INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Morality as a fundamental element in the decision-making process: A critical essay

Mauro Daniel Taborda Pereira de Sousa e Freitas

Master in Political Economy

Supervisor:
PhD Ana Cristina Narciso Fernandes Costa, Assistant Professor,
ISCTE-IUL

November, 2022

CIÊNCIAS SOCIAIS
E HUMANAS

Political Economy Department

Morality as a fundamental element in the decision-making process: A critical essay

Mauro Daniel Taborda Pereira de Sousa e Freitas

Master in Political Economy

Supervisor:
PhD Ana Cristina Narciso Fernandes Costa, Assistant Professor,
ISCTE-IUL

November, 2022

To my grandmother Astrilde,
whose love, support, and guidance
have led me to this moment

Acknowledgements

Words cannot express the deep feeling of appreciation and gratitude towards Professor Ana Cristina Costa, for her unending patience, precious insights, and most of all for taking on the challenge of helping me develop my dissertation on such an ambitious topic. Special thanks to my colleagues with whom I shared the rewarding task of writing this dissertation and with whose moral support, feedback, and discussions on the theme of this thesis, were decisive in the elaboration of this work. I would be remiss not to thank my family, especially my parents, my partner, and my very close friends. Their unending support and motivation gave me the strength to keep on working, particularly in those periods where things got tough. At last, I would also like to thank all those that came before me and without whose work I would be at a loss, unable to clearly argue my case.

Resumo

A Moralidade tem sido, nos últimos 2500 anos, uma dimensão central quando se discute o comportamento humano. A Economia, contudo, adotou um modelo simplificado de ator económico, o *Homo Economicus*, que ignora largamente a dimensão social e psicológica do processo de tomada de decisão em favor de um relato que salienta a maximização da utilidade como o objetivo final da decisão. Isto resultou em várias críticas que sublinharam, entre outras coisas, a falta de realismo ontológico deste modelo. Mais, contribuições no campo da psicologia e das ciências comportamentais sublinharam a importância da moralidade no processo de tomada de decisão. Nesta dissertação, irei exploras as contribuições de várias áreas de estudo para criticar o atual modelo de ator económico e discutir as implicações que estas conclusões têm para a elaboração de políticas.

Códigos JEL: A12, D91.

Palavras-chave: Economia Comportamental, Moralidade, Homo Economicus, Julgamento Moral, Teoria da Escolha Racional.

Abstract

Morality has been, for the past 2500 years, a central dimension when discussing human behaviour. Economics, however, adopted a simplified model of economic actor, the *Homo Economicus*, which largely ignores social and psychological dimension of the decision-making process in favour of an account that stresses utility maximization as the ultimate goal for the decision maker. This has resulted in several criticisms that have stressed, among other things, the lack of ontological realism of this model. Furthermore, contributions from the field of psychology and behavioural sciences have stressed the importance of morality in the decision-making process. In this dissertation, I will engage with contributions from several field of inquiry to criticise the current model of economic actor and discuss the implications that these findings have for policy making.

JEL codes: A12, D91.

Keywords: Behavioural Economics, Morality, Homo Economicus, Moral Judgement, Rational Choice Theory.

Index

Glossary of Acronyms

AR - Authority Ranking
CS - Communal Sharing
DPT - Dual Process Theory of Moral Judgment
DSG - Dyadic Solidarity Game
EM - Equality Matching
HE – Homo Economicus
MP - Market Pricing
RCT – Rational Choice Theory
RMT - Relational Model Theory
RRT- Relationship Regulation Theory
SES - Socio-economic status
SHH - Social Heuristics Hypothesis
SIG - Solitary Insurance Game
vmPFC - Ventromedial Prefrontal Cortex

Introduction

In a recent paper, Michael Sandel (2013) argues that market values and market reasoning have been expanding into realms that had previously been "governed by non-market norms", and cautions against quantifying every aspect of human life as it "erodes certain moral and civic goods worth caring about" (p. 121). In considering the conundrum that arises between the aspiration of economics to be a "value-neutral science of human behaviour and social choice" (Sandel, 2013, p. 122) and the necessity to propose measures which are in their nature normative/prescriptive, the author suggests that economists must turn to moral and political philosophy in order engage with the various challenges they're faced with. In a final thought, Sandel (2013) adds:

> To decide when to use cash incentives, or tradable permits, or other market mechanisms, economists must go beyond identifying the norms that inform social practices; they must also evaluate those norms. The more economic thinking extends its reach into social and civic life, the more market reasoning becomes inseparable from moral reasoning. If economics is to help us decide where markets serve the public good and where they don't belong, it should relinquish the claim to be a value-neutral science and reconnect with its origins in moral and political philosophy. (pp. 138-139)

Sandel brings to the fore the matter of economic analysis as a morally charged inquiry making a case for the necessity to take into account considerations from the fields of ethics to properly address problems that, at the core, are normative. As it stands his argument is significant on two accounts: first, (1) it deals with the transition from the realm of the descriptive to the realm of the normative - known as the humean is-ought problem that states that one cannot say what ought to be based on what is - and second it entails that humans are moral beings and as such only by acknowledging that both descriptive and normative claims must be subject to moral judgment do we offer a proper solution to the problems at hand.

It is precisely with this last point, the precise nature of humans as economic actors, that will concern me in this essay. The first point of contention will be that set of assumption about individual economic actors known has the *Homo Economicus* (HE). The concept of HE has been, since its introduction, a fundamental assumption in the field of economics and has extended to the whole of the social sciences through Rational Choice Theory (RCT) and has led to an appreciable number of theories and frameworks. Reconsidering HE's underlying assumptions will have far reaching implication for various theories and fields of inquiry in economics, but also in other fields in the social sciences.

To begin this analysis, we must consider the origin of the concept of Homo Economicus. The first proposal of the concept of HE is attributed to John Stuart Mill (1836) who stated that:

> [Economics] does not treat of the whole of man's nature as modified by the social state, nor of the whole conduct of man in society. It is concerned with him solely as a being who desires to possess wealth, and who is capable of judging the comparative efficacy of means for obtaining that end. It predicts only such of the phenomena of the social state as take place in consequence of the pursuit of wealth. *It makes entire abstraction of every other human passion or motive*; except those which may be regarded as perpetually antagonizing principles to the desire of wealth, namely, aversion to labour, and desire of the present enjoyment of costly indulgences.[1] (p. 321, emphasis added)

Since the introduction of this model of decision-maker, many economists have built on the idea of a self-interested rational utility-maximiser. During the Marginal Revolution, the concept of marginal utility as well as the law of diminishing marginal utility were introduced in economic theory, replacing the labour theory of value held by classical economists (Moscati, 2018), and became central to how economics understands economic behaviour. John Von Neumann and Oskar Morgenstern made two important contributions to the economic theory of decision-making. First, the team published *Theory of Games and Economic Behaviour* (Neumann & Morgenstern, 1944), a book that led to the establishment of the field of game theory in economics and then, in a second edition of the same book, the authors laid the groundwork for the theory of expected utility (Neumann & Morgenstern, 1947). These theories, as well as the concept of HE, are meant to be a description of how actual individuals go about making decisions. Nonetheless, as Herfeld (2018) notes in analysis of the normative turn in the Cowles Commission, this descriptive enterprise was soon abandoned when it became clear that economic actors didn't conform to HE's assumptions, and so, theorists started to employ rationality, understood as utility maximization, as the ultimate normative goal of action. The author concludes that:

> By 1954, the motivation behind studying rational behaviour had largely developed into a prescriptive undertaking, serving the end to find 'good' rules of reasoning. […] The investigation into, and the formulation of norms of rationality had become independent from the study of actual behaviour. The idea of calculating the optimal solution was too demanding or just not what economic agents in fact did. […] The rational course of behaviour could not only be suggested but also acquired as a skill and via practice. For actively improving decision-making and avoiding inconsistent behaviour, deviations from the rationality-ideal would have to be detected to subsequently bring people's decision-making on the 'rational' track. (Herfeld, 2018, pp. 45–46)

---

[1] Despite being the proponent of the concept of HE, Mill's understanding of the concept was different from the one we have today. To explain the varying behavioral proclivities of people from different countries, Mill (1882) argued that institutional arrangement influences the way people act, thus recognizing a broader set of factors influencing decision outcomes (p. 840).

Until this point we've talked about the concept of HE, particularly of its importance as a model of economic actor and a fundamental assumption in economics, but we have yet to properly define it. To this end, we refer to a recent review of critiques on the concept of HE by Urbina & Ruiz-Villaverde (2019), where the authors propose five dimensions that encapsulate the main assumptions of HE: individualism (I.e., individuals are "atomistically self-interested" (p. 65)), optimizing behaviour (I.e., constant maximization of utility), full-rationality (i.e., capacity to process all available information rationally), universality (i.e., the model of HE has universal validity, regardless of the context) and exogeneity of preferences (i.e., preferences are defined outside the process of decision and are exogenously given).

These underlying assumptions have been a point of contention for many social scientists. Max Weber (1921), for instance, argued that HE is better understood as an ideal type, but not as an accurate or even proximate description of human behaviour:

> The concepts and "laws" of pure economic […] state what course a given type of human action would take if it were strictly rational, unaffected by errors or emotional factors and if, furthermore, it were completely and unequivocally directed to a single end, the maximization of economic advantage. In reality, action takes exactly this course only in unusual cases, as sometimes on the stock exchange; and even then there is usually only an approximation to the ideal type. (p. 9)

One of the most fundamental criticisms came to the fore during the 1950's and was advanced by Herbert Simon (1957) in the book *Models of Man* where the author suggested that humans are limited in terms of the time available to make decision, the information that they can gather and their cognitive capacities. In recognizing these limitations, Simon argued that if human were to be seen as rational, the assumption could not be one of unlimited cognitive capacity, but rather it was a rationality that was constrained by personal and environmental factors. The author named this assumption bounded rationality and argued that if economics, as a field of study, is to properly understand and devise models of economic actors, it must take into consideration the psychological characteristics that motivate and constrain individual. Simon (1957) argues:

> If the principle [of bounded rationality] is correct, then the goal of classical economic theory to predict the behaviour of rational man without making an empirical investigation of his psychological properties is unattainable. […] He behaves rationally with respect to this [simplified] model, and such behaviour is not even approximately optimal with respect to the real world. To predict his behaviour, we must understand the way in which this simplified model is constructed, and its construction will certainly be related to his psychological properties as a perceiving, thinking, and learning animal. (p. 199)

Gigerenzer & Selten (2001) expand on the concept of bounded rationality presented by Simon (1957) and argue that bounded rationality is concerned with the cognitive processes that lead up to the decision and not just with the outcome of the decision, while doing away

with the notion of optimizing behaviour, preferring, alternatively, explanations based on heuristics, thus avoiding unrealistic assumptions that do not occur in “real-world situations” altogether (p. 4). Gigerenzer & Selten (2001) argue conclusively by saying that:

> Bounded rationality is ... not simply a discrepancy between human reasoning and the laws of probability or some form of optimization. Bounded rationality dispenses with the notion of optimization and, usually, with probabilities and utilities as well. It provides an alternative to current norms, not an account that accepts current norms and studies when humans deviate from these norms. Bounded rationality means rethinking the norms as well as studying the actual behavior of minds and institutions. … Bounded rationality is not an inferior form of rationality; it is not a deviation from norms that do not reflect the structure and representation of information in environments. Theories of bounded rationality should not be confused with theories of irrational decision making. (p. 6)

Despite the relevance of these contributions, economics still retains HE as both a normative and descriptive model of economic actor. Much of this “resistance” to integrate empirical findings in the field of psychology and cognitive sciences into more descriptively accurate models of human decision makers can be traced back to Milton Friedman (1953) methodological and epistemological essay, *The Methodology of Positive Economics*. In this essay, regarded by some scholars as the “most influential methodological tract of modern times” (Hausman, 2007, p. 143), Friedman proposed an instrumentalist approach to theory building in economic theory arguing that assumptions are instruments, neither true or false, and as such their truth-value (realism) is unimportant, since the goal of theory is to aptly predict relevant outcomes (Caldwell, 1980; Sent, 2004). Caldwell (1980), drawing from the field of ontology, criticizes this proposition on two accounts: (1) the proposition sidesteps the epistemological goal of scientific inquiry - if explanation is the goal of science, then Friedman’s instrumentalist approach is “considerably weakened” (p. 369) and (2) the truth-value of propositions is ignored - the author point out that “instrumentalists fail to comprehend that though we may not know whether a theory is true or false, it in fact is true or false” (pp. 369-370). Simon (1963) made a similar point previously proposing, in opposition to what the author coined as Friedman’s “principle of unreality”, a “principle of continuity of approximation”, that asserted that “if the conditions of the real world approximate sufficiently well the assumptions of an ideal type, the derivations from these assumptions will be approximately correct” (p. 231).

Behavioural economics became an independent sub-field of inquiry within economics, bringing about the necessary inputs regarding which psychological factors underlie our decision-making process, and how environmental and personal factor influence decision outcomes (Sent, 2004). However, when we speak of behavioural economics, we must concern ourselves with an important conceptual distinction, there is not one, but two ‘strains’ of

behavioural economics: old behavioural economics and new behavioural economics. While new behavioural economics retains HE as the normative standard and develops its models around it, old behavioural economics rejects HE not only descriptively, but also normatively. Gigerenzer (2018) argues that the “irrationality argument” has become the central contention of the new behavioural economics, and that “listing and modelling deviations from rational choice theory has become all that keeps it erect” (p. 329). The author adds, *contra* the heuristics and biases program, that “Heuristics, framing, defaults, and emotions may be rarely needed in a world of certainty but are essential tools for dealing with uncertainty” (Gigerenzer, 2018, p. 329).

We are left at a crossroads between the field of economics, that has become set in its ways and incapable of reforming its core assumption about economic actor, even if that signifies a lost in explanatory power, and the field of behavioural economics that has built its main theoretical frameworks around the same assumptions that had led to the foundation of this field in the first place. To this last point, Altman (2004) points to the work of Kahneman and Tversky, noting that hey developed and alternative theory (prospect theory) to RCT but they didn’t propose an “alternative normative theory”, rather maintaining the same neoclassical normative standard of rational behaviour as optimal behaviour, where “deviations from these norms are considered to be errors and biases and indicative of irrationality" (p. 8). Sent (2004) also makes this point, arguing that whereas Herbert Simon had insisted on the reformulation of the model of human behaviour, latter theorists, such as Kahneman and Tversky, "use the rationality assumption of mainstream economics as a benchmark from which to consider deviations" (p. 750). We can follow Gigerenzer & Selten (2001) and do away entirely with any normative theories, by adopting a descriptive stance, namely by describing the decision/making process rather than continuing to employ normative stances, that focus exclusively on the outcomes of the decision, such as is the case of the optimizing and utility maximizing HE. Here we would do well to look beyond classic rationality, and towards different explanations that look at cognitive processes and environmental features. Morality might just be the place to start.

The usefulness of a framework that properly describes how morality elicits certain responses from individuals has been shown to be of paramount importance. Brain imaging studies have shown that there is a significant overlap between the brain circuitry that process complex moral decisions and complex economic decisions (Shenhav & Greene, 2010). The same overlap is observable even in simple economic decisions (Hutcherson et al., 2015). This overlap might be the evolutionary result of a part of the brain that evolved to generate the necessary affective valence necessary to deal with the complexities of social life and

guarantee survival (Damásio, 2018) getting “repurposed” to deal with a new scenario and function for which it had not evolved to do, that is, to make economic decisions.

Such an approach might help us deal with problems that have plagued public and economic policy in the last decades. Bowles (2016) argues in his book *The Moral Economy: Why good incentives are no substitute for good citizens,* that some policies, informed by the assumption of HE, the introduction of monetary incentives has crowded out intrinsic motivation, thus depleting social capital and ethical motivations. The author argues that when we put a monetary price on certain practices, we signal individuals what exactly is the cost of incurring in that practice and thus turn the problem into a matter of cost-benefit calculations (i.e., Am I willing to pay X amount of euros to do this?) rather than a matter of intrapersonal and interpersonal deliberation. Regarding the possibility of these findings fostering a change in the way public policy, economic policy and even laws are made, the author offers the following consideration:

> How should policy makers respond to the realization that while both economic incentives and ethical and other-regarding motives are necessary for effective policy, the former may diminish the latter? If both sources of motivation are taken into account, then policy makers may reasonably consider giving economic incentives a more limited role in their policy packages. If incentives undermine social values, yet incentives and social values are both needed, then it would seem to follow that one ought to make less use of incentives than one would in the absence of this crowding-out problem. (Bowles, 2016, p. 5)

I will not engage with normative or prescriptive arguments about what ought to be done in each scenario, or what makes a determinate action moral or otherwise, as that is outside the scope of the present effort. Instead, I will call into question the assumptions that underlie the Homo Economicus model, mainly that of the exogeneity of preferences, and I will explore how moral considerations affect the decision-making process, by mobilizing theoretical contributions in the fields of social psychology. and behavioural sciences. This way will allow me to assess, albeit tentatively, the influence of morality in the decision process, namely how different moral motives affect our decisions and how different environmental and situational factors favour different types of response. This approach allows us, at least in theory, to describe and predict the behaviour of individuals, if we know which relational models and moral motives are relevant to the scenario in question. Furthermore, this contribution will allow a better understanding of the role of morality as a factor in the decision-making process. The conclusions drawn here have significant implications for the field of public policy and may open new avenues for intervention and investigation in this field, mainly, but not exclusively, where conflicts and moral considerations are particularly salient, as we will see later.

In the following sections of this essay, I will explore the major criticisms that have been made of the concept of HE, with a particular focus on the matter of preference exogeneity, as this assumption is in direct tension with the critique I want to develop here. Furthermore, I will look at recent theoretical contributions in the field of moral psychology and cognitive sciences to understand what the theoretical implication are for a model of human decision-maker that considers moral judgments and commitments. In the succeeding I will explore the implication of these findings for economic theory, public policy, and other scientific areas where the assumption of HE has been adopted. To conclude, I will ponder the implication of a new model of economic actor that that into account the moral dimension of decision-making. Furthermore, I will expand on the limitations of the approach I have conducted here and ways to overcome these limitations to build a decision-making framework that incorporates the moral and ethical dimension, as well as some of the possible theoretical approaches and lines of research that this decision maker model might generate.

Chapter 2

A (Very) Problematic Assumption – Exogeneity of Preferences

RCT's most problematic assumption, in terms of its methodological implications, is the assumption of the exogeneity of preferences (Fehr & Rangel, 2011; Krackhardt, 1998; Bazerman & Malhotra, 2006). The violation of this assumption, that underlies most Econometrics, and therefore most econometric models, results in "devastating effects on statistical tests" (Krackhardt, 1998, p. 247).

First, we must consider the implications that such an ontological commitment necessitates, such as the exclusion of institutions as an important agent in shaping decisions. Bazerman & Malhotra (2006) call our attention to a set of five economic assumptions - which they have labelled as "myths" - and argue that there is a growing body of evidence, especially in the field of social psychology, that directly contradict the by now classic dictum offered by Stigler & Becker (1977), that "[…] one does not argue about tastes for the same reason that one does not argue over the Rocky Mountains – both are there, will be there next year, too, and are the same to all men" (p. 76). The former authors argue that these "myths" have been used as policy guidelines and resulted in the destruction of "value in society" (Bazerman & Deepak Malhotra, 2006, p. 264). Although all five of these economic assumptions are relevant, two of them are of particular importance to the present critique: (1) the assumption of stability and consistency of preferences and (2) and the conception that the only outcome data, but not mechanism data, is acceptable as empirical evidence. Regarding the (1) stability and consistency of preferences, the authors conclude that the literature on this subject indicates that preferences are shaped by "framing effects, the omission bias, and the status quo bias" (Bazerman & Malhotra, 2006, p. 277) among others, meaning that preferences are situationally and relationally dependent and not defined *a priori*. Regarding the second point, (2) the authors argue that the insistence on the part of policymakers, economists as well as other political decision makers in presenting outcome data (i.e.: a 'smoking gun'), at the exclusion of mechanism data to identify a specific problem has resulted in detrimental, if avoidable, consequences. To further their point, Max H. Bazerman & Malhotra (2006) recall a testimony given in 2000 to the SEC by Bazerman and Loewenstein, wherein they warned the SEC Chairman, Arthur Levitt, that the structure of the accounting industry in the USA made auditor independence "an impossible goal" (p. 275) and advised the SEC to conduct a series of reforms to help mitigate conflicts of interest. The two economists illustrate this unwillingness to accept mechanism data, by pointing out the verdict of the SEC: "The SEC asked us and others to produce the smoking gun of a specific audit corrupted by conflicts of interest – they

wanted evidence of the economic effect of psychological bias; without one, they refused to institute needed reforms." (Bazerman & Malhotra, 2006, p. 275). It wasn't long until accounting scandals, such as Enron's and WorldCom's, broke out in the period between 2001-2002, laying bare the conflicts of interest between companies and accounting firms that compromised the veracity of audits. The authors list several examples stressing how cognitive processes are sometimes enabled by particular institutional and legislative environments, thus justifying mechanism data as credible empirical evidence that can help us prevent future negative outcomes and go as far as stating that "In many situations, psychological research on mechanisms predicts crises more efficiently than economic or other research on outcomes. Thus, the myth that outcome data are the only reasonable guide for action can be disastrous.

Fehr & Rangel (2011) explain that, in economics, constraint-based explanations of changes in choice behaviour have been preferred over preference-based explanations for one fundamental reason: preference-based explanations would introduce too many "free variables" (i.e.: degrees of freedom), explaining any changes in behaviour by invoking a "direct preference" (p. 398) for that behaviour, resulting in a theory able to explain everything and nothing at the same time. Fehr & Rangel (2011) argue that this problem can be entirely eschewed if we are able to constrain the set of preferences that can "reasonably be put forward as an explanans" and, to that end, two obstacles that have impeded on preference-based explanations, must be surpassed: (1) the inability to "infer causality from non-experimental data" due to inadequate econometric tools and methodologies; and (2) a poor understanding of the social and psychological categories that influence preferences stemming from the fact that the "standard approach – expected utility theory, exponential discounting theory and the assumption of purely self-interested preferences – failed to capture important preference patterns" (p. 398). The first point (1) has been mitigated by statistical, econometric, and methodological innovations (i.e.: quasi-experimental research designs such as regression discontinuity design, propensity score matching, instrumental variables), which have been developed in the last 20 years. In respect to the second point (2), a body of research has been developing, on matters such as risk, time, and social preferences, which allow us to provide "constraints on the types of preference that one can reasonably invoke" (Fehr & Rangel, 2011, p. 399). The authors also defend that the conventional approach used by economists also provides "many degrees of freedom", pointing to game theory's ability to 'explain' a variety of different phenomena by "tweaking assumptions about the extensive form of the game, such as the structure of moves, the choice variables […] or the asymmetry of information" (p.399), even when there is no evidence to support those change in the constraints. In this regard, the authors conclude their argument for a preference-based explanation of decision, with a claim that:

> Thus, taken together, there are neither methodological nor substantive arguments that favour changes in constraints over changes in preferences as explanatory factors. Ultimately, it is an empirical question which would provide the better explanation. However, explanations in terms of changes in preferences will be convincing only if we have clean measures of preference changes and if we gain a better understanding of the factors that affect preferences. In this regard, economics, and the social sciences more generally, are still in their infancy. We still know relatively little about how economic, social and biological factors, and the interactions between them, shape preferences (Fehr & Rangel, 2011, p. 400, emphasis added)

Findings in experimental economics conducted in cross-cultural samples agree with the arguments presented up until this point. Henrich et al. (2001), conducted a series of economic games[2] in 15 small-scale societies to determine whether the standard economic model of rational agent emerged in the distinct socio-cultural settings. The authors report five main findings: (1) the data didn't support the existence of the standard economic agent (i.e.: Homo Economicus) in any of the cultures studied; (2) there was "considerably more behavioural variability across groups" (Henrich et al., 2001, p. 73) than what was expected from previous studies; (3) a significant share of the behavioural variation across societies is explained by the level of market integration and variation in economic organization; (4) individual economic behaviour and demographic variables fail to explain within-group behaviour as well as across groups and (5) individuals rely on successful strategies they apprehended from their social settings and apply them in the context of economic games. These findings further substantiate the preference endogeneity hypothesis, given that even in one-shot economic games, where the monetary pay-off's would be maximized if players chose not to cooperate – or to cooperate as little as possible – while bearing no negative reputation costs, since interactions between players are restricted to the experimental setting, individuals contribute a substantial proportion of their entitlements, with the mean offer ranging from 26% in the low end to 58% in the high end, thus clearly deviating from the "ideal" standard economic agent. The researchers conclude this paper by pointing out that "preferences over economic choices are not exogenous […] but rather are shaped by the economic and social interactions […]." and label as "questionable" any attempts to predict outcomes of policy and institutional changes that fail to take into consideration behavioural change (Henrich et al., 2001, p. 77).

Hitherto I've been making the case that preferences are shaped by the endogenous factors (i.e.: cultural norms, values, moral codes) that surround us. Institutions, in this account, are of key importance and understanding in what ways they influence behaviour is fundamental. In a

---

[2] The authors employed a standard set of one-shot economic games (Ultimatum Game, Dictator Game, Public-Goods Game), that eliminate the relational component of the decision-making process, to observe if a pattern of *homo economicus* like behavior emerged.

literature review on the topic of the behavioural influence of institutions, Iris Bohnet (2006) argues that institutions perform six tasks: (1) create incentives, (2) help coordinate behaviour, (3) provide information regarding procedures (i.e.: individuals have different procedural preferences, and acceptance of different outcomes is conditional on the alignment between individual preferences and institutional procedures), (4) helping self-selection (i.e.: aid individuals and companies in choosing determinate institutional environments that are aligned with their own values) (5) and allowing causal attribution (i.e.: providing information about how outcomes came to be). Institutions can affect behaviour in one further manner: (6) they can directly influence preferences (e.g.: institutions can shift the locus of control from the inside to the outside, in turn reducing the importance of intrinsic motivation) rather than simply creating incentives for a certain behaviour (see Kleinjans & Gill (2018) for a recent exploration of the role of institutions in the emergence of an internal/external locus of control). This last task performed by institutions is of marked importance particularly for public policy since, as will be discussed in a latter chapter, there is strong evidence that economic incentives and behavioural interventions based in incentives fail to foster long lasting behavioural change and can crowd out intrinsic motivation, deepening the problem rather than solving it (Bohnet, 2006, pp. 225-228).

This function of institutions has gone largely unrecognised in the literature. Douglass D. C. North (1991), in a pivotal paper about institutions and institutional change defines institutions as:

> The humanly devised constraints that structure political, economic and social interaction. They consist of both informal constraints (sanctions, taboos, customs, traditions, and codes of conduct), and formal rules (constitutions, laws, property rights). Throughout history, institutions have been devised by human beings to create order and reduce uncertainty in exchange. […] Institutions provide the incentive structure of an economy; as that structure evolves, it shapes the direction of economic change towards growth, stagnation, or decline. (p. 97)

This definition proposes that institutions serve to help with problems of economic coordination by presenting formal and informal constraints on the set of possible choices. This definition has found wide acceptance in the field of political economy, however, as was latter recognized by North, the language of constraints and incentives doesn't fully capture how institutions influence our preferences to produce social and economic outcomes. This is the case particularly with informal institutions. In a critical overview of North's work institutions and institutional change, Faundez (2016) argued that despite recognizing the importance of

informal institutions in some early writings[3], North fails to present an explanation for their origin and normative power. More recently, Douglass North (2005) revisited the concept of informal institutions and in his book, *Understanding the Process of Economic Change*, proposed that informal institution influence our preferences through cognitive processes that are informed by cultural values. According to North (2005), the development of cognitive processes from cultural values can be explained by the interaction between mental models, belief systems and institutions. Faundez (2016) offers a simplified account of this process:

> Mental models are the predictions the mind makes about the environment; these models are revised or confirmed depending on the feedback individuals get from experience; mental models that are repeatedly validated become beliefs; and finally, when beliefs are widely shared they become belief systems. (p.391)

This process suggested by North, puts informal institutions front and centre in the discussion of decision making, and makes clear the relation between the expectations we formulate about the social world, the adaptive nature of these expectation given their success and their capacity to evolve beyond individually held norms into socially shared belief systems. This process begins as a cognitive process in which the individual must consider environmental factors to decide but at the same time, also has to consider previous experiences and draws from lessons learned and the applicability of such learnings to the present situation. North (2005) argues this socio-cognitive approach to learning and decision-making, emphasizes the "important cognitive role of social institutions" (p. 36). Furthermore, the author stresses the role of informal institutions:

> The intimate interrelationship of beliefs and institutions, while evident in the formal rules of a society, is most clearly articulated in the informal institutions – norms, conventions, and internally held codes of conduct. These informal institutions not only embody the moral codes of the belief system, which tend to have common characteristics across cultures, but also embody the norms particular to individual societies, which are very diverse across cultures. While formal institutions can be changed by fiat, informal institutions evolve in ways that are still far from completely understood and therefore are not typically amenable to deliberate human manipulation. (North, 2005, p. 50)

Similarly to North, Schmidtz (1994) argues that institutions have moral and normative powers that affect our preferences over the set of possible outcomes and helps in identifying which of them are desirable outcomes. The author argues:

---

[3] Faundez (2016) suggests that North had recognized but not fully explained the cognitive and behavioral influence of informal institutions in *Institutions, Institutional Change and Economic Performance.*

> Institutions act, if they act, by moving people, whereas people act by moving themselves. If institutions serve the common good at all, they do it not by acting but rather by influencing the behavior of human agents. Therefore, if morality serves a purpose, and especially if it serves a purpose as it works through interpersonal constraints embedded in social structure, it must serve its purpose through the effect it has on the opportunities, incentives, and expectations of individual agents. (Schmidtz, 1994, p. 237)

Our preferences are shaped by the laws and political institutions (formal institutions) but also by social norms, religious beliefs as well as moral values and codes (informal institutions) thus making it clear that endogenous factors have a prevalent effect over the decision-making process. If we are to properly understand the behavioural and cognitive influence of these institutions, in particular the role of morality in shaping decision outcomes, we have to engage with current theories in the field of moral psychology to understand the extent to which moral consideration factor into the decision process, whether such a relation is mediated by situational and relational factors and which factors relevantly impact decisions.

Chapter 3

Insights from Moral Psychology – Cognitive mechanism and psychological phenomenon

We have established in the previous chapter that morality is a cornerstone in classical political economy, and that its theoretical importance is pivotal, particularly if we are to fully understand the implications of the proposed theories, such as is the case of both of the previous authors. In doing so, we have identified a particular tradition in Political Economy that substantiates our claims, insofar as our claims are purely in terms of the relevance of the discussion for the field. However, if we are to present a model of decision making that recognizes the centrality of morality, the argument cannot occur purely at the normative level, as that would reduce this essay to a mere argumentative exercise. For that reason, in this chapter we will explore the existing scientific literature regarding the role of morality in the process of decision making.

3.1. Morality as Regulation – Relational Models Theory and Relationship Regulation Theory

The first question we ought to deal with is, first and foremost, the relative importance of moral motives in economic decision making. There is evidence to support the claim that morality is, in fact, a fundamental aspect of decision making (Brodbeck et al., 2013; Kugler et al., 2021). Brodbeck and his colleagues (2013), took issue with the inability of behavioural economics and game theory to account for indirect reciprocity, and sought to identify the underlying psychological mechanisms that influence economic decision-making, especially in one-shot economic games, by drawing from two theoretical framework in moral psychology namely Relational Models Theory (RMT) (Fiske, 1992) and Relationship Regulation Theory (RRT) (Rai & Fiske, 2011) as well as from Haidt's intuitionist approach to moral judgement (Haidt, 2001; Haidt, 2007).

RMT was first proposed by Alan Fiske (1991, 1992) as a framework for understanding human social cognition and behaviour. This theory is premised on the idea that the nature of relationships - the relational model in place - determines the way in which people interact with each other. This framework has been applied in many fields including psychology, anthropology, political science, sociology and, more importantly for the present work, economics (Brodbeck et al., 2013; Kugler et al., 2021; Stofberg et al., 2021). The theory

proposes four distinct relational models[4]: Communal Sharing (CS), Authority Ranking (AR), Equality Matching (EM) and Market Pricing (MP). Communal sharing relational models are premised on the notion of equal division and access to resources by all members of a given group. As stated in Fiske (1992) “the members of a group or dyad treat each other as all the same, focusing on commonalities and disregarding distinct individual identities” (p. 690). Authority ranking relational models are based on the notion that within a group hierarchy, individuals have differing levels of power and status. AR is markedly culture specific, with different cultures using different “space, time, magnitude and force metaphors for rank” (Fiske, 2004a, p. 95) but these categories are easily and intuitively apprehended. Equality matching relational models builds relations around “additive interval differences” (Fiske, 2004b, p. 5) that entail “one-to-one balanced in-kind reciprocity in social interactions” (Brodbeck et al., 2013, p. 3). Examples of EM include taking turns, side-by-side comparisons, and voting (Fiske, 2004a, p. 103). Market pricing relational models are based on a model of “proportionality in social relationships” (Fiske, 1992, p. 692), where people construe social reality in terms of ratios. In MP relational models it is common for individuals to subsume all the relevant features of a given interaction into a utility value, allowing them to perform cross-situational comparisons of many “qualitatively and qualitatively diverse factors” (Fiske, 1992, p. 692).

Later, Rai & Fiske (2011) developed RRT, which builds on RMT, by “integrating moral psychology into social relational cognition” (p. 57). The authors argued that this integration was necessary for two main reasons: (1) moral decisions need to be understood in the context in which they are made (i.e.: social-relational morality) and that (2) morality is an in-built characteristic of social relations (and so they go hand in hand). Regarding the first reason, (1) the author argue that while cognitive-developmental and rationalist approaches to scientific moral psychology focus on the individual’s capacity to reason morally, seeing deviations from idealized behaviours as social biases, social-interactionists make a distinction between moral judgements and social conventions, arguing that while the former are universal rules that ought to be enforced, regardless of the social context, the latter is context specific, and is “rule contingent” (Rai & Fiske, 2011, p. 58). The second reason, (2) is that morality is embedded in social relations. The authors argue that the belief that moral motives are fundamental in guiding our other-regarding judgements and behaviours in relation to prescriptive relational models for social relations is a fundamental claim of moral psychology. In such a relational-social paradigm of morality, morality “functions to facilitate the generation and maintenance of long-

[4] Fiske & Haslam (2005) performed a confirmatory factors analysis on the four models proposed by RMT and found that these models "represent distinct and generally coherent relational forms that may correlate systematically" (p. 248) thus confirming their validity as a taxonomy of social relationships.

term social-cooperative relationships with others" (Rai & Fiske, 2011, p. 59) (See also Fiske (2002) for an exploration of the role of moral emotions for the maintenance of social relations). Rai & Fiske (2011) propose a set of four moral motives – unity, hierarchy, equality, and proportionality – that map directly onto the four relational models that were proposed by RMT. The moral motive in CS relational models is Unity, and it is "directed toward caring for and supporting the integrity of in-groups through a sense of collective responsibility and common fate" (Rai & Fiske, 2011, p. 61). Unity promotes a strong feeling of belonging to the in-group and blurs the lines between group membership and self (e.g.: This feeling of profound belonging is well illustrated by the Ubuntu philosophy, whose namesake directly translates to "I am because we are"). The moral motive in AR relational models is Hierarchy, and it is "directed toward creating and maintaining linear ranking in social groups" (Rai & Fiske, 2011, p. 63). Even tough Hierarchy presupposes vertical enforcement of power, it recognizes that those at the top of the hierarchy (i.e.: leaders, superiors) have a "a sense of pastoral responsibility" (Rai & Fiske, 2011, p. 63) towards their subordinates. The moral motive in EM relational models is Equality, and it is "directed toward enforcing even balance and in-kind reciprocity in social relations" (Rai & Fiske, 2011, p. 63). This moral motive provides the motivation to maintain "tit-for-tat" relations, necessitating direct and proportional reciprocity. The moral motive in MP relational models is Proportionality, and it is "directed toward calculating and acting in accord with ratios or rates for otherwise distinct goods to ensure that rewards or punishments for each party are proportional to their costs, contributions, effort, merit, or guilt" (Rai & Fiske, 2011, p. 64). The Proportionality moral motive is the underlying motivation in consequentialist decision-making, as it allows individuals to interpret situation in terms of ratios (i.e.: how many are saved vs how many people are sacrificed in sacrificial dilemmas.)

Building on these theoretical frameworks, Brodbeck et al. (2013) present three (3) propositions regarding the behavioural influence of moral motives in one-shot economic games: the first proposition, (1) suggests that particular other-regarding behaviours in one-shot economic games are "determined by the kind of moral motive that is activated (or salient) within an actor's mind" (p, 4); the second preposition, (2) argues that the particular behaviour expressed in one-shot economic games is "determined by the kind of moral motive that is – consciously or unconsciously – activated (or salient) within an actor's mind" (p. 5), and, finally, the third proposition, (3) maintains that outside of a social setting, "decision making behaviour" (p. 5) is unaffected by moral motives (regardless of them being expressed consciously or unconsciously). To test whether their theoretical propositions had any bearing on particular behaviours expressed in one-shot economic games, the authors conducted a 2 x 2 research design, where the first independent variable was the decision situation, namely if it was in a

social setting or in a non-social setting (solitary setting), and the second independent variable was the way in which the moral motives were presented during the experiment, either by explicitly framing (conscious activation) the experiment in terms related to the pertinent moral motives or by subliminally priming participant. Furthermore, the authors introduced a "novel game paradigm" (Brodbeck et al., 2013, p. 5), that was termed Dyadic Solidarity Game (DSG), which is a simplified/altered version of the 3-player Solidarity Game (Selten & Ockenfels, 1998). In a DSG, the participants have a fixed 2/3 chance of winning and a 1/3 chance of losing "determinable financial resources" (p. 5) and given this information they can allocate a set amount of money either to "Amount A" (i.e.: the amount of money a participant will receive in case she wins) and "Amount B" (i.e.: the amount of money the other participant will receive in case they lose). In this game, participants have to consider and manage two different risks: (1) a probabilistic risk and (2) a relational risk. In the four experiments conducted for the purpose of this paper, Brodbeck et al. (2013) tested the effect of two of the four moral motives, Unity and Proportionality moral motives, and their associated relational models, CS relational models and MP relational models on the decision making in one-shot economic games. In experiment 1, which featured a social context where participant were exposed to framing effects, the authors verified that the outcome for both experimental conditions were "inconsistent with the maximum of expected utility" given that the Amount B was, on average, "significantly greater than 0" (Brodbeck et al., 2013, p. 7) and confirmed their first hypothesis, that is, participants framed with Unity moral motives allocated more money to Amount B than participants framed with Proportionality moral motives, thus proving that different moral motives elicit different behavioural responses in one-shot economic games. In experiment 2, which featured a social context where the participants were subliminally primed, the authors were able to confirm their second hypothesis, meaning that the behavioural effects of moral motives are induced either through conscious activation (framing) or subconscious activation (subliminal priming). Experiments 3 and 4 followed the same logic, with the former using framing and the latter using subliminal priming but employ a non-social (solitary) version of the DSG, known as Solitary Insurance Game (SIG), to test the third hypothesis, which maintains that decisions in the SIG are not influenced by moral motives, since there is no relational risk to consider, whereas the DSG is affected by moral motives. The results from both experiments entirely support the third hypothesis since the amount of money allocated to Amount B doesn't differ significantly from the control group to those exposed to both moral motives either through conscious or unconscious activation. This finding indicates that outside of social settings, moral motives, which are a relevant feature of relationship regulation, do not impact economic decisions and probabilistic risk becomes the only relevant factor for decision.

A recent study by Kugler and her colleagues (2021), which sought to replicate and extend Brodbeck et al. (2013) work, was able to reproduce the previous findings and assert that in social settings, moral motives are paramount for economic decision making. The authors have also found that in ongoing relationships the underlying moral motives took precedence over those that were introduced, through priming, during the experiment. There are two important differences between these studies, namely (1) the way in which the non-social situation was set up – in Kugler et al. (2021) this interaction was with a non-human participant while in Brodbeck et al. (2013) this "interaction" was with oneself – and (2) the introduction of a "prior interaction" condition to assess "which kind of information individuals used to infer morally acceptable behavior" (Kugler et al., 2021, p. 4), thus making a distinction between the effect of the moral framing in anonymous and non-anonymous social interactions. The authors followed a 3 x 2 study design. The first independent variable is the decision situation, operationalized as three "relationally different situations" (Kugler et al., 2021, p. 7): (a) anonymous social one-shot interactions, (b) non-anonymous social ongoing interactions and (c) anonymous non-social one-shot interaction (this relational situation entails knowingly playing a DSG with an algorithm). The second independent variable is the framing of the situation regarding the moral motive (either Unity condition or Proportionality condition). Accordingly, the authors defined five experimental hypothesis all of which were corroborated by the results of the study. Hypothesis 1a, 1b and 1c pertain to the effects of the framing of different moral motives, such that Hypothesis 1a proposed that in anonymous social one-shot interactions Unity framing evokes more solidarity than Proportionality framing; Hypothesis 1b proposed that in non-anonymous social ongoing interaction the framing has no effect on economic decision; and Hypothesis 1c also suggests that framing has no effect on economic decision in anonymous non-social one-shot interactions. Hypothesis 1d proposed that there is an interaction effect between the decision situation and the situational moral motive framing, and this hypothesis was confirmed given that the effects of moral motive framing were only found in anonymous social one-shot interactions where there were no previously established relationship, which would be a much more salient cue than moral motives (Rai & Fiske, 2011).The last hypothesis, Hypothesis 2, proposed that levels of solidarity (i.e.: the amount of money allocated to Amount B) varied across situation and framing, with expected high levels of solidarity for participant in non-anonymous social ongoing interactions, regardless of the moral motive framing, and participants in anonymous social one-shot interactions with a unity frame would show high levels of solidarity, and expected low solidarity for participants in anonymous non-social one-shot interactions, regardless of the moral motive framing, and participants in anonymous social one-shot interactions with a proportionality frame.

Both previous studies, make important contributions to the discussion on decision making, as they go beyond the common conceptualization of the individual as a rational, self-interested and utility maximizing entity, to give an account of an individual that takes into account situational and relational factors as well as the customary "probabilistic risk" associated with making the decision that benefits one the most at the expense of social others. In the account presented by Brodbeck et al. (2013) and Kugler et al. (2021), moral motives play a key role in anonymous social one-shot settings, where pre-existing relational cues do not exist, and this finding is of extreme relevance particularly in fields such as public policy (e.g.: helping in alleviating problems related to collective action), with a particular focus where interactions between different actors are of the aforementioned nature.

Another fundamental point is that there are differences between the proportionality moral motive and pure self-interest. The authors have found that the difference between the choices made by participants in the control condition and the Proportionality condition, were not significantly different (Brodbeck et al., 2013; Kugler et al., 2021), meaning that the "default" response is subject to market pricing relation model rationale[5]. This finding, however, does not constitute an equivalence between selfishness and the Proportionality moral motive, because the latter and the related MP relational models, create "commitments to (the terms of) relationships […] [where] violations of these symbolic acts evoke strong emotions, arouse moral sentiments, and motivate sanctions." (Fiske, 2004a, p. 115). This means that unlike selfishness, which is by definition asocial, Proportionality is necessarily social as it creates commitments that must be fulfilled. To this point Brodbeck et al. (2013) argue that:

> The constructs explicitly exclude any supposition that self-interest or maximization of individual benefit is a defining, necessary, or distinctive feature of the Market Pricing relational model or the Proportionality motive. The idea that self-interest or the maximization of individual benefits is intrinsic to Market Pricing relational models or Proportionality moral motives seems plausible from folk psychology and from economic theory, but it is not part of RMT and RRT. (p. 117)

Furthermore, the participants in this study did not behave in the way self-interested agents would (i.e.: defecting and thus, not contributing to funds to the other player), allocating a significantly higher amount of money to the other participant than what was predicted by a "zero solidarity" (selfish) condition. Kugler et al. (2021) argue that due to their research paradigm, they cannot completely rule out that the "proportionality framing was confounded with pure self-interest or egoistic behaviour" (p. 14), but comparisons between the allocation

[5] Brodbeck et al. (2013) have argued that the salience of money (the more salient the monetary cue is) in economic games might signal the participants as to what is the most appropriate moral motive to adopt in a given situation (p. 16).

of money to “Amount B” in anonymous social one-shot interactions with proportionality framing and allocations of money to “Amount B” in both framings of the anonymous non-social one-shot interactions, have provided evidence that in the first condition, participants allocated a significantly higher amount of money than those on the latter conditions, reinforcing the notion that moral motives, and not selfishness, affect the choices of participants.

3.2. Intuitive and Reflective Morality – Dual process of moral judgement

As we’ve discussed in the previous section, we can look at morality in terms of its sociopsychological manifestations, particularly in the setting-up and maintenance of relationships and institutions. This offers us precious insight into the influence of moral considerations (moral motives) to the decision-making process. Explanations offered by social psychology advance our comprehension of how relational and situational factors influence our perception of a given phenomenon and how we factor these considerations into a decision. However, moral judgement and moral reasoning can be studied, and thus understood, at a lower level, at the level of the individual and the functioning of the brain. To that effect, we can follow in the footsteps of researchers in the field of neuroeconomics (Bechara & Damasio, 2005; Fehr & Rangel, 2011) and consider findings in the field of neurology and neuropsychology regarding the relation between cognitive processes, morality and decision-making and what are the implication for theorizing in economics.

In a landmark study, H. Damásio et al. (1994) conducted a “neuroanthropological” (p. 1103) imaging study on the skull of the now famous Phineas Gage, a nineteen-century railroad worker that experienced dramatic behavioural change in the aftermath of an accident that damaged his brain. After running various simulations of the possible trajectory of the tamping iron, the instrument responsible for the brain lesion, the authors established that both the brain damage that was observed and the coincident and abrupt behavioural change that was reported by Phineas’s peers and family, fit the pattern of patients with frontal lobe damage, particularly in the “anteromedial region of both frontal lobes” (H. Damásio et al., 1994, p. 1104). This type of brain damage compromises the ability to make “rational decision in personal and social matters, […] [as well as the] processing of emotion” while the “ability to tackle the logic of an abstract problem, to perform calculations, and to call up appropriate knowledge and attend to it remains intact” (H. Damásio et al., 1994, p. 1104). These findings are extremely relevant, precisely because they corroborate the hypothesis that emotions are central to the decision-making process. The authors argue, in line with their findings that “emotion and its underlying neural machinery participate in decision making within the social domain and … [raises] the possibility that the participation depends on the ventromedial frontal region” (H.

Damásio et al., 1994, p. 1104). In a follow up study by Bechara et al. (2000), vmPFC patients were asked to play a gambling game to assess why individual with this type of brain lesion present impaired decision-making capabilities. The authors argue that this impairment comes down to an inability to control cognitive impulsiveness:

> It is possible that VM (ventromedial prefrontal cortex) lesion patients have cognitive impulsiveness. That is, when the patients are presented with a deck of cards with a large immediate reward but with delayed costs, the patients seek the reward. These VM lesion patients seem unable to delay the gratification of the reward for too long, as indicated by their tendency to return quickly and more often to the decks that yield high immediate reward but an even larger future loss." (Bechara et al., 2000, p. 2199)

More recently, Young et al. (2010) studied moral judgements in ventromedial prefrontal cortex (vmPFC) patients and found out that despite preserving "general intelligence, logical reasoning, and declarative knowledge of social and moral norms" (p. 846) and presenting similar moral judgements to the control group, these patients responded abnormally in situations where they had to judge harmful actions based on their intentionality (attempted harm *vis-à-vis* accidental harm). Unlike healthy individual, vmPFC patients judged instances of accidental harm more harshly than instances of attempted harm, showing a "neglect [for] the protagonist's negative intention, focusing instead on the action's neutral outcome", resulting in "unusually lenient moral judgments of failed attempts to harm" (Young et al., 2010, p. 848). The inability of vmPFC patients to account for intentionality in moral judgements, where intent is a decisive factor in decision-making for healthy individuals, indicates that due to damages in the vmPFC, an area of the brain responsible for encoding the "affective value of emotional stimuli" (Winecoff et al., 2013, p. 11037) and "processing intentions with high emotional content" (Young et al., 2010, p. 849), prevents these patients from generating the negative emotional content that influences judgements of intent and actual cause of harm. Given the results of their study the authors conclude that vmPFC "participants' abnormal responding to attempted harms may be mediated by a specific deficit in triggering a sufficiently robust emotional response to these representations, in this case, an aversive response to harmful intent" (Young et al., 2010, p. 849).

The study of the behavioural correlates of brain damage in vmPFC patients or, as Greene (2014) puts it, the study of "bad brains" (p. 1014-1015), affords us the opportunity to understands what the differences in terms of the cognitive processes in damaged brains are *vis-à-vis* the cognitive processes in healthy brains and to formulate more accurate theories of decision-making. Multiple studies, such as the ones previously addressed, have established that different areas of the brain are responsible for different decision-making tasks (i.e., the anterior cingulate cortex is associated with emotional appraisals and the dorsolateral prefrontal

cortex is associated with utilitarian judgements while the ventromedial prefrontal cortex integrates both appraisals and generates an "overall value judgement" (Hutcherson et al., 2015, p. 12604). Some studies (Hutcherson et al., 2015; Shenhav & Greene, 2010) highlight that moral decisions and economic decisions share a similar architecture in terms of the regions of the brain involved in the decision-making process. This large overlap between brain structures that produce valuation for economic and moral purposes indicates that findings in studies of moral judgements may have non-negligible implications for the study of economic decisions.

In an attempt to integrate these emotivist insights - that is, findings that have highlighted the role of emotions in the decision-making process – with traditional rationalist models in moral psychology, Greene et al. (2001) analysed differences in responses to sacrificial dilemmas, more specifically differences in responses to the trolley dilemma *vis-à-vis* the footbridge dilemma, to ascertain why individuals feel more conflicted in some scenarios than in others. The trolley dilemma, when individuals are asked whether they would push a lever to divert the course of a trolley thus killing one person instead of five, the answer is, most of the times, affirmative, but in a similar scenario, the footbridge dilemma, when the individuals are asked if they would push a person of the bridge and into the tracks thus causing the trolley to stop and saving five lives, the answer is overwhelmingly negative (Greene et al., 2001, p. 2105). The problem is similar, and the outcome, for all intents and purposes, is the same: one individual dies so that five other survive. What, then, explains this difference in choices in this set of dilemmas? Greene and his colleagues suggest that the difference in responses can be attributed to the salience of agency (i.e.: one's responsibility over given outcome), creating a distinction between "moral personal" dilemmas, where it is expected a stronger emotional activation since the agent has direct involvement in the outcomes (e.g.: pushing a person off the bridge into the tracks), and "moral impersonal" dilemmas where the individual is expected to feel less emotional activation since agency is less salient (i.e.: pushing a lever to divert a trolley). The authors observed that there were clear and consistent differences in the patterns of brain activity between the moral personal condition, where brain areas associated with emotional processing displayed increased activity, and the moral impersonal condition, where the patterns of brain activity resembled those of non-moral decision-making (Greene et al., 2001, p. 2107). In a following study, Greene et al. (2004) found that utilitarian judgments were associated with increased activity in the dorsolateral prefrontal cortex, when compared to deontological judgements. In this study, only high conflict "personal" dilemmas were employed, thus extending the logic of distinct brain processes occurring in different strains of moral judgements beyond a personal/impersonal paradigm.

These finding have led Greene to propose a Dual Process Theory of Moral Judgement (DPT). DPT posits that there are two systems that are involved in moral judgements: an intuitive system that is fast, automatic, and unconscious, making extensive use of (heuristics) and emotions; and a reflective system that works at a slower, more deliberate, and conscious pace relying on logical reasoning and rigorous examination of facts. Greene (2013) argues that the "automatic setting makes […] thinking efficient" while the "manual mode makes our thinking flexible" (p. 171). Furthermore, Greene's DPT makes a clear association between controlled cognition and consequentialist moral judgements while intuitive responses are associated with deontological judgements "that are naturally justified in terms of right or duties" (Greene, 2014, p. 1016).

Building on the DPT framework, Rand et al. (2014) posited the existence of heuristics that are informed by underlying social environment and coined this the Social Heuristics Hypothesis (SHH). SHH conjugates insights from word in the field of heuristics (Gigerenzer, 2004; Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1982), as well as the literature on dual-process theory (Greene et al., 2001; Greene et al., 2004) and proposes, *contra* RCT, that individual's preferences are best understood not as an utility function, but rather as "two competing sets of preferences, one more intuitive and the other more reflective" (Rand et al., 2014, p. 2). This theory accounts for situational (i.e.: context in which the decision is made) and individual factors (i.e.: prior experiences) when making predictions about decisions. SHH proposes that the intuitive process, our "generalized default response" (also the "social heuristic"), is shaped by previous experiences and should generally favour cooperation "given the pervasiveness of mechanisms that make cooperation advantageous in daily life in the long run" (Rand, 2016, p. 2). Alternatively, the reflective process enables us to adapt our behaviour in response to the unique social context we are in at any particular time, and cooperation (or non-cooperation) is contingent on the specifics of the situation. SHH makes two predictions:

> First, experimentally promoting intuition over deliberation should increase pure cooperation, on average, because in social dilemmas, intuition can favor either cooperation or non-cooperation (depending on the individual), whereas deliberation always favors noncooperation. Second, experimentally promoting intuition over deliberation should have no overall effect on strategic cooperation, because both intuition and deliberation may favor either cooperation or noncooperation, depending, respectively, on the individual's experiences and his or her explicit expectations about what strategy will be payoff maximizing in the current context. (Rand, 2016, p. 2)

To test these two predictions the authors ran multiple one-shot economic game experiments (Public Goods Games) that involved the "application of time pressure/delay to cooperation decisions" (Rand et al., 2014, p. 3), and found that when time pressure was

applied – reduced time to make a decision will prompt participants to behave instinctively – participants displayed higher levels of cooperation, rather than defection (Rand et al., 2014, pp. 8–10). The authors also found that this intuitive response was conditioned by the institutional, social, and cultural milieu from which the participant came[6], such that if a given participant had developed its intuitions in an institutional setting that favoured cooperation those intuitions would, most likely, be cooperative rather than non-cooperative (Rand et al., 2014, pp. 1–2). This study also found, in line with the second prediction, that those with experience in economic games, had created new context-specific institutions that favoured non-cooperation in the context of one shot-economic game.

In a 2016 paper, Rand conducted a meta-analysis on experimental studies that looked at the relation between intuition, deliberation, and cooperation, to ascertain whether existing experimental data supported the SHH. In pure cooperation experiments, Rand (2016) found, in line with SHH's predictions, an increased level of cooperation (17.6% more cooperation) in the intuitive condition *vis-à-vis* the deliberative condition, and that this increase was robust to the inclusion of individuals that failed to comply with the experimental manipulations (4.2 % more cooperation). This increase is not explained by competing alternatives such as decision conflict – there is a negative, instead of positive, correlation between the "level of cooperation in the baseline (more deliberative) condition, and the difference in cooperation between manipulation conditions (the intuition effect)" (Rand, 2016, p. 10) – or random play – the levels of cooperation are not close to 50% as randomness in play would entail. In strategic cooperation experiments, there were no statistically significant effects of increased intuitive processing, confirming the predictions advanced by the SHH. The author concludes by stating that:

> These findings highlight the importance of social heuristics for human cooperation. Critically, the key finding is not simply that promoting intuition over deliberation increases cooperation. Rather, the key finding is that this occurs when the SHH predicts it will (pure cooperation) and does not occur when the SHH predicts it will not (strategic cooperation). Not only does the SHH predict these empirical findings, but it also provides an explanation for why intuition and deliberation should have come to function as they do: A simple process of maximizing long-run payoffs (be it via evolution, social learning, or strategic reasoning) can explain the observed pattern of behavior as the result of an on-average optimal set of responses. (Rand, 2016, p. 12)

SSH puts in question the assumption of preference exogeneity, since test subjects based their intuitive decisions in internalized strategies that reflected their institutional background,

[6] The authors note that Indian participants contributed "17% less of the endowment spent on cooperating on average compared with subjects residing in the United States" (Rand et al., 2014, p. 2)

thus making institutions a crucial element to understand decision behaviour. Other assumptions challenged by this study are the assumption of rationality and the assumption of self-interest, since test subjects, who were not familiarized with experimental economic game paradigms, made use of a “social heuristic” rather than effortful reasoning to make decisions, even if adopting other strategies, namely defection, would have increased their expected utility.

Chapter 4

Implications of a new model of human decision-maker

In the previous chapters we have explored critiques of the assumptions that underlie the concept of HE, namely the assumption of preference exogeneity and perfect rationality, and determined that these assumptions stand on dubious epistemological grounds, with empirical evidence pointing to individuals that have more modest and flawed cognitive capacities and rules of thumb - heuristics - to make decisions. We have also seen that individuals' preferences over outcome are, to a great degree, influenced by their previous experiences and reflect the environment in which they were developed (Rand et al., 2014). Moral norms, values and motives constitute a fundamental environmental and motivational factor in the person-situation interaction (Fleeson & Noftle, 2008). They offer behavioural stability, where moral motives are already defined (i.e., ongoing relations) and opportunity to modulate and adapt behaviour where moral motives are yet to be established (i.e., in circumstances where there are no previous relations) (Brodbeck et al., 2013; Kugler et al., 2021).

The implications of these findings are widespread and might offer new solutions to enduring problems. One key problem pertains to the efficacy of incentive-based policy instruments. Some studies point to a more nuanced relation between outcomes and incentives, where the latter might have, in certain situations, a perverse effect on individual's intrinsic motivation to pursue certain goals. Intrinsic motivation can be defined as the willingness to perform and activity when there are no "apparent rewards except the activity itself" (Deci, 1971, p. 105). Regarding the subject of crowding out of intrinsic motivation, Frey & Jegen (2001) explain that "[The] 'crowding-out effect' suggests the opposite of the most fundamental economic 'law', that raising monetary incentives increases supply. If the crowding-out effect holds, raising monetary incentives reduces, rather than increases, supply" (p. 590). The notion that monetary incentives might perversely affect intrinsic motivation to perform an action was first suggested by Titmuss (1971) and Deci (1971). Titmuss (1971) argued that the commoditization of blood donations, undermines the altruistic nature of the act, and depletes the stocks of intrinsic motivation among donors, resulting in a smaller volume of donations once the economic incentives have been put in place. As Frey & Jegen (2001) point out, the author didn't present "serious empirical evidence", nonetheless, his arguments garnered a great deal of attention (p. 589). Deci (1971), on the other hand, conducted a series of laboratory experiments with undergraduate students to test if external rewards had any effect on intrinsic motivation. These experiments consisted of three sessions of puzzle-solving where the

experimental group[7] was exposed to incentive-based rewards for the completion of puzzles during the second session and then, in the last (third) session, the rewards were removed. To measure motivation, the groups of students were left unattended, but monitored for the duration of the experiment, and the measure of motivation would be the time spent working on the puzzles. Deci (1971) found that in the reward condition, when offered money[8] participants would spend more time doing the task. Surprisingly, in the third session, when incentives were removed, the participants in the experimental group spent less time solving puzzle then the control group.

In a telling study, Gneezy & Rustichini (2000) selected a random group of kindergartens in Haifa, Israel, and during a 20-week period, they tracked late arrivals, and implemented a fine - from the fourth to the seventeenth week - that was to be applied to those parents who failed to comply with the schedules to pick up their children. The fine resulted not in a decrease, but in an increase (double the number of late arrivals when compared to the control group and the pre-fine baseline) of late arrivals. Moreover, after the fine was removed, the trend endured. These results indicate that when a tacit understanding, such as the social norms that dictate that one ought to pick-up their children on time from school in order not to inconvenience the professor, gets transformed into a tangible arrangement, a fine for late arrivals, we are putting a price on non-compliance, commoditizing the breaking of social norms and allowing for an easier cost-benefit analysis that is not so straight forward to do with social and moral norms.

Bohnet (2006) conducted a review on the matter of how different institutional arrangements elicit determinate behavioural responses from individuals. In one section, Bohnet (2006) reviews the literature on the topic of how institutions influence the levels of trustworthiness by either rewarding those who can be trusted or by punishing those who cannot. Regarding this matter the author notes that stronger legislation (i.e., more punitive) does not translate into overall greater levels of trust and can have negative downsides such as the loss of intrinsic motivation to be trustworthy. The author says:

> Stronger institutional constraints or more law do not necessarily increase contract performance even if they increase the expected cost of betrayal. If extrinsic incentives are large enough to deter and outweigh any loss in intrinsic motivation, performance increases. However, when incentives

---

[7] The control group was never promised or offered any rewards for the completions of the puzzles and the experimental condition - for the experimental group - was only introduced in the second session, allowing the experimenters to set a baseline of performance.

[8] In the money-as-reward condition, exposure and the removal of rewards lead to a loss of intrinsic motivation, however when other kinds of rewards were used, such as “verbal reinforcement and positive feedback” (Deci, 1971, p. 114), the participants showed improved levels of intrinsic motivation, even during the third session where rewards were removed.

> do not satisfy this condition, they may lead to less performance than without any external intervention because they destroy the intrinsic motivation to be trustworthy. (Bohnet, 2006, pp. 226–227)

In a paradigmatic study, Frey et al. (1996), highlight the interplay between morals and markets, as they investigate the factors that influence support/non-support for the construction of a nuclear waste management facility in Switzerland. The authors conducted interviews with locals to assess the level of support for the construction of the facility and found that support was higher when individuals were acting out of a sense of civic duty (i.e., the non-reward condition) compared to lower levels of support when offered a monetary reward for the construction of the facility. The researchers justify this this crowding-out of civic duty by arguing the "monetary rewards deprive individuals of the possibility of indulging in altruistic feelings" (Frey et al., 1996, p. 1301). The crowding-out effect was being driven by a "bribe effect" which the authors define as "individuals [incurring] moral costs by publicly showing that their approval can be bought … [reducing] the willingness to accept the noxious facility by imposing moral costs" (Frey et al., 1996, p. 1300). The author caution, that decisions, particularly those made in the field of politics, are often times enshrined in moral consideration, and such an understanding might lead us to question the suitability of current models of voting behaviour (See Geys (2006) for a critique of the rational voter theory). One's sense of self-image becomes tied to a moral commitment to a belief that could be better stated by the colloquial expression: morality *isn't up for sale*.

Fiske (1992) expands on RMT (originally presented in Fiske, 1991), and presents a set of predictions regarding various aspects of social life and how different relational models elicit certain types of responses. If we recover the previous example, choosing a location for the nuclear waste management facility, we could argue, using Fiske (1992) framework, that individuals in the no-reward condition - where Frey et al. (1996) offered that they were acting altruistically out of a sense of civic duty - were looking at the problem through a communal sharing relational model frame. This relational model interprets the building of this facility in their town not as a burden for which there should be monetary reparations, but rather as a needed sacrifice for the good of the whole[9]. Fiske (1992), regarding the nature of relationship in a communal sharing relational model, maintains that "in norms or moral ideals, people account for their sense of solidarity and identity with the "we" group, and the special care and altruistic sacrifice they provide for their CS (communal sharing) partners, in terms of immanent, pre-existing, natural facts" (p. 699). The use of monetary rewards, which is associated with

[9] See Fiske (1992), pp. 694 - 696 for the theoretical predictions regarding various social outcomes such as decision-making, social identity and relational self, motivation.

another relational model, namely Market Pricing, creates a conflict between the former relational model, based on a sense of community and kinship, and the latter, which understands value as something that can be subsumed to a monetary sum. This mismatch between the relational model employed by the policymakers and the citizens creates a backlash that results in lower levels of acceptance when monetary incentives are introduced. In a recent paper, Gallus et al. (2022) establish a connection between Fiske's RMT and incentives schemes, identifying the proper incentive scheme associated with each relational model. Here the authors further explore the ramifications of the mismatch between the incentive schemes used and the relational models in place. Gallus et al. (2022) state:

> Certain means may be cues of one relational model or another - for example, money for market pricing […]. This […] suggests that 'economic incentives' (money) may backfire by signaling that the relational coordination is a market pricing transaction and not a "social exchange" and by reducing the giver's ability to signal her knowledge of the recipient's preferences. (p. 591)

Through creating institutions that convey the proper[10] relational model and the underlying moral motive, we can devise incentive schemes that promote compliance and cooperation and avoid the crowding-out of intrinsic motivation and the related negative effects, namely the substitution of intrinsic motivation by extrinsic motivation, which can have enduring motivational effects with the possibility to spill over to otherwise unrelated problems (e.g., pursuing monetary incentives in tightly knit communities might adversely affect the intrinsic motivation borne out of a spirit of community).

There are also important implications for economic theory. Theories on household saving behaviour, for instance, stand to gain from the developments in the fields of psychology and cognitive sciences mentioned until now. In understanding household saving behaviour, we must first expand on the concepts of delayed gratification and social trust, both of which are fundamental to understand this phenomenon. The psychological process of delay of gratification, was first explained by Walter Mischel during the 60's and 70's (Mischel, 1974; Mischel et al., 1989; see Mischel, 1961; Mischel & Ebbesen, 1970) and it is a process wherein an individual chooses to delay a reward in the present for the possibility of a greater reward in the future. Veldhoven & Groenland (1993) relate the concept of delayed gratification to saving behaviour and argue that "saving requires cognitive planning and structuring of activities within time, and toward goals. At the same time saving requires one to refrain from direct gratification, and to postpone benefits and outcomes to the (near) future" (p. 510). Social trust, on the other and refers to a generalized feeling of trust towards others and institutions. Welch et al. (2005)

[10] The term "proper" refers to matching the incentive scheme to the relational model frame of the intervening parts.

define social trust as "[…] the mutually shared expectation, often expressed as confidence, that people will manifest sensible and, when needed, reciprocally beneficial behaviour in their interactions with others" (p. 457). Taken together, we can conceivably propose that social trust might influence our capacity to delay gratification, in such a manner that low levels of social trust might implicate weak capacity to delay gratification and have an impact in the saving behaviour of households. A recent study by Michaelson et al. (2013), established precisely that relation, that is, perceived trustworthiness predicts social trust, in one-on-one interaction but also in interactions with multiple individuals. The authors suggest that "emphasizing social trustworthiness might be important in interventions for delaying gratification" (Michaelson et al., 2013, p. 6). The problem of household savings becomes a wicked problem when we consider the predicament of individuals with low socio-economic status (SES). On one hand, lack of trust can be explained by a low SES. Stamos et al. (2019) have found that children from low SES households are less trusting than those from high SES households. This lack of trust is interpreted by the authors as a "life-history strategy" - a strategy that is adopted when growing up in a given environment to increase fitness - arises from lack of resource availability and low sense of control (Stamos et al., 2019, p. 24) and that it influences decisions throughout life. On the other hand, trust in not only an outcome of low SES, but also a precedent. Nooteboom (2007) argues that trust is "both an outcome and an antecedent of relationships. It forms a basis for relationships, and thus generates social capital. It may be based on institutions, and it may be built from relationships, and then it arises from social capital" (p. 29). In this manner, people from low SES, become trapped in poverty. The lack of resources and perceived lack of control leads individuals to distrust others, which in turn leads them to use whatever resources they have when they have them, which precludes them from, in the long-turn, create savings and acquire assets that would allow them to climb the socio-economic ladder, thus maintaining these individuals in a perpetual state of lack of trust.

Rand et al. (2014) SHH provide an important insight into just how we can go about increasing social trust and, consequently, cooperation and possibly change household saving behaviours. The authors found that individuals from different institutional backgrounds display different levels of cooperation throughout multiple experiments. The authors suggested (as Stamos et al. (2019) did), that institutions have a relevant and lasting impact in the way we intuitively think about cooperation and the degree to which we trust social others (i.e.: institutions, individuals). The group argued that:

> Subjects whose intuitions were shaped in contexts where cooperation is not supported will internalize defection as their default response. Thus for these subjects, no cognitive conflict exists: both intuition and reflection favour selfishness, and cooperation should be relatively low regardless of whether subjects decide intuitively or reflectively. These could be people who, for example,

> developed under social norms that allow or promote selfishness, or corrupt institutions where free-riding is incentivized. (Stamos et al., 2019, p. 2)

These insights must inform the way in which we develop public policy and design public institutions. Where the policy goals might have once focused on providing individual-level responses, these contributions force us to acknowledge the role of wide ranging, population-level interventions that go beyond monetary and financial incentives and subsidies, to establishing, for instance, “shared ethics of conduct” (Nooteboom, 2007, p. 48), to create comprehensive understandings between those involved by framing and moderating expectations of appropriate behaviours and outcomes.

Chapter 5

Conclusion

HE has taken a central place, first in economics and then in all of the Social Sciences. Since the work of Herbert Simon, and later with the developments that culminated in Behavioural economics, there have been a greater number of contributions that recognize the central role of psychological processes that intervene in the decision-making process. Despite the progress in this direction and the theoretical diversity that exists in the new approaches associated with behavioural economics and, by extension, behavioural sciences, the approach that has become prevalent has been that proposed by Kahneman and Tversky. This approach, in turn, fails to present a new normative theory, continuing to use the same normative standard proposed in neoclassical economics, that is, rationality as utility maximization. Here I argue, in line with Gigerenzer & Selten (2001), that a proper normative theory is premised on the notion of bounded rationality, a notion that recognizes interplay between the mind and the environment and which assesses the rationality of a decision depending on the context in which it is taken.

The critique that has been developed in this dissertation, that morality is a central and defining factor in the decision-making process, has taken into account developments in the behavioural sciences, mainly moral psychology and neuroscience, to argue that the current models of decision maker and decision-making need to incorporate not only behavioural but also situational and motivational elements. The theories discussed here, namely RRT/RMT as well as dual-process theory of moral judgement, consider the relevance of morality as an inter-personal and intra-personal phenomenon, and establish mechanisms by which different moral motives and or psychological processes intervene in the decision-making process. RRT and RMT treat morality primarily as an interpersonal phenomenon. These theories suggest that different relationships have different underlying relational models, and these are associated with different moral motives. Individuals intuitively recognize the different relational models, acting in accordance with the associated moral expectations and commitments, taking into account the trade-offs associated with the relational model in which they find themselves to make decisions. Dual-process theory of moral judgment, on the other hand, addresses the intra-personal component of morality by looking at the psychological and cognitive processes that occur during the moment of decision. There are two concurrent psychological processes: one process is intuitive and automatic and the other is reflective and deliberative. These concurrent and competing processes favour different strategies, the former produces, mainly, deontological judgments, while the latter produces, mainly, utilitarian judgments.

From these two distinct theoretical contributions, we were able to stipulate the implications of incorporating the moral component in the decision-making process for the design and implementation of public policies. On the one hand, RMT/RRT define the type of strategy suitable for the pursuit of a given objective, taking into account the relational model and the associated moral motive in such a way that policymakers can fulfil short- and medium-term objectives without compromising the intrinsic motivation that individuals have to carry out vital social commitments. On the other hand, the dual-process theory of moral judgment offers us important clues on how to mitigate problems associated with household savings behaviours, particularly through the creation and implementation of institutions, formal and informal, that foster social capital and social trust, since these seem to favourably influence the increase and maintenance of higher savings rates.

The present dissertation does not constitute a detailed critique of the concept of Homo Economicus, focusing exclusively on the relationship between morality and the decision-making process and the implications that such a conception has for this concept. As such, the relevance of this criticism is pertinent only to this point, not having the scope that a more generalized criticism would have. Nevertheless, this critique constitutes an exploratory exercise, an effort to summarize and synthesize the literature on the subject of morality and its relationship with the decision-making process and, more broadly, as an alternative to the current model of economic actor. Because it is an exploratory exercise, and given the complexity and vast number of theoretical contributions on this topic, this dissertation sometimes lacks conciseness, resulting from the attempt to condense and synthesize the various theoretical contributions into a coherent and representative narrative of the state of art. This is undoubtedly one of the weakest points of this work that will require additional efforts to be developed in subsequent works. In this sense, this critique configures an alternative approach to that of orthodox economics and even behavioural economics, avoiding problems related to normative theories of action by presenting itself as a descriptive exercise of the decision-making process as a cognitive process influenced by situational and environmental factors. I was unable to properly develop here the implications that adopting a different normative theory of human behaviour, in this case, bounded rationality, would have in our understanding of human action. Such an exercise would require a different approach to this topic, one more centred in contributions from philosophy, and as such would constitute an entirely different effort, but a necessary and important one to compliment and advance the ideas discussed here.

In the future, the contributions discussed here may generate related studies. The framework proposed by RMT/RRT, based on relational models and moral motives, can inform public policy experiments in problems of collective choice and collective action. Problems such

as the construction of an airport, or other facilities of public interest, are often contentious, and these theories offer answers to these problems by suggesting different strategies based on the prevailing relational model. If we can understand the frames and relational models/moral motives, we can improve the efficacy of public policy interventions. As far as the dual-process theory of moral judgment is concerned, it may help to solve problems associated with social cooperation. Studies on this topic could focus on how time pressure, or framing effects, could lead individuals to demonstrate higher levels of cooperation. More interestingly, it will also be possible to assess which institutional configurations most favour cooperation to enhance cooperative intuitions *vis-à-vis* non-cooperative intuitions. Based on these insights, some studies could be developed in comparative politics, to understand which kinds of political and social institutions favour intuitive cooperation and which favour free-riding and non-cooperation. Alternatively, longitudinal case studies could be conducted to understand how changes in formal and informal institutions throughout time impacted individuals' cooperation in a given society.

References

Altman, M., 2004. The Nobel Prize in behavioral and experimental economics: a contextual and critical appraisal of the contributions of Daniel Kahneman and Cernon Smith. Review of Political Economy 16, 3–41. doi:10.1080/0953825032000145445

Bazerman, M.H., Malhotra, D., 2006. Economics Wins, Psychology Loses, and Society Pays, in: De Cremer, D., Zeelenberg, M., Murnighan, J.K. (Eds.),. Lawrence Erlbaum, Mahwah, N.J, pp. 263–280.

Bechara, A., Damasio, A.R., 2005. The somatic marker hypothesis: A neural theory of economic decision. Games and Economic Behavior, Special Issue on Neuroeconomics 52, 336–372. doi:10.1016/j.geb.2004.06.010

Bechara, A., Tranel, D., Damasio, H., 2000. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. Brain: A Journal of Neurology 123 (Pt 11), 2189–2202. doi:10.1093/brain/123.11.2189

Bohnet, I., 2006. How institutions affect behavior: Insights from economics and psychology. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 213–237.

Bowles, S., 2016. The moral economy: why good incentives are no substitute for good citizens, The castle lectures in ethics, politics, and economics. Yale University Press, New Haven; London.

Brodbeck, F.C., Kugler, K.G., Reif, J.A.M., Maier, M.A., 2013. Morals Matter in Economic Games. PLOS ONE 8, e81558. doi:10.1371/journal.pone.0081558

Caldwell, B.J., 1980. A critique of friedman's methodological instrumentalism. Southern Economic Journal 47, 366374.

Damásio, A., 2018. A estranha ordem das coisas: A vida, os sentimentos e as culturas humanas. Círculo de Leitores, Lisboa.

Damásio, H., Grabowski, T., Frank, R., Galaburda, A.M., Damásio, A.R., 1994. The return of phineas gage: Clues about the brain from the skull of a famous patient. Science 264, 1102–1105. doi:10.1126/science.8178168

Deci, E.L., 1971. Effects of externally mediated rewards on intrinsic motivation. Journal of Personality and Social Psychology 18, 105–115. doi:10.1037/h0030644

Faundez, J., 2016. Douglass North's Theory of Institutions: Lessons for Law and Development. Hague Journal on the Rule of Law 8, 373–419. doi:10.1007/s40803-016-0028-8

Fehr, E., Rangel, A., 2011. Neuroeconomic Foundations of Economic Choice–Recent Advances. Journal of Economic Perspectives 25, 3–30. doi:10.1257/jep.25.4.3

Fiske, A.P., 1991. Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing. Free Press, New York, NY, US.

Fiske, A.P., 1992. The four elementary forms of sociality: Framework for a unified theory of social relations. Psychological Review 99, 689–723. doi:10.1037/0033-295X.99.4.689

Fiske, A.P., 2002. Socio-moral emotions motivate action to sustain relationships. Self and Identity 1, 169–175. doi:10.1080/152988602317319357

Fiske, A.P., 2004a. Four Modes of Constituting Relationships: Consubstantial Assimilation; Space, Magnitude, Time, and Force; Concrete Procedures; Abstract Symbolism, in: Haslam, N. (Ed.),. Lawrence Erlbaum, Mahwah, NJ, pp. 61–146.

Fiske, A.P., 2004b. Relational Models Theory 2.0, in: Haslam, N. (Ed.),. Lawrence Erlbaum, Mahwah, NJ, pp. 3–26.

Fiske, A.P., Haslam, N., 2005. The four basic social bonds: Structures for coordinating interaction. Guilford Press, New York, NY, US, pp. 267–298.

Fleeson, W., Noftle, E., 2008. The End of the Person-Situation Debate: An Emerging Synthesis in the Answer to the Consistency Question. Social and Personality Psychology Compass 2, 1667–1684. doi:10.1111/j.1751-9004.2008.00122.x

Frey, B.S., Jegen, R., 2001. Motivation Crowding Theory. Journal of Economic Surveys 15, 589–611. doi:10.1111/1467-6419.00150

Frey, B.S., Oberholzer-Gee, F., Eichenberger, R., 1996. The old lady visits your backyard: A tale of morals and markets. Journal of Political Economy 104, 1297–1313. doi:10.1086/262060

Friedman, M., 1953. The Methodology of Positive Economics. University of Chicago Press, Chicago, pp. 3–46.

Gallus, J., Reiff, J., Kamenica, E., Fiske, A.P., 2022. Relational incentives theory. Psychological Review 129, 586–602. doi:10.1037/rev0000336

Geys, B., 2006. 'Rational' Theories of Voter Turnout: A Review. Political Studies Review 4, 16–35. doi:10.1111/j.1478-9299.2006.00034.x

Gigerenzer, G., 2004. Fast and frugal heuristics: The tools of bounded rationality. Blackwell Publishing, Malden, pp. 62–88. doi:10.1002/9780470752937.ch4

Gigerenzer, G., 2018. The Bias Bias in Behavioral Economics. Review of Behavioral Economics 5, 303–336. doi:10.1561/105.00000092

Gigerenzer, G., Selten, R., 2001. Rethinking rationality, in: Bounded Rationality: The Adaptive Toolbox. The MIT Press, Cambridge, MA, US, pp. 1–12.

Gneezy, U., Rustichini, A., 2000. A fine is a price. The Journal of Legal Studies 29, 1–17. doi:10.1086/468061

Greene, J.D., 2013. Moral tribes: Emotion, reason, and the gap between us and them. The Penguin Press, New York.

Greene, J.D., 2014. The Cognitive Neuroscience of Moral Judgment and Decision Making, in: Gazzaniga, M.S., Mangun, G.R. (Eds.),. MIT Press, pp. 1013–1019.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. The Neural Bases of Cognitive Conflict and Control in Moral Judgment. Neuron 44, 389–400. doi:10.1016/j.neuron.2004.09.027

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science (New York, N.Y.) 293, 2105–2108. doi:10.1126/science.1062872

Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychological Review 108, 814–834. doi:10.1037/0033-295x.108.4.814

Haidt, J., 2007. The new synthesis in moral psychology. Science 316, 998–1002.

Hausman, D.M., 2007. The philosophy of economics: An anthology, 3rd ed. Cambridge University Press, Cambridge. doi:10.1017/CBO9780511819025

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In search of homo economicus: Behavioral experiments in 15 small-scale societies. The American Economic Review 91, 73–78.

Herfeld, C., 2018. From theories of human behavior to rules of rational choice: Tracing a normative turn at the cowles commission, 194354. History of Political Economy 50, 1–48. doi:10.1215/00182702-4334997

Hutcherson, C.A., Montaser-Kouhsari, L., Woodward, J., Rangel, A., 2015. Emotional and Utilitarian Appraisals of Moral Dilemmas Are Encoded in Separate Areas and Integrated in Ventromedial Prefrontal Cortex. Journal of Neuroscience 35, 12593–12605. doi:10.1523/JNEUROSCI.3402-14.2015

Kahneman, D., Tversky, A., 1972. Subjective probability: A judgment of representativeness. Cognitive Psychology 3, 430–454. doi:10.1016/0010-0285(72)90016-3

Kahneman, D., Tversky, A., 1973. On the psychology of prediction. Psychological Review 80, 237–251. doi:10.1037/h0034747

Kleinjans, K.J., Gill, A., 2018. Institutions, parental selection, and locus of control. Applied Economics Letters 25, 1041–1044. doi:10.1080/13504851.2017.1391998

Krackhardt, D., 1998. Endogenous preferences: A structural approach, in: Halpern, J., Stern, R.N. (Eds.),. Cornell University Press, pp. 239–247.

Kugler, K.G., Reif, J., Petersen, G.K., Brodbeck, F.C., 2021. The impact of moral motives on economic decision-making. Journal of Dynamic Decision Making 7. doi:10.11588/jddm.2021.1.77559

Michaelson, L., Vega, A. de la, Chatham, C., Munakata, Y., 2013. Delaying gratification depends on social trust. Frontiers in Psychology 4.

Mill, J.S., 1836. On the definition of political economy; and on the method of philosophical investigation in that science. The London and Westminster Review IV, 1–29.

Mill, J.S., 1882. A system of logic ratiocinative and inductive, being a connected view of the principles of evidence, and the methods of scientific investigation. Harper & Brothers.

Mischel, W., 1961. Delay of gratification, need for achievement, and acquiescence in another culture. The Journal of Abnormal and Social Psychology 62, 543–552. doi:10.1037/h0039842

Mischel, W., 1974. Processes in Delay of Gratification, in: Berkowitz, L. (Ed.),. Academic Press, pp. 249–292. doi:10.1016/S0065-2601(08)60039-8

Mischel, W., Ebbesen, E.B., 1970. Attention in delay of gratification. Journal of Personality and Social Psychology 16, 329–337. doi:10.1037/h0029815

Mischel, W., Shoda, Y., Rodriguez, M.L., 1989. Delay of gratification in children. Science 244, 933–938. doi:10.1126/science.2658056

Moscati, I., 2018. Measuring utility: From the marginal revolution to behavioral economics, in: Moscati, I. (Ed.),. Oxford University Press, pp. 25–47. doi:10.1093/oso/9780199372768.003.0003

Neumann, J.V., Morgenstern, O., 1944. Theory of games and economic behavior. Princeton, NJ, USA: Princeton University Press.

Neumann, J.V., Morgenstern, O., 1947. Theory of games and economic behavior, 2nd Revised Edition. ed. Princeton University Press, Princeton, NJ, US.

Nooteboom, B., 2007. Social capital, institutions and trust. Review of Social Economy 65, 29–53.

North, D.C., 2005. Understanding the Process of Economic Change. Princeton University Press.

North, D.C., 1991. Institutions. Journal of Economic Perspectives 5, 97–112. doi:10.1257/jep.5.1.97

Rai, T.S., Fiske, A.P., 2011. Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. Psychological Review 118, 57–75. doi:10.1037/a0021867

Rand, D.G., 2016. Cooperation, Fast and Slow: Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation. Psychological Science 27, 1192–1206. doi:10.1177/0956797616654455

Rand, D.G., Peysakhovich, A., Kraft-Todd, G.T., Newman, G.E., Wurzbacher, O., Nowak, M.A., Greene, J.D., 2014. Social heuristics shape intuitive cooperation. Nature Communications 5, 1–12. doi:10.1038/ncomms4677

Sandel, M.J., 2013. Market Reasoning as Moral Reasoning: Why Economists Should Re-engage with Political Philosophy. Journal of Economic Perspectives 27, 121–140. doi:10.1257/jep.27.4.121

Schmidtz, D., 1994. The institutions of morality. Nomos 36, 228–243.

Selten, R., Ockenfels, A., 1998. An experimental solidarity game. Journal of Economic Behavior & Organization 34, 517–539. doi:10.1016/S0167-2681(97)00107-8

Sent, E.M., 2004. Behavioral economics: How psychology made its (limited) way back into economics. History of Political Economy 36, 735–760. doi:10.1215/00182702-36-4-735

Shenhav, A., Greene, J.D., 2010. Moral Judgments Recruit Domain-General Valuation Mechanisms to Integrate Representations of Probability and Magnitude. Neuron 67, 667–677. doi:10.1016/j.neuron.2010.07.020

Simon, H.A., 1957. Models of man - social and rational. John Wiley & Sons, Inc., New York.

Simon, H.A., 1963. Discussion: Problems of methodology. The American Economic Review 53, 229–231.

Stamos, A., Altsitsiadis, E., Dewitte, S., 2019. Investigating the effect of childhood socioeconomic background on interpersonal trust: Lower childhood socioeconomic status predicts lower levels of trust. Personality and Individual Differences 145, 19–25. doi:10.1016/j.paid.2019.03.011

Stigler, G., Becker, G., 1977. De gustibus non est disputandum. American Economic Review 67, 76–90.

Stofberg, N., Bridoux, F., Ciulli, F., Pisani, N., Kolk, A., Vock, M., 2021. A Relational-Models View to Explain Peer-to-Peer Sharing. Journal of Management Studies 58, 1033–1069. doi:10.1111/joms.12523

Titmuss, R.M., 1971. The gift relationship: From human blood to social policy, 1st American ed. ed. Pantheon Books, New York.

Tversky, A., Kahneman, D., 1982. Judgment under uncertainty: Heuristics and biases. Cambridge University Press, Cambridge ; New York.

Urbina, D.A., Ruiz-Villaverde, A., 2019. A Critical Review of Homo Economicus from Five Approaches. American Journal of Economics and Sociology 78, 63–93. doi:10.1111/ajes.12258

Veldhoven, G.M. van, Groenland, E.A.G., 1993. Exploring saving behaviour: A framework and a research agenda. Journal of Economic Psychology 14, 507–522. doi:10.1016/0167-4870(93)90030-O

Weber, M., 1921. Economy and society: An outline of interpretive sociology. University of California Press, Berkeley.

Welch, M.R., Rivera, R.E.N., Conway, B.P., Yonkoski, J., Lupton, P.M., Giancola, R., 2005. Determinants and Consequences of Social Trust*. Sociological Inquiry 75, 453–473. doi:10.1111/j.1475-682X.2005.00132.x

Winecoff, A., Clithero, J.A., Carter, R.M., Bergman, S.R., Wang, L., Huettel, S.A., 2013. Ventromedial Prefrontal Cortex Encodes Emotional Value. Journal of Neuroscience 33, 11032–11039. doi:10.1523/JNEUROSCI.4317-12.2013

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., Damasio, A., 2010. Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. Neuron 65, 845–851. doi:10.1016/j.neuron.2010.03.003