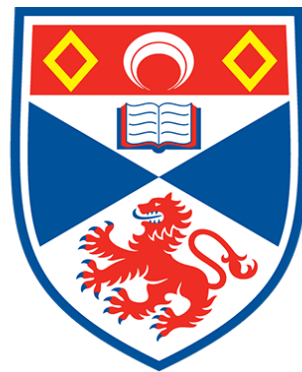# Moral Decision-Making: Essays from Philosophy and Economics

Philipp Schönegger

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

September 2022

Abstract

This thesis investigates moral decision-making from the two disciplinary angles of philosophy and economics. Moral decision-making includes moral judgements (e.g., judgements as to the moral (im-)permissibility of actions) and moral behaviour (e.g., charitable giving). The topic choice throughout this thesis was primarily motivated by global priorities research, ranging from population ethics to effective charitable giving. The first three chapter primarily focus on experimental philosophy. In them, I (a) investigate the relationship between the dark triad personality traits, psychopathy, Machiavellianism, and narcissism and anti-natalist views, finding that those high on dark triad traits are significantly more likely to endorse anti-natalist views. I also find that this relationship is mediated by depression. Then, I (b) study reflective equilibrium behaviour in the context of population ethics, finding that in accordance with theory, concrete case judgements play a revisionary role with respect to endorsements of general moral principles. Further, I (c) argue that research in psychology and experimental philosophy has not adequately dealt with the issue of incentivisation. I then go on to conduct an empirical showcase of the Bayesian Truth Serum in this context, demonstrating impacts on response behaviour. The last three chapters focus on experimental economics. I (d) analyse charitable giving behaviour under normative uncertainty and show that randomly allocated expert advice is undervalued by donors, though can impact donation behaviour and reduce uncertainty. Then, I (e) investigate the effect of morally demanding charitable solicitations on donor behaviour, finding that while moral arguments raise donations, increases of moral demandingness do not. Lastly, I (f) analyse the predictors of donating to probabilistic and ambiguous charities as opposed to more reliable ones, failing to find an impact of risk and ambiguity uncertainty. Overall, I hope that the work presented in this thesis is able to advance global priorities research into these topics.

Candidate's declaration

I, Philipp Schönegger, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 74,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2019.
I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

13 September 2022

Date   Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

13 September 2022

Date   Signature of supervisor   Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not

being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Philipp Schönegger, have obtained, or am in the process of obtaining, third-party copyright permissions that are required or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of 5 years on the following ground(s):

• Publication would preclude future publication

Supporting statement for printed embargo request

Uncertainty over whether publishing the thesis might preclude further publication.

Electronic copy

Embargo on all of electronic copy for a period of 5 years on the following ground(s):

• Publication would preclude future publication

Supporting statement for electronic embargo request

Uncertainty over whether publishing the thesis might preclude further publication.

Title and Abstract

• I agree to the title and abstract being published.

13 September 2022

Date   Signature of candidate

13 September 2022   *Miguel Costa Gomes*   _____

Date   Signature of supervisor   Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Philipp Schönegger, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

13 September 2022

Date   Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

No embargo on underpinning research data or digital outputs.

13 September 2022

Date   Signature of candidate

13 September 2022

Date   Signature of supervisor   Signature of supervisor

General acknowledgements

This work would not have been possible without the support of my parents Sigi and Franz, my brother Jakob, and my girlfriend Indre.

This thesis and I also would not be what they are today without my extraordinary advisors Theron Pummer and Miguel Costa-Gomes.

Research Data/Digital Outputs access statement

Research data underpinning this thesis are available at https://osf.io/pj5wt/ and https://osf.io/w9gfu/.

Introduction

This doctoral thesis consists of a collection of six individual papers all centred around the theme of moral decision-making. Each paper is roughly aligned with the growing field of global priorities research. Global priorities research, broadly, aims to provide answers to the question what one should do with a limited set of resources if the aim is to do the most good possible. In this introduction, I outline the six papers of my doctoral thesis and give a quick overview over the main topics covered in each chapter, their motivations, methods, findings, and relevance. All individual papers are, at least in part, of inter- and multidisciplinary nature, covering philosophy and economics, and also touching on psychology. However, the first three papers are best characterised as (experimental) philosophy and the latter three as (experimental) economics, though there are influences from the respective others in some of them. Two of the six chapters were co-authored with Ben Grodeck, both of which were co-authored jointly in all aspects (chapters 2 and 5). At the moment of submission, four of the six chapters are published or accepted for publication (chapters 1-3 as well as chapter 6) and two are at the Revise & Resubmit stage (chapters 4-5).

The overarching theme of this thesis is moral decision-making. Broadly understood, this encompasses moral judgements (e.g., judgements regarding the moral (im-)permissibility of an action) and moral behaviour (e.g., behaviours relating to moral areas such as charitable giving). This thesis investigates moral decision-making in a variety of contexts, ranging from evaluating anti-natalist principles to reflective equilibrium reasoning in revising one's views on population ethics to charitable decision-making in a variety of different environments. The choice of topics was in part motivated by global priorities research, which describes research on how to do as much good as possible. The work on moral judgements, for example regarding population ethics, aims to contribute to this research programme by providing a better understanding of intuitions about the value of life and the way that we reason about it that may then inform further theorising regarding global priorities. The work on moral behaviour that almost exclusively focuses on charitable giving more directly aims to inform (meta-)charity's operations such that they may be able to increase donations to effective charities that strive to do as much good as possible with their interventions and donations. This work aims to inform such (meta-)charitable organisations by investigating what best predicts risky charitable giving or by attempting to determine the impact of expert advice on donor choices. Below, I outline

all six chapters in more detail. The remainder of this thesis consists of these six stand-alone chapters.

The first chapter is titled 'What's Up with Anti-Natalists? An Observational Study on the Relationship between Dark Triad Personality Traits and Anti-Natalist Views'. It has been published in *Philosophical Psychology* in 2021. The central motivating question of this chapter is what personality features may best explain anti-natalist views, i.e., views that procreation is morally wrong. This is centrally important as much of public policy would look quite differently if anti-natalist inclined individuals were guiding it or were substantially influential in steering public policy. Such anti-natalist views of a lay population had not been formally studied prior to the publication of this article, though there had been an emerging literature showing the relationship between several socially aversive moral judgements (such as sacrificial utilitarian decisions in moral dilemmas) and the dark triad of personality – the topic of this chapter. The dark triad is the cluster of personality traits of psychopathy, Machiavellianism, and narcissism. This chapter introduces a measure of lay anti-natalist beliefs and shows that (sub-clinical) psychopathy and Machiavellianism (but not narcissism) predict agreement with anti-natalism in two studies that also indicate strong test-retest validity. Further, depression is found to be mediating this relationship between dark triad personality traits and anti-natalist views. Importantly, this paper does not by itself advance an argument against or in favour of the claim that because those higher in agreement with anti-natalism are also more likely to feature these personality traits or show symptoms of depressive mood, we should discount anti-natalism. Rather, this paper provides data and argumentation that can be a first step in developing such an argument and that may advance understanding into anti-natalist views in the broader public, with a focus on how this might then be applied to public policy.

The second chapter is titled 'Concrete Over Abstract: Experimental Evidence of Reflective Equilibrium in Population Ethics'. It has been accepted for publication in the Routledge Series 'Issues in Experimental Moral Philosophy' and has been co-authored with Ben Grodeck. This chapter investigates narrow reflective equilibrium reasoning of a lay population in the context of population ethics. The underlying question that is investigated in this chapter is how individuals react when abstract moral principles that they themselves endorse conflict with intuitions and judgements about concrete cases that they make. The context that this question is studied in is population ethics, with a focus on totalism and

averagism. As such, this research most directly contributes to our understanding of how reflective equilibrium functions in a lay population in the context of population ethics, which may be crucial for political decision-making about future spending (that may increase or decrease population size or welfare). This study allows us to investigate exactly this dynamic as participants are asked to endorse several population ethical views (if they wish) while also making judgements on a number of population ethical concrete cases, some of which directly violate these views. Once their choices regarding the abstract principles and concrete cases are made, participants are asked to resolve any potential conflicts by either rejecting the previously endorsed principle, revising their judgement of a case, or continuing with the conflict. The results indicate that participants are significantly more likely to revoke their endorsement of the general moral principles rather than their judgements about concrete cases. This suggests that, as has been hypothesises by some, that case judgements play a central revisionary role in reflective equilibrium reasoning, specifically in the context of population ethics. This may then go on to build the basis for a better understanding of how public reasoning about moral issues evolves and how it may be impacted.

The third chapter is titled 'Experimental Philosophy and the Incentivisation Challenge: A Proposed Application of the Bayesian Truth Serum'. It has been published in *Review of Philosophy and Psychology* in 2021. This chapter is concerned with the methodological point that much of previous research in experimental philosophy, social psychology, and other cognate disciplines has failed to seriously consider how to properly incentivise participants, especially in the contexts of subjective research matters like moral judgements, compared to economics, where considerations of incentivisation are paramount. In contexts like experimental philosophy and social psychology, however, other standard incentivisation mechanisms like paying participants for how close they are to the correct response are not applicable as the researchers have no access to the truth due to the nature of data collected. After all, it is difficult to conceive how experimenters could grade agreements with philosophical statements as true or false. The main claim advanced in this paper is that simply paying participants who partake on platforms such as MTurk or Prolific a fee to complete the study risks incentive-incompatibility regarding them expressing their true views and preferences and answering honestly. This might lead to participants simply maximising their payoffs by reducing time spent per question and increasing the numbers of surveys they complete, thereby failing to answer honestly. To address this, this chapter proposes the

application of the Bayesian Truth Serum, an incentive compatible mechanism introduced by Prelec that has already been used in economics and marketing. The mechanism rewards participants based on how surprisingly common their answer are. This criterion builds on the Bayesian claim that because ones' own view is treated as an 'informative sample of one', one would expect individuals to overestimate the population frequency of one's own view. In order to show that this approach would be tractable in the context of experimental philosophy and social psychology, this chapter also presents an empirical study, showcasing that the Bayesian Truth Serum meaningfully shifts response patterns in a variety of contexts commonly used in the fields that standardly employ Likert-scales. The main conclusion of this chapter is that practitioners of experimental social sciences ought to seriously consider adopting this mechanism to incentivise their research participant's honest expressions of their views, especially if they want their research to inform important policy making outwith the context where failing to think about incentivisation is deemed acceptable.

The fourth chapter is titled 'Charitable Giving under Normative Uncertainty: Experimental Evidence on The Behavioural Impact of Normative Expert Advice' It has received a *Revise & Resubmit* from *Oxford Economic Papers*. It presents one of the first empirical investigations of individual donor behaviour plays out in conditions of normative uncertainty (i.e., under uncertainty about the correct standard of evaluation). Plausibly, almost all charitable decisions are made under this type of uncertainty. After all, we can never be wholly certain about the correct standards of evaluation of any given action; there always remains some level of doubt and uncertainty. Additionally, these choices are also often made under more mundane (descriptive) uncertainty, i.e., uncertainty over the empirical world. Both types of uncertainty, their interplay, and strategies to guide behaviour under them are crucially important when aiming to guide (meta)-charitable decisions and strategy. This chapter presents an experiment in which participants are faced with a choice between five charities that all draw on fundamentally distinct normative claims. They are then able to bid for descriptive information, normative expert advice, or both via a Becker-DeGroot-Marschak mechanism as a means to (partially) resolving that uncertainty. The results indicate that descriptive information (e.g., about a charity's administrative costs) is preferred to normative expert advice. However, when the normative expert advice (by ethicists and economists) is received and has been paid for, this impacts charity choice to a larger extent than if it is provided for free (and at comparable levels to descriptive information). Randomly provided information and

advice also impacted donation choice but failed to increase (or decrease) donation amounts. These results suggest that in situations of normative uncertainty, expert advice can have a significant impact in guiding individual donors, though that this effect is not present if the advice is provided for free, suggesting that a sunk cost effect may be at play here. This provides important evidence for charities aiming to do the most good regarding their provision of expert testimony in their public facing materials.

The fifth chapter is titled 'Demanding the Morally Demanding: Experimental Evidence on the Effects of Moral Arguments and Moral Demandingness on Charitable Giving'. This chapter has been co-authored with Ben Grodeck and has received a *Revise & Resubmit* at *Journal of Behavioral and Experimental Economics*. It asks the general question whether one should confront people with morally stringent demands to perform certain actions if one wants the action to be performed. More specifically, this chapter investigates this question in the context of charitable giving, where this is a question that all charities inevitably face in constructing their public facing materials. Many have worried before that confronting potential donors with highly morally demanding statements regarding to their obligation to give to charity may reduce their frequency of donation and/or the size of their donations. We conduct two experiments in which we vary the level of moral demandingness, ranging from a mere inspirational message to a heavily morally demanding claim. We find that compared to no moral demand at all, moral arguments (about the state of global poverty and one's ability to help) increase the amount donated by 51.7%. However, increasing levels of moral demandingness above and beyond the baseline argument did not lead to an increase in giving or in the frequency in giving. Yet, we find a gender effect in that in the condition of highest moral demandingness, women are more likely to donate while men donate less. This suggests that demanding the morally demanding may have disparate effects on different populations and may as such not be straightforwardly applicable to policy makers or (meta-)charitable organizations. However, in a follow-up study we fail to replicate this gender effect. As a result, the data are unclear as to what the optimal communication strategy for (meta)-charitable organisations is. However, our data allow for the claim that at least moderately demanding communication strategies are very unlikely to lead to backfire effects and are thus, at least, not a mistake to engage in. This presents some actionable take-aways from this work.

The sixth and last chapter of this thesis is titled 'Sure-Thing vs. Probabilistic Charitable Giving: On the Role of Risk and Ambiguity Attitudes'. It has been accepted for publication at

PLOS ONE. It studies the charitable giving context in which charities differ primarily along the axis of probability. The central question is how behaviour in these situations can be predicted from a number of individual difference measures. In other words, the main question investigated here is what best predicts individual donor choices in situations where they are presented with a charity that has a highly reliable intervention that is mostly unambiguous and one that has an intervention that is both highly risky and ambiguous. This question in itself is highly important for global priorities research as many charities and interventions that are often thought of as crucially important for reducing existential risks and safeguarding the long-term future are such that they are, almost by nature, highly risky and extremely ambiguous; think of a charity working to reduce nuclear war risk. This chapter provides first data on this decision and investigates whether risk and ambiguity attitudes play into this behaviour, as well as whether other individual differences (relating to donor type, optimism, empathy, etc) impact behaviour to understand how these individual differences may impact choices. The data provide strong evidence in favour of a null effect, both from equivalence tests and from Bayesian analyses, in that the none of the individual difference measures studied did provided meaningful predictive power as to the choices between sure-thing and probabilistic charities. This suggests that further research into this area is needed to properly disentangle and understand donor choices in this context, and that given the results present now, individual differences may not play a large role in this choice.

Because of the multitude of questions asked and data analysed in this portfolio-style Ph.D. thesis, there are no easy and straightforward take-aways that can sum up the full project. However, I hope that through this work presented here, there are now better arguments and data available for the questions raised, ranging from charitable giving to reflective equilibrium reasoning in population ethics. Furthermore, the work here lays the groundwork for a variety of follow-up work, some of which I have already conducted, but much more of which is still yet to be done. While I have already conducted follow-up work, such designing a reliable and valid scale to measure anti-natalist attitudes (following up on Chapter 1) or further investigating the mechanisms underlying the Bayesian Truth Serum in this context (following up on Chapter 3), there are several other research projects that may directly build upon this. For example, building on Chapter 2, one may want to use this same mechanism of capturing reflective equilibrium reasoning in different contexts to test whether the pattern of data presented here also generalises to other contexts. Additionally, starting at our results from

Chapter 5, one may want to operationalise the moral demand that we studied as variations of moral language as variations of amount requested instead. This may give us an additional angle to investigate this question and provide further data on the same type of research question. These are just a few of the potential ways one could build on the work presented in this thesis, and I hope that the argument and data from all six chapters will make their way into even more further academic research in one form or another.