980) *Essays on Aristotle's Ethics*, Berkeley, Los Angeles, London: University of California Press.
- Rosati, C.S. (1996) 'Internalism and the Good for a Person', *Ethics*, 106(2), 297-326.
- Ross, D. (2002) 'Why people are atypical agents', *Philosophical Papers*, 31(1), 87-116.
- Ross, D. (2005) *Economic theory and cognitive science: Microexplanation*, Cambridge, MA and London: MIT Press.
- Ross, D. (2006) 'Evolutionary game theory and the normative theory of institutional design: Binmore and behavioral economics', *Politics, Philosophy & Economics*, 5(1), 51-79.
- Ross, D. (2012) 'The Economic Agent: Not Human but Important' in Mäki, U. (ed.) *Philosophy of Economics*, Amsterdam: Elsevier, 691-736.
- Rothschild, E. (2004) 'Dignity or meanness', *The Adam Smith Review*, 1, 150-164.
- Rubinfeld, J. (2001) 'The New Unwritten Constitution', *Duke Law Journal*, 51, 289.
- Runde, J. (2009) 'Dissecting the black swan', *Critical Review*, 21(4), 491-505.
- Ryan, R.M., Deci, E.L. (2000) 'Intrinsic and extrinsic motivations: Classic definitions and new directions', *Contemporary educational psychology*, 25(1), 54-67.

Samuelson, P.A. (1937) 'A note on measurement of utility', *The Review of Economic Studies*, 4(2), 155-161.

Samuelson, P.A. (1938) 'A note on the pure theory of consumer's behavior', *Economica*, 5(17), 61-71.

Samuelson, P.A. (1947) *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.

Sandven, T. (1999a) 'Autonomy, Adaptation, and Rationality—A Critical Discussion of Jon Elster's Concept of "Sour Grapes," Part I', *Philosophy of Social Sciences*, 29(1), 3-31.

Sandven, T. (1999b) 'Autonomy, Adaptation, and Rationality—A Critical Discussion of Jon Elster's Concept of "Sour Grapes," Part II', *Philosophy of Social Sciences*, 29(2), 173-205.

Sarasvathy, S. D. (2002) 'Entrepreneurship as economics with imagination', *Business Ethics Quarterly*, 3, 95-112.

Satz, D., Ferejohn, J. (1994) 'Rational Choice and Social Theory', *The Journal of Philosophy*, 91(2) 1994), 71-87.

Schelling, T.C. (1980) 'The intimate contest for self-command', *The Public Interest*, Vol. 60, 94.

Schelling, T.C. (1984) 'Self-command in practice, in policy, and in a theory of rational choice', *The American Economic Review*, 74(2), 1-11.

Schelling, T.C. (2006) *Micromotives and macrobehavior*, New York: WW Norton & Company.

Schenck, David (1976) 'Re-casting the 'Ethics of Virtue/Ethics of Duty Debate'', *Journal of Religious Ethics*, 4, 269-286.

Schneewind, J.B. (1998) *The invention of autonomy: A history of modern moral philosophy*, Cambridge: Cambridge University Press.

Schnellenbach, J. (2011) 'Some Notes on the Nudge: The Political Economy of Libertarian Paternalism in Democratic Societies'. Available at SSRN: <https://ssrn.com/abstract=1854670> or <http://dx.doi.org/10.2139/ssrn.1854670> (Accessed: 12 April 2017).

- Schnellenbach, J. (2012) 'Nudges and norms: On the political economy of soft paternalism', *European Journal of Political Economy*, 28(2), 266-277.
- Schooler, J.W., Ariely, D., Loewenstein, G. (2003) 'The pursuit and assessment of happiness can be self-defeating', *The Psychology of Economic Decisions*, 1, 41-70.
- Schubert, C., Cordes, C. (2013) 'Role Models that Make You Unhappy: Light Paternalism, Social Learning, and Welfare', *Journal of Institutional Economics*, 9(2), 131–159.
- Schubert, C. (2015) 'On the ethics of public nudging: Autonomy and Agency'. Available at SSRN: <https://ssrn.com/abstract=2672970> or <http://dx.doi.org/10.2139/ssrn.2672970> (Accessed: 10 April 2017).
- Schumpeter, J.A. (1954) *History of Economic Analysis*, New York: Oxford University Press.
- Sen, A. (1973) 'Behaviour and the Concept of Preference', *Economica*, 40 (159), 241-259.
- Sen, A. (1982) 'Rights and agency', *Philosophy & Public Affairs*, 11(1), 3-39.
- Sen, A. (1999) *On ethics and economics*, Oxford: Oxford University Press.
- Sen, A. (2004) *Rationality and freedom*, Cambridge, MA: Harvard University Press.
- Sen, A. (1977) 'Rational fools: A critique of the behavioral foundations of economic theory', *Philosophy & Public Affairs*, 6(4), 317-344.
- Sen, A. (2009) *The Idea of Justice*, London: Penguin Books.
- Shackle, G.L.S. (1979) *Imagination and the Nature of Choice*, Edinburgh: Edinburgh University Press.
- Shafir, E., Simonson, I., Tversky, A. (1993) 'Reason-based choice', *Cognition*, 49(1), 11-36.
- Shefrin, H., Thaler, R. (1981) 'Gift-Giving and Mental Accounts', Mimeo, Santa Clara University.
- Sherman, N. (1989) *The Fabric of Character: Aristotle's Theory of Virtue*, Oxford: Clarendon Press.
- Shibutani, T. (1974) *Society and personality: An interactionist approach to social psychology*,

New Brunswick and London: Transaction Publishers.

Shiffrin, S. (1991) 'Moral autonomy and agent-centred options', *Analysis*, 51(4), 244-254.

Shiffrin, S.V. (2000) 'Paternalism, Unconscionability Doctrine, and Accommodation', *Philosophy of Public Affairs*, 29(3), 205-250.

Simon, D., Krawczyk, D.C., Holyoak, K.J. (2004) 'Construction of preferences by constraint satisfaction', *Psychological Science*, 15(5), 331-336.

Simon, H.A. (1955) 'A behavioral model of rational choice', *The Quarterly Journal of Economics*, 69(1), 99-118.

Slovic, P. (1995) 'The construction of preference', *American Psychologist*, 50(5), 364.

Small, A. (1907) *Adam Smith and modern sociology: A Study in the Methodology of the Social Sciences*, Kitchener: Batoche Books.

Smart, J.J.C., Williams, B. (1973) *Utilitarianism: For and against*, Cambridge: Cambridge University Press.

Smith, A. (1976) *An Inquiry into the Nature and Causes of the Wealth of Nations*, in Campbell, R.H., Skinner, A.S., Todd, W.B. (eds.) *The Glasgow Edition of the Works and Correspondence of Adam Smith*, Vol. 2. Oxford: Clarendon Press.

Smith, A. (1976) *Theory of Moral Sentiments*, in Raphael, D.D., Macfie, A.L., Campbell, R.H., Skinner, A.S. (eds.) *The Glasgow Edition of the Works and Correspondence of Adam Smith*, Vol. 1. Oxford: Oxford University Press.

Smith, Adam (1759) *The Theory of Moral Sentiments*, 6th edn. London: A. Millar.

Smith, Adam (1982) *Lectures on Jurisprudence*, Indianapolis: Liberty Press.

Smith, Adam (1904) *An Inquiry into the Nature and Causes of the Wealth of Nations*, London: Methuen & Co.

Smith, V.L. (1976) 'Experimental economics: Induced value theory', *The American Economic Review*, 66(2), 274-279.

- Smith, V.L. (2012) 'Adam Smith on humanomic behaviour', *The Journal of Behavioral Finance & Economics*, 2(1), 1-20.
- Sobel, D. (2016) *From Valuing to Value: Towards a Defense of Subjectivism*, Oxford: Oxford University Press.
- Stewart, H. (1992) 'Rationality and the market for human blood', *Journal of Economic Behavior and Organization*, 19, 125-43
- Stewart, H. (1995) 'A Critique of Instrumental Reason in Economics', *Economics and Philosophy* 11(1), 57-83.
- Stigler, G.J., Becker, G.S. (1977) 'De gustibus non est disputandum', *The American Economic Review*, 67(2), 76-90.
- Stigler, G.J. (1982) *The Economist as Preacher, and other essays*, Chicago: University of Chicago Press.
- Stiglitz, J.E. (2002) 'Information and the Change in the Paradigm in Economics', *The American Economic Review*, 92(3), 460-501.
- Storr, V.H. (2018) 'The Impartial Spectator and the Moral Teachings of Markets' in Schmidtz, D., Pavel, C.E. (eds.) *The Oxford Handbook of Freedom*. New York: Oxford University Press.
- Strotz, R.H. (1955) 'Myopia and inconsistency in dynamic utility maximization', *The Review of Economic Studies*, 23(3), 165-180.
- Sugden, R. (1985) 'Why be consistent? A critical analysis of consistency requirements in choice theory', *Economica*, 52(206), 167-183.
- Sugden, R. (2002) 'Beyond sympathy and empathy: Adam Smith's concept of fellow-feeling', *Economics & Philosophy*, 18(1), 63-87.
- Sugden, R. (2006) 'Hume's non-instrumental and non-propositional decision theory', *Economics & Philosophy*, 22(3), 365-391.
- Sugden, R. (2008) 'Why incoherent preferences do not justify paternalism', *Constitutional Political Economy*, 19(3), 226-248.

Sugden, R., Gold, N. (2006) *Beyond Individual choice: Teams and Frames in Game Theory*, Princeton, NJ: Princeton University Press.

Sunstein, C. R. (1993) 'Democracy and Shifting Preferences' in Copp, D., Hampton, J., Roemer, J.E. (eds.) *The Idea of Democracy*, Cambridge: Cambridge University Press, 196-230.

Sunstein, C.R. (2000) *Behavioural Law and Economics*, New York: Cambridge University Press.

Sunstein, C.R. (2013) 'Deciding by Default', *University of Pennsylvania Law Review* 162(1), 1-57.

Sunstein, C.R., Thaler, R.H. (2003) 'Libertarian paternalism is not an oxymoron', *The University of Chicago Law Review*, 70, 1159-1202.

Sunstein C.R., Thaler R. (2005) 'Libertarian Paternalism' in Sunstein, C. (ed.) *Laws of Fear: Beyond the Precautionary Principle*, Cambridge: Cambridge University Press.

Sunstein, C.R. (2012) The Storrs lectures: Behavioral economics and paternalism, *Yale Law Journal*, 122, 1826.

Sunstein, C.R. (2014) *Why nudge?: The politics of libertarian paternalism*, New Haven: Yale University Press.

Sunstein, C.R. (2015a) 'Nudges, Agency, Navigability, and Abstraction: A Reply to Critics', *Review of Philosophy and Psychology*, 6(3), 511-529.

Sunstein, C.R. (2015b) 'Nudges Do Not Undermine Human Agency', *Journal of Consumer Policy*, 38, 207-210.

Sunstein, C.R. (2015c) 'The Ethics of Nudging', *Yale Journal on Regulation*, 32(2), 413-450.

Sunstein, C.R. (2016) 'Fifty shades of manipulation', *Journal of Marketing Behavior*, 1(3-4), 293-306.

Szulczewski, G. (2015) 'Krytyka 'etyki sympatii i 'problem A. Smitha w świetle lektury Teorii uczuć moralnych', *Prakseologia*, 157, 79.

Taylor, C. (1983) *Social theory as practice*, Delhi: Oxford University Press.

- Taylor, C. (1985a) 'Self-interpreting animals' in Taylor, C. (ed.) *Human agency and language*, (Philosophical papers: volume 1), Cambridge and New York: Cambridge University Press, 45-76.
- Taylor, C. (1985b) 'Philosophy and the Human Sciences' in Taylor, C. (ed.) *Philosophical papers: volume 2*, Cambridge and New York: Cambridge University Press.
- Taylor, C. (1989) *Sources of the self: The making of the modern identity*, Cambridge and New York: Cambridge University Press.
- Teichgraeber, R. (1981) 'Rethinking Das Adam Smith Problem', *Journal of British Studies* 20(2), 106-123.
- Teschl, M. (2006) 'The impact of identity on economics' in Clary, B.J., Dolfsma, W., Figart, D.M. (eds.) *Ethics and the Market: Insights from Social Economics*, London and New York: Routledge, 84-97.
- Thaler, R.H., Shefrin, H.M. (1981) 'An economic theory of self-control', *Journal of Political Economy*, 89(2), 392-406.
- Thaler, R. H., Sunstein, C. (2003) 'Libertarian paternalism', *American Economic Review*, 93, 175–179.
- Thaler R, Sunstein C. (2006) 'Preferences, paternalism, and liberty', *Philosophy*, 59, 233–264.
- Thaler, R.H., Sunstein, C.R. (2008) *Nudge: Improving decisions about health, wealth, and happiness*, New Haven: Yale University Press.
- Thaler, Richard H., Sunstein, C.R. (2009) *Nudge: improving decisions about health, wealth, and happiness* [Rev. and expanded ed.] New York: Penguin Books.
- Thoron, S. (2016) 'Morality Beyond Social Preferences: Smithian Sympathy, Social Neuroscience and the Nature of Social Consciousness, Psychology and Economics', *Historical Perspective*, 6(2), 235-264.
- Todorov, T. (1996) 'Living alone together', *New Literary History*, 27(1), 1-14.
- Trout, J.D. (2005) 'Paternalism and cognitive bias', *Law and Philosophy*, 24(4), 393-434.

- Tversky, A., Kahneman, D. (1986) 'Rational choice and the framing of decisions', *Journal of Business*, 59(4), 251-278.
- Tversky, A., Thaler, R.H. (1990) 'Anomalies: preference reversals', *The Journal of Economic Perspectives*, 4(2), 201-211.
- Valdman, M. (2010) 'Outsourcing self-government', *Ethics*, 120(4), 761-790.
- Van Staveren, I. (2001) *The values of economics: An Aristotelian perspective*, London and New York: Routledge.
- Vanberg, V.J. (2006) 'Rationality, rule-following and emotions: on the economics of moral preferences', Papers on Economics and Evolution no. 0621. Available at: https://www.researchgate.net/publication/45139272_Rationality_rule-following_and_emotions_on_the_economics_of_moral_preferences (Accessed: 3 February 2016).
- Vanberg, V.J. (2008) 'On the economics of moral preferences', *American Journal of Economics and Sociology*, 67(4), 605-628.
- Veatch, H.B. (1962) *Rational man: A modern interpretation of Aristotelian ethics*, Indianapolis: Liberty Fund.
- Venkataraman, S. (1997), 'The Distinctive Domain of Entrepreneurship Research' in Katz, J., Brockhouse, R. (eds.) *Advances in entrepreneurship, firm emergence and growth*, Greenwich, CT: JAI Press, 119-138.
- Viner, J. (1927) 'Adam Smith and laissez faire', *Journal of Political Economy*, 35(2), 198-232.
- Viner, J. (2015) *The role of providence in the social order: an essay in intellectual history*, Princeton, NJ: Princeton University Press.
- Von Weizsäcker, C.C. (1971) 'Notes on endogenous change of tastes', *Journal of Economic Theory*, 3(4), 345-372.
- Wagner, R.E. (2006) 'States and the crafting of souls: Mind, society, and fiscal sociology', *Journal of Economic Behavior & Organization*, 59(4), 516-524.

- Wagner, R.E. (2010) *Mind, Society, and Human Action: Time and Knowledge in a Theory of Social-Economy*, London and New York: Routledge.
- Walsh, V. (2000) 'Smith after Sen', *Review of Political Economy* 12(1), 5-25.
- Wartenberg, T.E. (1990) *Forms of Power: From Domination to Transformation*, Philadelphia: Temple University Press.
- Waszek, N. (1984) 'Two concepts of morality: A distinction of Adam Smith's ethics and its stoic origin', *Journal of the History of Ideas*, 45(4), 591-606.
- Werhane, P. (1991) *Adam Smith and his legacy for modern capitalism*, Oxford: Oxford University Press.
- West, E.G. (1969) 'Adam Smith's Philosophy of Riches', *Philosophy*, 44(168), 101-115.
- White, M.D. (2011) *Kantian ethics and economics: autonomy, dignity, and character*, Stanford, CA: Stanford University Press.
- White, M. (2013) *The manipulation of choice: Ethics and libertarian paternalism*, New York: Palgrave Macmillan.
- White, M.D. (2015) 'Judgment: Balancing principle and policy', *Review of Social Economy*, 73(3), 223-241.
- Whitebook, J. (1978) 'Pre-market economics: The Aristotelian perspective', *Dialectical Anthropology*, 3(3), 197-220.
- Whitman, D.G. (2004) 'Meta-Preferences and Multiple Selves', Mimeo. Available at econfaculty.gmu.edu/pboettke/workshop/archives/f03/whitman.doc (Accessed: 12 May 2017).
- Wiggins, D. (1980) 'Deliberation and Practical Reason' in Rorty, A. (ed.) *Essays on Aristotle's Ethics*. Berkeley, Los Angeles, London: University of California Press.
- Wiggins, D. (1988) *Truth, Invention, and the Meaning of Life* in Sayre-McCord, G. (ed.) *Essays on Moral Realism*. Cornell University Press, 127-65.
- Williams, B. (1973) 'A Critique of Utilitarianism' in Smart, J.J.C., Williams, B. (eds.) *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

- Williams, B., Sen, A. (1982), *Utilitarianism and beyond*, Cambridge: Cambridge University Press, 39-62.
- Wilson, B.J. (2010) 'Social preferences aren't preferences', *Journal of Economic Behavior & Organization*, 73(1), 77-82.
- Wilson, T. (1976) 'Sympathy and Self-interest' in Wilson, T., Skinner, A.S. (eds.) *The Market and the State. Essays in Honour of Adam Smith*, Oxford: Clarendon Press, 74-112.
- Wincewicz-Price, A. (2016) 'Homo Economicus Moralis' in Gorazda, M., Kwarcinski, T., Hardt, Ł. (eds.) *Metaekonomia: zagadnienia z filozofii ekonomii*, Torun: Copernicus Center Press.
- Winch, D. (1992) 'Adam Smith: Scottish moral philosopher as political economist', *The Historical Journal*, 35(1), 91-113.
- Winston, G.C. (1980) 'Addiction and backsliding: a theory of compulsive consumption', *Journal of Economic Behavior & Organization*, 1(4), 295-324.
- Winston, G.C. (1982) *The Timing of Economic Activities*, Cambridge: Cambridge University Press, 193-197.
- Wood, J.C. (1984) *Adam Smith: Critical Assessments I*, London and New York: Routledge.
- Wood, J.C. (1991) *Joseph A. Schumpeter: Critical Assessments*, London and New York: Routledge.
- Wright, J.D., Ginsburg, D.H. (2012) 'Behavioral law and economics: Its origins, fatal flaws, and implications for liberty', *Northwestern University Law Review*, 106(3), 1-58.
- Young, J.T. (1985) 'Natural Price and the impartial spectator: a new perspective on Adam Smith as a Social Economist', *International Journal of Social Economics*, 12(6/7), 118-121.
- Young, J.T. (1992) 'Natural morality and the ideal impartial spectator in Adam Smith', *International Journal of Social Economics*, 19(10/11/12), 71-82.
- Young, Jeffrey T. (1997) *Economics as a Moral Science: The Political Economy of Adam Smith*, Cheltenham, U.K.: Edward Elgar.

Zimmerman, D. (2003) 'Sour Grapes, Self-Abnegation and Character Building: Non-Responsibility and Responsibility for Self-Induced Preferences', *The Monist*, 86(2), 220-241.