UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN & MATHEMATICAL SCIENCES

Social Statistics and Demography

Three Papers on Decision Theoretic Agent-Based Modelling in Demography

by

Jonathan Gray

Thesis for the degree of Doctor of Philosophy

May 2017

For Jen and Poppy.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN & MATHEMATICAL SCIENCES

Social Statistics and Demography

Doctor of Philosophy

THREE PAPERS ON DECISION THEORETIC AGENT-BASED MODELLING IN DEMOGRAPHY

by Jonathan Gray

This thesis consists of three papers, which address Agent-Based Modelling (AM) as a methodology in demography, focusing on the modelling of decision making processes. The discussion begins by assessing the utility of AM as a methodology and some of the issues peculiar to it, and argues that the modelling of choice is of special significance in the demographic context. Of the three papers, the first one outlines an approach to model development combining aspects of game theory, and decision theory. It then contrasts the effect of four choice models on the behaviour of a simulation based on qualitative accounts of the disclosure behaviours surrounding alcohol misuse in pregnancy. The second paper applies this approach to help-seeking in older adult care, drawing on survey data to parameterise and validate the model. Simulation results, and variance-based sensitivity analysis indicate that a model of decision making which incorporates a representation of the interactions between agents are necessary to reproduce observed rates of caregiving. The third paper reports experiments designed to validate the choice behaviour of agents in the older adult care model. I examine human decision making about paired gambles from experience, where the pair has some features common to both choices. I report results for eight decision problems undertaken by 20 participants, and contrast the predictive ability of four models of decision making. I then estimate parameters to maximise the fit where possible, and find that while the best performance is offered by decision models with a representation of the problem, they do not offer a significant advantage over heuristic methods. I discuss the implications of this, in the context of the original agent-based model, and for agent-based modelling more generally.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Jonathan Gray , declare that the thesis entitled *Three Papers on Decision Theoretic Agent-Based Modelling in Demography* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: (Gray et al., 2016)

Signed:...

Date:..

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/H021698/1 Care Life Cycle, funded within the Complexity Science in the Real World theme, and EPSRC Doctoral Training Centre grant (EP/G03690X/1). The use of the Experimental Economics Laboratory, IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work is also gratefully acknowledged. I also wish to thank my supervisors Jakub Bijak, and Seth Bullock for their guidance and support, my examiners David Banks and Steven Glautier for their insightful comments, my parents for endless help and boundless enthusiasm, and my colleagues Jason Hilton, Chris Osowski, Vincent Marmion, Nick Hill, and David Arden for the many useful conversations and cups of coffee instrumental to this process.

Acronyms

ABCD	Agent Based Computational Demography. 17
ABM	Agent-Based Model. 1, 2, 4, 7, 11, 12, 14, 15, 17, 30–32, 81, 93, 94, 107, 108, 113–116
ADHD	Attention Deficit Hyperactivity Disorder. 33
ADL	Activity of Daily Living. 67
AI	Artificial Intelligence. 7
ALSPAC	the Avon Longitudinal Study of Parents and Children. 49
AM	Agent-Based Modelling. v, 3, 4, 6
AML	acute myeloid leukaemia. 33
ARA	Adversarial Risk Analysis. 7, 22
AUC	area under the curve. 103, 104
AUDIT	Alcohol Use Disorders Identification Test. 36
BACCO	Bayesian Analysis of Computer Code Outputs. 8, 51, 62, 81
BDI	Belief-Desire-Intention. 11
CPT	Cumulative Prospect Theory. v, vii, 25, 26, 29, 44, 47, 49, 56, 57, 59–62, 74, 81, 84, 85, 87, 98–104, 106, 114, 129, 133–141, 145–151
DU	Discounted Utility. 25
ELSA	English Longitudinal Survey of Ageing. 67, 68

ESS	Evolutionarily Stable Strategy. 19
ESS	European Social Survey. 69, 70
EU	expected utility. 102–104, 106, 107
FAS	Foetal Alcohol Syndrome. 33, 34
FFH	Fast and Frugal Heuristic. 23, 24, 44
fMRI	function Magnetic Resonance Imaging. 21
GEM	Gaussian Emulation Machine. 8, 30
GEM-SA	Gaussian Emulation Machines for Sensitivity Analysis. 51
GRM	Graded Response Model. 69, 70, 74
IAPT	Improving Access to Psychological Therapies. 35
IQR	interquartile range. 51, 52
IRT	Item Response Theory. 69
MI	mutual information. 82, 87
MM	Maximax. 102, 103
NICE	the National Institute for Health and Care Excellence. 25, 30, 34–36, 49
NM	Natural Mean. 102–104, 107
ODD	Overview, Design concepts, and Details. 4, 5
OFC	orbito-frontal cortex. 21, 22
ONS	Office of National Statistics. 70
OPN	Opinions and Lifestyle Survey. 70
PB	Payoff Bayes. v, vii, 96, 97, 100–104, 106, 107, 114, 135, 137, 139, 141, 147, 149, 151
PND	Postnatal Depression. 38

RCOG	Royal College of Obstetricians and Gynaecologists. 34
RCT	randomised control trial. 37, 38
T-ACE	Tolerance, Annoyance, Cut down, Eye-opener. 36, 40
TFR	Total Fertility Rate. 14
TPB	theory of planned behaviour. 15–17
TTB	take-the-best. 23, 24

Chapter 1

Introduction

This thesis consists of three papers, which address the psychological plausibility of Agent Based Models (ABMs) in demography, emphasising the modelling of decision making. This encompasses a variety of challenges, since while there are many alternative models for decision making, there is no single approach which is universally used. In fact, we might comfortably say that there are essentially as many models of decision making as there are agent-based models, since the tendency is to conjecture a situation-specific decision model to suit the modelling question.

In these papers, I argue for an approach based on models of decision making which are drawn from existing theories of choice behaviour, which are desirable because they are both generally applicable, and can be falsified in isolation from the specific simulation model they are applied in. I first show that such an approach is reasonable by demonstrating an agent-based model based on these principles which reproduces stylised facts about the disclosure of alcohol misuse by pregnant women. I then show that such approaches can be used in concert with a strong empirical grounding for the simulation, by using survey, and local authority data to parametrise a simulation of older adult care. This also allows me to generate not only a synthetic population, but one with plausible synthetic psychologies, whose beliefs and attitudes are representative of the population under study. Having successfully calibrated the resulting model, I then examine the individual level in more detail in order to assess the validity of the agent decision models, using controlled lab experiments on human participants. This is considerably easier to achieve, because the structure of the simulation model is first framed in the abstract as a game, and the realisation applies models of agents which can be utilised outside the specific domain context in which they are initially developed. This allows me to show that the parametrisation of the agent-based models used in the simulations is likely to be incorrect, suggesting that the successful calibration exclusively at the macro level is misleading. This amply demonstrates both the need for controlled experimental validation of the individual level of agent-based simulations, and the possibility of doing so, where the model is developed with facilitating this process in mind. The overarching aim then, is to show that calibration and validation of agent-based models at both the micro, and macro scale is necessary, and to show a path for doing so.

1.1 Why agent-based modelling?

Agent-based modelling, sometimes referred to as Individual-based modelling is an approach to modelling systems as composed of many autonomous, interacting 'agents', with individual characteristics. In a population sciences context, this equates to a bottom-up approach to modelling based on the individuals within the titular population.

Although there are an increasing number of examples of the use of agent-based modelling, and simulation in the social sciences more generally (Niazi and Hussain, 2011), it is still not a mainstream technique in population sciences. Considerable efforts in this area have been undertaken, beginning largely with Billari et al. (2003b), who applied agent-based modelling to study the evolution of norms surrounding the age of marriage. Billari et al. (2003a) also argued for agent-based models as a new approach to research in population sciences, a call subsequently reiterated by Billari et al. (2006), and again by Billari (2015), where the authors argue that agent-based models are a crucial tool in understanding population chance. Despite these considerable efforts, the use of simulation modelling in demography remains limited, and the question of why simulation, as opposed to more tried and tested approaches to demography, is the route I have taken is worth addressing.

In part the answer to the question, is that the concern is with the methodology itself, and how it can be improved. There are, after all, reasons to not be altogether comfortable with agent-based modelling. Axelrod (2006) gives an account of the resistance the author experienced when attempting to publish an ABM examining the evolution of sex (Hamilton et al., 1990). He suggests that the challenges arose in part from an unfamiliarity on the part of the reviewers with agent-based modelling as a technique, but also from a degree of mistrust about the robustness and generality of the simulation results. In fact, Axelrod (2006) notes that the mistrust is not wholly unjustified, but arises from drawing a faulty analogy between simulation, and analytical methods. Part of this challenge has been eroded by time, and familiarity, in that agent-based models are now considerably better known. Concerns about robustness, and generality persist, and perhaps not without good reason.

More recently, Waldherr and Wijermans (2013) surveyed colleagues working with agent-based modelling to discover what the barriers to acceptance by the mainstream were. They identified ten broad categories of objection to agent-based modelling, some of which are in a similar spirit to those encountered by Axelrod (2006). Key concerns were that models were unrealistic, with arbitrary assumptions and parameters, and that results were designed into the model. These are fundamentally concerns about the robustness and validity of agent-based modelling as a technique. In fact, that such concerns continue to be present is a prompt to the modeller to address them more fully. Other causes for scepticism of Agent-Based Modelling (AM) have been discussed by Leombruni and Richiardi (2005), who discussed challenges presented by the interpretation, generalisation, and calibration of simulation models in economics. All three relate to the perceived lack of rigour of simulation in comparison to traditional analytical, or statistical approaches, which the authors attribute once more to unfamiliarity with AM. To address these concerns, I perform global sensitivity analyses of the simulation models, the approach to which is discussed more extensively in chapter 3.

A second class of objections is that models are too simple, or excessively complex, or perhaps both at once (Waldherr and Wijermans, 2013). This is more challenging to address, since the right level of complexity is to some extent a matter of personal taste. The objection from oversimplification arises from a rejection of the idea that simple models can ever hope to be informative about the richness of reality. This objection is clearly not specific to agent-based modelling, since it could be applied to any modelling approach to science. The model is *not* the modelled, it is a more tractable and readily understandable representation of it. Concerns about models being excessively rich come from the opposite perspective. In this view, agent-based models are typically considerably more complicated than analytical ones, ergo *the analytical modelling approach is better*. This is in some ways inarguable, in that by contrast to analytical approaches, agent-based modelling tends towards the inclusion of more features at the level of the individual. A modelling paradigm which dwells on the individual components of a system is of course sometimes necessary, because it facilitates the capturing of heterogeneity, feedback loops, and interactions at multiple levels (Axelrod, 1997).

The ability to incorporate complexity is a great part of the appeal of AM as a methodology, but also presents tremendous risk, in the opportunity to arbitrarily decide what is included in a simulation, and hence implicitly encode what is important. This applies not only to information, but to human behaviour, often leading to the introduction of an *extempore* theory of human behaviour by the modeller. This introduces the potential for implicit axioms underpinning the results of simulation, which as I show in chapters 3, and 4 can manifestly impact overall behaviour. To some extent this can be seen as a particular manifestation of the challenge and opportunity characteristic of simulation models, which tend towards the expressive as analytical approaches do towards concision. This richness is powerful in complex domains, where non-linearity and emergent phenomena are common and more compact traditional approaches struggle, but can lead to models barely more tractable than reality. This introduces significant challenges to both validation, and verification in that a given model may represent the composition of many conjectured processes.

This should of course imply that agent-based modelling is not always the right, or only, tool for a particular task. In fact, I argue that agent-based modelling is most useful when combined with analytical, and statistical approaches to a problem. Nonetheless, striking the balance between too simple, and over-complicated is a challenge, particularly given that the nature of ABMs is to invite the observer to consider 'what would happen if. . . ' (see Epstein (2008)).

Perhaps the most damaging concern is that models are not simply over-complicated, but too complicated and too convoluted to understand. The potential for ABMs to become 'black boxes' which can neither be understood, nor replicated, has also been expressed by Lorek and Sonnenschein (1999). The authors discuss the excessive technical demands placed upon the reader, by the dual nature of simulation models, and the many programming languages and ecosystems they can exist in. If the simulation is effectively the only expression of the model, then the reader is faced with the responsibility of understanding the implementation to be able to evaluate the model. One mooted suggestion for resolving the opacity of AM is to use a standard for documentation, such as the Overview, Design concepts, and Details (ODD) protocol (Grimm et al., 2006). This provides a standardised framework for the reporting of ABMs, with the aim of both formalising such description, and making models more readily understood by the interested reader. Grimm et al. (2010) report considerable success in the use of the ODD protocol, but do note that it has been criticised as containing redundant elements, being rather excessive for simpler models, and not fit for every purpose.

My approach is to address model development rather than *post hoc* description, by employing the language of game theory which provides a developed and well understood framework for abstracting complex real world interactions. This facilitates transformation into a form amenable to simulation, with the added benefit of translating readily into real-world experiments. This approach also carries an additional advantage in common with the ODD approach, in that it requires explicit statement of assumptions. In a sense this approach positions the game theoretic model at one extreme, and reality at the other. The game is the simplest, most abstract representation which we hypothesise captures the important characters and interactions. The agent-based model, as a realisation of the game, can gradually reintroduce aspects of reality, while at the same time giving us access to the process which produces the observed phenomena. This is also advantageous, in that the underlying game structure supports experimental validation of each increment towards reality in a controlled setting. Naturally the process is not as entirely linear as this characterisation suggests, since the accretion of elements of reality is not simply additive, and does not progress in a strict ordering of 'now time, now time and mortality' and so forth. This is another key reason to make sensitivity analysis a central part of the modelling process, because it supports the modellers' judgement about which aspects of reality are most important to explaining the phenomena of interest.

1.2 Significance of Decisions in Demographic Simulation

The need to address decision making specifically in the agent-based modelling context arises when we consider that the ability to make choices is indivisible from the quality of agency. From this follows the maxim of the intelligent agents community, which roughly stated is 'agents are not objects' (Wooldridge, 1999). The distinction in this case rests on the ability to make decisions, since an object in the software context may have requests made upon it which it will always fulfil. An agent, by contrast, invariably has the option of declining a request, should it decide to do so.

In the context of demography, where we typically consider large numbers of agents this idea might be extended to 'agents are not atoms'. To see why, a brief digression into statistical mechanics is necessary. Rather than a population of a country, we might consider a gas sealed in a jar. To determine the pressure, temperature, and other characteristics of the gas, we can in most circumstances largely ignore the behaviour of the individual particles which make up the gas. Instead, we can comfortable employ the ideal gas law, which is based on aggregate information about particles which behave like billiard balls (Adkins, 1968). This ceases to be true at the extremes of heat, and pressure. At these places, the behaviour of the individual particles within the gas becomes significant, and the macro scale law is no longer able to predict.

The analogy to gas is apt, if we consider that macroscale prediction in that case is dependent on a degree of equilibrium. This has parallels in the social sciences, and indeed, a central argument for the use of AM in social science is that existing methods break down away from equilibrium (Arthur, 2005). Returning momentarily to the gas analogy, we can allow that where equilibrium reigns, the macro scale behaviour of the gas can successfully be predicted. However, the model that allows this does not contribute to our understanding of the particles of the gas, and moreover, relies on unnatural assumptions about them, since particles are a really nothing like billiard balls.

This raises the question of whether modelling at a single scale is sufficient. If we take the view that society is the macro scale manifestation of the actions, reactions, and interactions of heterogeneous individuals, then we must also acknowledge that behaviour occurs in the context of society. In other words, there is a reflexive interscale relationship, and it follows that at least tacit acknowledgement of that is likely to be beneficial. This is naturally part of the strength of agent-based modelling, since it allows such feedback loops to be captured (Billari et al., 2003a, 2006; Billari, 2015).

If we are to consider a phenomena at multiple levels, the appropriate level of abstraction for them depends on the nature of the inquiry, for example, a particulate abstraction of individuals is sufficient to capture emergent crowd behaviour (Silverberg et al., 2013). By contrast if, as is typically the case in demography, the focus is on processes of population change, we may need agents rather than atoms. This is because the decision to jump in a particular direction in time to a beating drum is qualitatively different to the decision to, for example, migrate to another country. Migration is deliberative decision, whereas the motions of dancing, having begun to dance, are instinctive[1] and immediate.

[1] While this is generally the case, there are those of us for whom dancing is very much a conscious and deliberative process.

Taking a popular population statistic as an example, crude birth rates are an aggregate snapshot of a plethora of individual decisions, and interactions of which actual mating is only a minor part. These decisions are not taken in isolation, but are part of a complex web of choices at the individual level, which are influenced by the wider social, economic, and political conditions at the time.

The reasons for drawing on decision theory to give agency to the agents parallel those for suggesting game theory as an approach to developing the underlying analytical model for an ABM, in that the decision rules are well specified, and grounded in a strong theoretical base which I explore in more detail in chapter 2. The most significant argument for incorporating the decision-theoretic approach, however, is to ameliorate the substantial limitation that the lack of consensus on modelling decision making imposes (Klabunde and Willekens, 2016). In fact, this is also beneficial in the other direction, since it offers the potential to test the capability of theories of decision making in an entirely new way, at the scale of populations.

The success of an ABM in explaining a large scale social phenomena does not of necessity imply that the decision model used by the agents works, nor does an ABM which performs poorly as such falsify it. However, the process is informative about how the decision model performs in a more complex and interactive context than is typically the case, which invites further development of the theory of decision making to address these factors. Situating this in the context of demography is by the nature of the field also placing it in a strongly empirical environment, which offers the potential to collect data at massive scales. As suggested by Van Bavel and Grow (2016), this can also help to address the issue of developing behavioural models based on over representation of 'weird people'[2] (Henrich et al., 2010b,a).

An approach which combines game theory and decision theory is not in itself a novelty, and is both familiar to the Artificial Intelligence (AI) community (Parsons and Wooldridge, 2002), and similar to that of what Ríos Insua et al. (2009) term Adversarial Risk Analysis (ARA). Here, it is used in a context where rather than seeking optimal decisions, as in the former, the desire is for plausibly human choices in a quasi-collaborative, rather than strictly adversarial environment as in the latter.

The use of simulation in concert with statistical modelling can be a powerful tool to predict and

[2]There is a notable over representation of experimental participants from Western, Educated, Industrialized, Rich, and Democratic societies in behavioural experiments.

explain social systems, and understanding individual choice behaviour is a critical component to this not adequately addressed by current ABMs. I argue that incorporating decision theory into agent-based modelling can help resolve this, and forms a symbiotic relationship, strengthening the descriptive strand of choice modelling while enhancing the plausibility of simulation. Applying this in the environment of demographic modelling bolsters the theoretical component thereof by providing a theoretical grounding for the micro-level of populations, which can then be built from to better understand the macro. This also provides a more general benefit, by providing an invaluable arena to critically evaluate models of decision making.

1.3 Outline of the Thesis

The remainder of this thesis proceeds to provide background (chapter 2) covering agent-based models as applied to demography (section 2.1); game and decision theory with a focus on signalling games (section 2.2); followed by normative (section 2.3), heuristic (section 2.4), and descriptive (section 2.5) accounts of decision making. This review aims to situate the papers which follow in the domain context, to wit, demographic methodology, as well as the theoretical context from which the approach to modelling behaviour and interaction is drawn.

This is followed by the first paper (chapter 3), which expounds the decision theoretic agent approach, and demonstrates the effect of different choice models on the overall model dynamics. This is motivated by an example exploring the disclosure of alcohol misuse in pregnancy, which attempts to incorporate insights about barriers to help seeking drawn from qualitative research. This paper also introduces the Bayesian Analysis of Computer Code Outputs (BACCO) approach to sensitivity analysis of simulation models, using Gaussian Emulation Machines (GEMs), which is applied again in the second paper.

The second paper (chapter 4) provides a more refined elucidation and implementation of the model and the wider agent-based approach, and applies it in a context where I am able to confront, and infuse the simulation with data (section 4.2). This moves the substantive domain to older adult care, which illustrates the potentially broad applicability of the disclosure model, and more generally the utility of the decision-theoretic agent-based approach. This domain also offers considerably greater quantitative data to parametrise, and calibrate the model. The paper draws on several large-scale social surveys, in addition to local authority expenditure data, used to empirically ground parameters where possible. A global sensitivity analysis is then performed, followed by calibration of those parameters which cannot be evidenced using the metamodel produced in the course of the sensitivity analysis. The model is successfully calibrated for two decision rules, and a further perturbation analysis is used to explore the impacts of possible policy interventions.

The final paper (chapter 5) demonstrates the possibility of validating agent-based models at multiple levels, by exploiting the computable nature of simulations. The structure of the model used in chapter 4 is shorn of the context, and becomes a monetary decision problem. Simulations of the agent decision models attempting to solve the problem are then used to generate variations on the decision problem which maximise the differentiation in their predicted actions, and the behaviour of human participants when confronted with the resulting problems is then tested. The paper then examines the performance of the agent-models in predicting real behaviour, and compares them to other alternative decision models. This identifies a potential issue with the apparently successfully calibrated older adult simulation model, and highlights the pragmatic trade-offs of using simulations as a tool in demography.

Lastly, chapter 6 summarises the results of the three papers and their collective contribution, before discussing some important limitations of the work, and the approach more generally. The benefits and trade-offs of different decision models are discussed in relationship to the modelling context, in addition the relationship between simulation design and the ability to validate a simulation. Finally, potential avenues for future research into models of decision making are suggested.

This cycle, where qualitative work informs simulation, which can then be linked to quantitative data, and subsequently inform further experimental and data gathering work, which further enhances modelling, is suggestive of the potentially synergistic role for simulation in the social sciences. The overarching argument I have given here is in the line of constructionism, in the tradition of Braitenberg (1986), and Resnick (1994). This holds that, if we want to understand a thing, it is instructive to try to build it, and to do that it is indispensable to know how it is expected to behave (Franck, 2002).

Chapter 2

Background

This chapter provides a broad overview of the context in which this work is situated. It reviews relevant aspects of previous work covering agent-based models as applied to population studies and the role of agent-based modelling in demography, as well as providing background on signalling games, and in addition normative, heuristic, and descriptive/positive theories of decision making.

2.1 Agent-based Modelling in Demography

This section reviews agent-based modelling as applied to demography, and population studies an area pioneered by Billari et al. (2003a), encompassing both individual models, and work examining the role of the methodology in this domain.

Demography is fundamentally the study of aggregated life courses (Billari, 2015): the manifestations at population level of key events in the lives of individuals. This places an emphasis on particular classes of events, namely births, deaths, partnership formation, and migration. Death, as an event which is rarely a matter of agency on the part of the individual, is typically incidental to the focus of ABMs, although Mesoudi (2009) has used agent-based modelling to examine the role of social learning in 'copycat suicides'. The impacts of mortality, as opposed to mortality as an activity, have been investigated by Ewert et al. (2003), in a simulation of pre-modern mortality crises. The model is somewhat unusual in including a mixture of behavioural models for agents, with agents driven by a Belief-Desire-Intention (BDI) model, as well as utility maximisation. Because the simulation is not linked in any quantitative way to the empirical realm, it lies very much in the tradition of simulations as 'opaque thought experiments' (di Paolo et al., 2000), serving as a hypothetical disaster simulator.

In the following sections I review ABMs applied to key focus areas of population science: partnership formation, fertility, and migration; before discussing and critiquing the role of agent-based models in population studies.

2.1.1 Partnership Formation

Partnership formation is a key focus of demographic interest, as a significant life event, and more critically because of the role it plays in fertility, and hence population change. Todd et al. (2005) present a rather abstract model which simulates patterns in the distribution of age at (first) marriage, demonstrating that population level behaviours can be produced using simple rules operating at the individual level. The model represents a considerable departure from earlier approaches, by assuming that agents lack complete knowledge, and use heuristic methods to decide who to choose as a mate. Todd et al. (2005) note that a key finding is that considerable heterogeneity is required amongst the agents to satisfactorily reproduce the observed patterns in the data. The subsequent 'Wedding Ring' model, by Billari et al. (2007) addresses the same topic, but focuses instead on the role of an individual's social network on the decision to marry. This addresses a shortcoming in the Todd et al. (2005) model, but largely dispenses with any modelling of the decision process, making marriage a function of mutual circumstance, rather than an active decision on the part of the agents. Both models also tend towards the abstract, with neither population structures or agent characteristics grounded in empirical data. Bijak et al. (2013) have subsequently extended the 'Wedding Ring' model to include a semi-artificial population based on the characteristics of the United Kingdom, arguing that linking the model to real population data augments predictive capability. This follows the general line of argument expressed by Silverman et al. (2011), who suggested that agent-based modelling could be used to ameliorate the limitations imposed on demography by the need for ever greater amounts of data, by enhancing the theoretical aspect. Similar perspectives have been espoused by Billari et al. (2006), and Epstein (1999), who have argued for bottom up, 'generative' approaches to social science.

A different aspect of partnership formation – the transmission of norms surrounding age-at-marriage – has also been examined using agent-based modelling, by Billari et al. (2003b). They employ a stylised model of marriage, where possible partners are constrained by inherited age bounds. While the model is very abstract (agents have only one sex, and no demographic characteristics), the authors identify a variety of conditions under which norms persist, or die out, and where heterogeneity in norms may be sustained. A subsequent extension by Aparicio Diaz and Fent (2006) addresses the stylised form of the original, by incorporating additional demographic features in the form of age, and sex. They also introduce heterogeneous norms between the sexes which for female agents bound the age at which they consider marrying, and for males the acceptable age bracket of potential partners. This added complexity allows the model to offer a potential explanation for an observed U-shaped curve in historical mean age at first marriage in six European nations. As with the 'Wedding Ring' model, the actual decision to marry is unexplored here, and takes the form of an exogenous matching procedure contingent on the norms of the agents, although the authors acknowledge that the reality is more complex.

2.1.2 Fertility

In addition to the formation of partnerships, the fertility outcomes of those partnerships are also of great interest to demographers. Aparicio Diaz et al. (2011) provides the earliest example of the application of agent-based modelling to understanding the drivers of fertility behaviour. The authors use a social network model, inspired by that developed by Watts et al. (2002), combined with a simple agent model to investigate the impact of an individual's social ties on how they balance their desire to attain an education, and their desire to become a parent. The resulting model is able to reproduce changes occurring in the probability of having a first child which occurred between 1984, and 2004 in Austria, and the authors subsequently apply the model to forecast age specific fertility rates up to 2016. This forecast suggests a shift towards later fertility, and predicts that this will be slightly more pronounced than is indicated by existing methods. This is particularly interesting, in that the accuracy of the prediction can now be quantitatively evaluated, because the change in age specific fertility rates is now known. Perhaps because the focus in this instance is primarily on the role of peer influence, the agent-models are a simple threshold function. The most obvious criticism of the model is the conceptualisation of intended education, drawn from empirical data, as a static, determining factor in the construction of an agent's social network, rather than in any way a function of it. This makes intended education level an exogenous driving force, where it might be expected to be in part a function of the social interactions, and to vary across an individual's lifetime.

Kashyap and Villavicencio (2016b) have examined a different aspect of fertility, looking at choice about the sex of children, rather than the decision to have them. The authors use a two-stage model, to demonstrate that increased availability of the technology necessary to perform sex-selective abortions can account for changes in parity specific fertility in South Korea. The authors perform an extensive sensitivity analysis, and calibration of the model against empirical data, with the primary finding that relatively low levels of son preference can produce a skewed sex ratio in combination with early and consistent increased access to sex-selective abortion. The agents in these simulations use a series of thresholded stochastic production rules to determine whether an abortion takes place, driven by a son preference level drawn from based on survey data, while conception itself is probabilistic. This leads to an interesting scenario, where there are two sexes of agent, but only one is active – male agents are essentially a commodity, and do not contribute to the decision making process, or play a role in conception. In contrast to Aparicio Diaz et al. (2011), the agents in these simulations have a global perspective, with decisions motivated in part by the whole system state – for example, the agents' 'readiness' to use sex selective abortion is a function of the Total Fertility Rate (TFR) at that time for the whole system. A potential extension then, would be to explore whether the model were still able to perform when agents have access to only local, or noisy information about the state of the system. Kashyap and Villavicencio (2016a) have also successfully applied the model to India, but in both cases the authors note that the availability of information about son preference levels is a limiting factor in applying it to other regions, as is an absence of spatial heterogeneity in preference levels within the model.

2.1.3 Migration

A major focus for ABMs in demography is migration, as both a significant driver of population change (Bongaarts and Bulatao, 2000b), and, particularly in recent years, an increasingly vexed topic in the arenas of public opinion, and political discourse. Projections of migration flows are also subject to a significant degree of uncertainty (Bijak, 2006; Bongaarts and Bulatao, 2000c), and error, for example Bongaarts and Bulatao (2000a) noted that World Bank and UN forecasts of net international migration rates leading up to the year 2000 were only directionally correct in 60% of cases. This suggests that classical approaches face serious challenges, and makes modelling of migration an attractive area for agent-based modelling.

Migration, as an application domain, covers a broad remit, addressing not just international migration, but also internal, and intra-city migration. Wu et al. (2008) have examined the latter, comparing the ability of an agent-based model to represent student populations in Leeds, versus microsimulation approaches. Agents in the model use a set of heuristics to decide whether to remain in an area, or to move to another, with differing strategies dependent on their stage of study. They find that by using heuristics which capture the desire of students to live near their fellows, and to live near the university, a more realistic distribution of population across the city can be produced. A potential issue with the model however is that the possible migration sites are drawn from known student accommodation, which may suggest that the results of simulation in terms of spatial distribution are rather predetermined. This is a recurring concern about agent-based modelling approaches, identified by Waldherr and Wijermans (2013), who found that those sceptical of agent-based modelling often felt that results were ‘built into the model’.

At an international scale, return migration has been examined by Klabunde (2014), who developed an ABM of return migration in Mexican emigrants to the USA, based on microdata from the Mexican Migration Project. In this model, agents exist within a social network, and use local information about employment conditions to probabilistically migrate. Return migration is a function of their individual desire to return home, social ties, and current wealth. The decision mechanisms in both case are thresholded random number generators, embodying heuristics which correspond to the hypotheses to be tested. The model successfully reproduces stylised facts, and the author performs extensive calibration and robustness analysis. While a potential criticism arises in the specificity of the behavioural rules, their role as a heuristic representation renders this less troubling, and they could be readily expressed in terms of a more general mechanism.

More recently, Klabunde and Willekens (2016) have reviewed a wide range of ABMs focused on migration of various kinds, with an emphasis on how the models address individual decision making. They identify choosing a behavioural model as one of the most significant challenges in the design of ABMs, and suggest that discrete choice theory is a natural candidate for modelling agent decision behaviour. This is perhaps an excessively general categorisation, given the variety of models which could be classed as discrete choice, a number of which are reviewed in sections

2.3, 2.5, and 2.4. In addition the authors do not clearly distinguish between models of behaviour, and models of decision making, conflating utility maximisation with higher level models such as the theory of planned behaviour (TPB) (Ajzen, 1991). While this is not a crisp distinction, TPB does not as such address the process of making a decision, but rather the process of generating and valuing potential actions, and sequencing the stages of the process. While both processes - generating options, and choosing between them - are critical components of a full model of behaviour, they are not sufficient in isolation.

Willekens (2016) focuses on (e)migration as an individual decision, with considerable focus on the stages of decision making. The author introduces a 'process model' of decision making, based on TPB, which delineates the decision to emigrate into stages of acquiring an interest in doing so, developing an intention to emigrate, and the actual act. Transitions between states are a function of social factors, and characteristics of the individual agent, mediated by a fixed transition probability. The model is able to produce stylised facts about emigration, and suggests that many more people intend to emigrate than actually do. Concerns similar to those expressed about Wu et al. (2008) can be raised, in that the transition probabilities are set to produce these stylised facts. Absent a sensitivity analysis then, it is perhaps unclear how far the dynamics of the model rest on these probabilities rather than the decision model. The author also notes that the model disregards return migration, suggesting that this explains some of the discrepancies between the model predictions and real world populations.

2.1.4 Critique

The role of agent-based modelling in demography is worth discussing, particularly in the context of a science with a heavy emphasis on forecasting and prediction. It is germane then, to consider where on the prediction-explanation spectrum agent-based models lie. While it may seem that any model which purports to offer a sound explanation of a phenomena should also be able to make predictions, Epstein (2008) has argued that prediction is only one of many reasons to use modelling in science, emphasising the distinction between prediction and explanation. Epstein points to plate tectonics as an example of a model which explains, but does not predict specific earthquakes. The suggestion that prediction is not a necessary component of a sound explanation has been criticised by Thompson and Derr (2009), who argue that a pre-emptive defence of agent-based modelling is misguided. The authors point out that there is a clear distinction between a complete prediction, for example of earthquakes as point events, and prediction in a more general sense, i.e. that earthquakes will occur and are more probable in certain regions.

The thrust of this position is that agent-based modelling is a 'generative' approach (Epstein, 1999), which permits us to put to the test whether a minimal set of individual level rules can produce a phenomena of interest at the macroscale. From this perspective, part of the utility of a simulation model is as a constructive (Epstein, 1999), or existence (Axelrod, 1997) proof. The implication of this claim is that the existence of the simulation model is proof that, if we accept the assumptions the model is built on, then the mechanisms included in it can produce the phenomena observed. This follows in the tradition of Braitenberg (1986), and Resnick (1994), who suggest that the process of constructing a replica, even if the effective implementation differs from the original, can be instructive. Epstein and Gang (2006) have subsequently argued that the generative approach is not without limitations, pointing out that the ability to generate a phenomena is not identical with an explanation of it, the former being a necessary, but not sufficient condition for the latter. Agent-based modelling has also been criticised by Grüne-Yanoff (2009), who argues that the majority of ABMs are incapable of providing full explanations because they are not validated, or are not amenable to validation. The root of this objection returns to the question of whether explanation is, or should be identical with prediction. Elsenbroich (2012) has responded to this, arguing in essence that a mechanistic explanation which does not perfectly reproduce reality is still a useful explanation, although the authors do acknowledge that agent-based modelling has been unsuccessful as a predictive tool.

As can be seen from this selection, agent-based modelling in demography has been applied to a diverse range of problems. A lack of consideration of the decision making process of agents is a common feature, with considerable reliance on case specific heuristic approaches. Where more of the modelling effort has focused on this facet, as in the migration model by Willekens (2016), the result is more general, in the sense that the framework of TPB is not specific to migration. However, the actual implementation is as a thresholded random number generator which is highly idiosyncratic to the scenario at hand.

In the specific context of demography and population studies, Billari et al. (2003a) have argued for what they term Agent Based Computational Demography (ABCD), suggesting that it could potentially bolster what they perceive to be a weak explanatory tradition in the field. Burch (2003a) has extended this idea, and argues that the strongly empiricist approach of the majority of demography underlies the weakness of the theoretical aspect, suggesting that the effective relaxation of empiricism imposed by agent-based modelling approaches are, in this respect, beneficial. Billari et al. (2006) have also argued that part of the role of agent-based modelling is as a theoretical counterpart to microsimulation approaches, offering the potential to link the micro and macroscale, while facilitating the modelling of heterogeneity in populations, and the presence of feedback loops. Recent work by Courgeau et al. (2016) builds on these trains of reasoning, and considers the future role of agent-based modelling in demography. The authors advocate for what they term a 'functional-mechanistic' approach, which embeds computational modelling within a cyclical process of data collection, analysis and inference about the underlying structure and drivers of the observed world, and conceptual and mathematical modelling of phenomena. The authors also argue for the role of controlled experiments, and interdisciplinary engagement in this process, noting similar calls by Conte et al. (2012) when discussing computational social science more generally.

In the following sections, I review more general approaches to modelling interactions and mechanisms for decision making.

2.2 Signalling Games

This section reviews literature focusing on signalling games, which form a key part of the modelling process used throughout this thesis.

In some cases, the games studied in game theory are those with 'perfect', or 'complete' information, where every player knows the actions by the others, and has access to all the information used to decide on those moves. Moves by players are assumed to be perfectly rational, and account for the equally rational moves of their opponents. An alternative situation which is perhaps more common in reality, is that players information which is incomplete in some way. They may lack information about the possible actions available to their opponents, or to themselves, or alternatively may not have complete knowledge of the possible payoffs in the game. This second case was essentially unexplored, until Harsanyi (1967) introduced the concept of a Bayesian game. This treats scenarios where players have incomplete information as separate games in their own right, so that each possible variation on the rules of the game is in effect a subgame. This adds an additional player - nature - to the game, where nature takes the first move thereby deciding which subgame is played. Nature is assumed to choose its move by lottery, and where the probability distribution governing the lottery is known to all players, the game can then be formulated as one of complete information. Ríos Insua et al. (2009) also address this scenario, but weaken the assumption that the probability distribution is known to all players and advance a practical sampling-based method to permit players to reason about their opponent's actions.

Here, we are specifically interested in signalling games (Spence, 1973; Kreps and Cho, 1987), a type of Bayesian game where one player holds some private information which may be communicated (or not) by means of a signal, known as their 'type'. This basic form has been widely applied, with substantial interest in what conditions permit honest signalling as a Nash equilibrium, where no rational player would willingly change their strategy because they reason it would make them worse off (Nash, 1951). Others have considered the conditions under which honest signalling would represent an Evolutionarily Stable Strategy (ESS), where if a population adopts the strategy, natural selection will suffice to preserve it against alternative strategies (Smith and Price, 1973). A critical factor in this occurring has been found to be that signalling comes at a cost. For example, in the context of animal behaviour, Grafen (1990), following from a suggestion by Zahavi (1975), proposed that if signals intended to indicate mate quality exacted a cost on the signaller (e.g. peacock tail feathers), then honest signalling would constitute an ESS. Similar results have also been demonstrated in a game of job market signalling by (Spence, 1973), who found that honest signalling would constitute an equilibrium only where the signal (in this case, obtaining a level of education) was more costly for those who were less capable.

Costly signalling has also been suggested as an explanation of behaviour that at first glance appears counter-intuitive, for example Godfray (1991) applied the idea to the food solicitation behaviour of chicks, where a stronger signal (i.e. more, or louder chirping) carries a risk of being eaten by predators. The logic in this case being that the risk of death forces chicks to only demand food when they truly need it. Moving beyond animal behaviour, Sosis (2003) considered the implications of ritual behaviour, in the context of religion, representing a costly signal. This idea has subsequently been extended by Henrich (2009) to include cultural transmission, and Wildman and Sosis (2011) to introduce group differentiation.

Other work augments the signalling game model, for example Austen-Smith and Fryer Jr. (2005) add a second 'peer group' audience signalling game to the original Spence game in an effort to explain poor academic performance in some social groups. In essence the model attempts to capture the balance between the social requirement to be ‘cool’, and the longer term economic benefits of academic success. Fryer Jr. and Torelli (2010) have subsequently examined this empirically, using longitudinal survey data to construct a measure of adolescent social status. They found significant differences in the relationship between academic success and popularity between ethnic groups, which was particularly pronounced in diverse student bodies.

On a similar tack, Feltovich et al. (2002) introduced additional, noisy, type information to signalling games. This means that signalling is not the only source of information about a player’s type, and it can be inferred from observation. They found that in scenarios where type was interpreted as indicative of quality that this could explain apparently counter-intuitive observed behaviour where actors engaged in ‘counter signalling’, by claiming to be of low quality when they were not.

2.3 Normative Decision Theory

The following three sections discuss theories of decision making, which form the basis of the agent decision models used throughout this thesis. In particular, this section examines normative approaches to decision theory.

While game theory addresses rational *strategic* decision making, decision theory deals instead with rational decision making (Peterson, 2009). Taken literally, this leads to normative decision theory, where the focus is on giving the rational answer to a decision problem. An alternate perspective, suggests that decisions are rational only in ecological context, and heuristic in nature (Gigerenzer and Goldstein, 1996) (section 2.4). Finally, a complementary view - descriptive, or behavioural decision theory, holds that the focus should instead be on giving an account of human decision making performance, complete with observed deviations from perfect rationality, which I address in section 2.5.

The conceptual underpinning of all of these theories is the central idea of expected utility, originated by Bernoulli (1738) and later formalised by Von Neumann and Morgenstern (1953). Utility in this sense is the subjective value of an outcome, often expressed in monetary terms. The notion of expected utility incorporates the uncertainty of obtaining that outcome through multiplying the probability of the outcome occurring, by the subjective value of it. The expected utility of an action is then simply the weighted (by probability) sum of the possible outcomes of it. From this, Von Neumann and Morgenstern (1953) provide an axiomatic definition of the rational decision maker, resting on *Completeness*, meaning a decision maker can always decide which of any two preferences they prefer; *Transitivity*, which guarantees that choices are consistent such that if the decision maker prefers A to B, and B to C, they must also prefer A to C; *Independence*, meaning that irrelevant options do not affect the decision makers preferences, and *Continuity*, which states that if the decision maker prefers gambles A and B, to gambles B and C, there is a combination of A and C for which the decision maker is indifferent between that and gamble B.

The extent to which expected utility provides a good model of human behaviour is subject to some debate, and I review several alternative models in sections 2.4, and 2.5. One of the earliest criticisms is the Allais paradox (Allais, 1953), which claims that most people's choices in hypothetical lotteries violate fundamental axioms of expected utility. Harrison (1994) has however argued that this does not occur when dealing with real, as opposed to hypothetical lotteries, highlighting experimental results by Conlisk (1989), which are consistent with expected utility if real monetary outcomes are used. Burke et al. (1996) have subsequently replicated these experiments, and their findings support the general conclusion that real monetary outcomes do have an effect. However, they point out that the result is not that no violations occur, only that they are reduced. Oliver (2003) also found evidence for the paradox in a non-monetary context, showing that participants violated the independence axiom of expected utility when considering choices over health outcomes. More broadly, expected utility has been described as 'grossly inadequate' (Tversky, 1975) as a theory of decision making under uncertainty, for anything beyond normative uses. It should be noted, however, that none of the objections raised dispute *per se* the suggestion that people compute some approximation of utility and use this in decision making. Objections instead concern the way in which utility is calculated in the basic model, suggesting it is flawed or incomplete.

Recently, several studies have explored biological correlates of aspects of expected utility. The fundamental concept, that all outcomes are comparable in a universal currency has been supported by evidence of neural correlates of decision variables (Platt and Glimcher, 1999), and following from this results from Padoa-Schioppa and Assad (2006, 2008) showing neuronal firing in the orbito-frontal cortex (OFC) corresponding to revealed preferences in monkeys. Additionally, some support for neural representation of value, and risk aversion was found by Christopoulos et al. (2009), who used function Magnetic Resonance Imaging (fMRI) to monitor brain activity in human subjects playing gambling games. In addition, work by Soon et al. (2008) has examined similar processes in human decision making, by using fMRI to examine active brain areas during a motor-decision task. They found that the outcome of a decision was encoded in brain activity as much as ten seconds in advance of the individual becoming aware of the decision. Some concerns about the validity of fMRI as a technique have been raised, most notably by Bennett et al. (2010), who found measurable neural activity in dead salmon. However, that the outcomes of decisions are encoded in neural activity is relatively uncontroversial at this point. Arguably the most significant finding by Soon et al. (2008) is the substantial delay between the decision being made, and the individual becoming aware of it. This raises the question of how far introspective accounts of decision making are to be trusted, since there is reason to suspect that they may be post hoc rationalisations.

The models presented in chapters three, and four or this thesis make an explicit assumption that social decisions utilise the same process, and while this is less well supported there is some evidence to suggest involvement by the same brain region, since damage to the OFC has been shown to impair social judgements in both primates (Watson and Platt, 2012), and humans (Willis et al., 2010).

An alternative normative model of decision making is Bayesian decision theory, proposed by Robbins (1964), which is essentially the application of Bayesian probabilities to the expected utility model. This allows probabilities used in reasoning to be subjective, which may allow for a better account of decisions from experience (see Hertwig et al. (2004); Hau et al. (2008) for results elucidating the distinction, and comparing the performance of several non-Bayesian models). This model has seen notable successes in practical problems, such as, developing optimal foraging behaviour in uncertain environments (McNamara and Houston, 1980), planning sensor actions in partially unknown environments (Kristensen, 1997), and effective management of natural resources (Dorazio and Johnson, 2003).

Both Tenenbaum et al. (2006), and Griffiths et al. (2010) have also suggested that Bayesian inference can constitute an effective top-down model of human learning, arguing that the brain is effectively near optimal, and can be modelled as such. Bowers and Davis (2012) have strongly criticised this, arguing that empirical evidence for Bayesian models of learning is weak, and that the flexibility of priors and utility functions introduces unreasonable challenges to the falsifiability of such models. Gallistel (2012) has also addressed six specific features of animal learning around extinction conditioning from a Bayesian inference perspective. Gallistel (2012) claims that Bayesian inference is used to reason about the relationship between causes and effects, thereby guiding conditioning. This model has been criticised by Miller (2012), primarily on the grounds that it leads to optimal behaviour which simply does not occur in practice. Miller (2012) also expresses concern that the fitted model is not compared to any others, making it difficult to say how well the model performs relative to alternatives. Chater and Oaksford (2008) have also argued that a probabilistic approach to modelling cognition sheds light on how humans deal with uncertainty. They note that there are clearly demonstrable circumstances where people behave in a way which violates these models, but suggest that this may arise, in essence, because the brain has limitations which the abstract probabilistic models do not capture, or that the brain implements them in an approximate fashion – Ríos Insua et al. (2015b) discuss level-*k* thinking, which another approach to this in the context of ARA. Chater and Oaksford (2008) contentend that this does not prevent such models from being informative, and note approaches similar to those I discuss in section 2.5 as addressing the underlying difficulty.

2.4 Heuristic Decision Making

In this section, heuristic approaches to decision making are discussed. This area is distinguished from normative decision making, and descriptive decision making, which I discuss in the following section, by a focus on rationality as existing in an ecological context.

As noted, heuristic decision making stems from a contention that Von Neumann and Morgenstern type rationality ignores the context of decision making, and that there is a lack of correspondence between predicted and actual human decisions. Arguably, this notion begins with Simon (1956), who suggested that humans do not attempt to make optimal choices, but instead simply choose the first option which is perceived to be 'good enough', a process Simon terms "satisficing". While the author notes that in most cases, the satisfactory option is in fact likely to also be the optimal one, the claim is that humans exhibit bounded rationality (Simon, 2000) arising from inherent limits to cognition.

Gigerenzer and Goldstein (1996) take the concept of bounded rationality further, and argue for what they term 'Fast and Frugal Heuristic' (FFH). This recasts rationality as bound to the context of the behaviour: a rational approach to choosing the right mate might well require checking every possible partner, but given finite time, memory, and so on, rapidly becomes unachievable. On this basis, they contend that the rationality of any given decision rule can only be determined in the context of the environment (Todd and Gigerenzer, 2003), which implies that heuristics are task-specific. They provide a number of heuristics for varying decision problems. The *recognition heuristic* (Goldstein and Gigerenzer, 2002), which suggests that given two city names, people will assume the more familiar one is larger, is one example. The *take-the-best* (TTB) *heuristic* (Gigerenzer and Goldstein, 1999), which concerns how to select a single cue for making a decision is another, and finally the central idea of single reason decision making applied in the binary case, the *priority heuristic* (Brandstätter et al., 2006), which the authors claim predicts many observed deviations from rationality. This last argues for an approach to decision making which considers the minimum gain of the two actions, the probability of that minimum gain, and the maximum gain. The heuristic considers each in order, and if either option is clearly better, chooses the superior option.

The Fast and Frugal Heuristic (FFH) approach has been criticised on several counts. Newell et al. (2003) examined the take-the-best (TTB) heuristic using a share buying experiment, where participants had the option to buy clues about the value of shares. The authors argue that despite creating an environment strongly favourable to the heuristic approach, only a third of participants' behaviour reflected the TTB approach. Oppenheimer (2003) has also examined the recognition heuristic, replicating the city size prediction task, but forcing comparisons between a known city, and a fictional counterpart. They found that where participants knew that the real city was small, they tended to predict that the fictional city was larger. This casts considerable doubt on the frugality of decision making in this case, suggesting that decision making is not based on a single cue. Further doubt has been cast on the central assumption of the approach, that the alternative to a fast and frugal approach is cognitively costly, by Bröder and Newell (2008). The authors review empirical work within the FFH paradigm, and conclude that the costs associated with decision strategies which consider more information have been overstated. Finally, Hilbig (2010) critically reviews studies examining the empirical basis of FFHs, in part from concern that previous reviews have tended to ignore, or misinterpret the evidence in an overly favourable light. Hilbig finds that the evidence is less than favourable, suggesting that the most charitable conclusion is that some heuristics explain some behaviour, sometimes. They also note that in the case of the priority heuristic, the evidence is overwhelmingly against it. Hilbig (2010) also argues that the central assumption of higher cognitive costs being associated with more complex strategies is fundamentally flawed.

2.5 Descriptive Decision Theory

This section reviews the field of descriptive decision theory, which attempts to capture human decision making, with its accompanying deviations from classical rationality, and finite power.

While heuristic theories arguably fall under the purview of the descriptive, the wider tendency is towards what are in essence "patches" to normative models. The most influential models in this class derive from Prospect Theory (Kahneman and Tversky, 1979), which combines a set of heuristics based on observed decision behaviour (Tversky and Kahneman, 1974), with distortions to the perception of probability, and the value of outcomes (Kahneman and Tversky, 1984; Tversky and Kahneman, 1986). Tversky and Kahneman (1992) subsequently addressed issues present in their original formulation by introducing Cumulative Prospect Theory (CPT), which allows for non-binary decisions, but dispenses with the heuristic aspects of the original formulation. The essence then, is that high and low probabilities are treated differently, and the subjective value of a loss differs from the equivalent gain[1] (losing your shirt is perceived as more of a loss than winning a shirt is a gain). This last effect, known as the framing effect is particularly significant, see for example work by Toll et al. (2007) examining the relationship between loss and gain framings and success rates in giving up smoking, and the National Institute for Health and Care Excellence (NICE) guidance on framing of treatment options (National Institute for Health and Care Excellence, 2007).

CPT has been successful in explaining a number of anomalous results in decision tasks (see Camerer (2000) for a review), and Thaler (2000) comments to the effect that the theory is promising, albeit incomplete, lacking for example any explanation of how frames are constructed. While remaining an effective account of decision behaviour under risk, the theory does not attempt to resolve apparent inconsistencies that arise when outcomes are delayed, i.e. in situations of intertemporal choice, where preferences may alter if outcomes occur at different times. Historically, Discounted Utility (DU) (Samuelson, 1937), which effectively claims that

[1] Figure 5.3, in Chapter four illustrates these distortions.

the value of a thing now is exponentially greater than the promise of the same thing at some future date, has been applied to explain this. More recently, Ainslie (1991) has suggested that discounting of future outcomes is hyperbolic, rather than exponential. However neither model is complete, in that both fail to account for results from Thaler (1981) showing differing temporal discounting rates for losses and gains. Loewenstein and Prelec (1992) report additional failings in classic DU models, and propose a modified form of CPT which they suggest is able to handle both immediate, and intertemporal choice.

An additional issue with both CPT, and Prospect Theory, is that they were developed and evaluated in the context of decisions from description. Such decisions are, in reality, relatively rare, and while the model has been extended by Fox and Tversky (1998) to incorporate subjective probabilities, this remains an important caveat.

Baltussen et al. (2006) also report experiments where participants behave inconsistently with CPT in mixed gambles task, with moderate probabilities. The authors suggest that this may arise from an issue with the parameterisation of the probability distortion applied by the CPT model when gambles are mixed, and probabilities do not sit at the extremes. This is not entirely controversial, in that there a variety of functional forms have been suggested both for the value, and probability weighting parts of CPT[2] (Stott, 2006), as well as considerable debate about the parametrisation of them (Booij et al., 2009). Birnbaum (2008) offers a more substantial criticism of Prospect Theory, CPT, and all models in the same line. The author presents eleven 'paradoxes' where real behaviour violates CPT, with the incidental comment that this also applies to the priority heuristic. Birnbaum suggests an alternative model, which views decisions as trees, such that multiple occurrences of the same outcome are distinct. Branches are then weighted according to the degree to which the decision maker pays attention to them, which is dependent on how risk averse they may be. Birnbaum (2005) presents a direct comparison of the model with CPT on predicting behaviour in gambling tasks, finding that it significantly outperforms CPT in predicting participants actual choices.

Another noteworthy characteristic in all the approaches reviewed here, is that they deal only with a single aspect of decision making - the actual mechanics of making the choice. The surrounding infrastructure – the multiple processes necessary to produce a decision problem from observations of the world – is taken as given. This means the models do not attempt to

[2]The functional forms used throughout this work are those of Tversky and Kahneman (1992), and are given in full in equations (3.8-3.16) in Chapter three.

explain how discrete options are arrived at, how they are valued, or other complex issues such as the role of perception and attention in decision making. While this is perhaps amongst the most minor of simplifications of reality in the context of agent-based modelling, it must nevertheless be borne in mind when considering the explanatory power of such approaches.

Chapter 3

Deciding to Disclose: A Decision Theoretic Agent Model of Pregnancy and Alcohol Misuse

An abridged version has been published as Gray et al. (2016).

Abstract

I draw together methodologies from game theory, agent based modelling, decision theory, and uncertainty analysis to explore the process of decision making in the context of pregnant women disclosing their drinking behaviour to their midwives.

I employ a game theoretic framework to define a signalling game. The game represents a scenario where pregnant women decide the extent to which they disclose their drinking behaviours to their midwives, and midwives employ the information provided to decide whether to refer their patients for costly specialist treatment. This game is then recast as two games played against "nature", to permit the use of a decision theoretic approach where both classes of agent use simple rules to decide their moves. Four decision rules are explored - a lexicographic heuristic which considers only the link between moves and payoffs, a Bayesian risk minimisation

agent that uses the same information, a more complex Bayesian risk minimiser with full access to the structure of the decision problem, and a CPT rule.

In simulation, I recreate two key qualitative trends described in the midwifery literature for all the decision models, and investigate the impact of introducing a simple form of social learning within agent groups. Finally a global sensitivity analysis using GEMs is conducted, to compare the response surfaces of the different decision rules in the game.

To aid in replication and extension, the model has been implemented as a Python module, and is freely available under the Mozilla Public License from `https://github.com/greenape/disclosure-game-module`.

3.1 Introduction

The case in favour of Agent-Based Model (ABM) as a general analytical approach has been made numerously, and elegantly (e.g. Epstein and Axtell (1994); Resnick (1994); Axelrod (1997); Gilbert (1999); Macy and Willer (2002); Silverman et al. (2011, 2013), and Epstein (2014), amongst others). As such I will not belabour the point, and instead turn to addressing some of the concerns expressed about the approach. In this instance I focus on the perception of ABM as ad hoc in nature, reflecting the assumptions of the modeller rather than empirically or theoretically grounded (Waldherr and Wijermans, 2013). To ameliorate this concern, I draw on decision theory to produce simple rule based, learning, decision making agents and show that they are able to play a form of signalling game[1] (Kreps and Cho, 1987) with a basic form of intragroup social learning. Four decision models of varying complexity, and behavioural plausibility are contrasted, by way of demonstrating the significance of the operationalisation of decision making in ABM.

This exercise is framed in the context of disclosure decisions, taking drinking patterns in pregnant women as a motivating example. Alcohol consumption in the antenatal period is a significant issue in itself, although there is not a clear consensus on associated risk. In terms of official guidance in the UK, the National Institute for Health and Care Excellence (NICE) acknowledge

[1] In a signalling game, one player (the signaller), has some piece of information that is known only to them which affects the outcome of the game for both players. The signaller has a choice as to what they tell the other player about this hidden information, and the responding player as to what they believe the information to be.

that evidence of harm to the fetus is less than conclusive, but advise not drinking at all, or significant moderation (National Institute for Health and Care Excellence, 2010b), with similar advice from the UK Department of Health (2008).

Turning more specifically to disclosure of alcohol use by women to healthcare professionals during their pregnancy, research is relatively sparse, although qualitative trends are reported by Phillips et al. (2007) and Alvik et al. (2006). The former explored factors impacting disclosure through a small case study, highlighting the need to build up rapport between woman and midwife over several appointments; the latter compared post-partum reports of consumption with contemporaneous accounts, finding apparent underreporting during pregnancy which was amplified by increased drinking. The simulation model described in this paper is able to replicate both qualitative trends, i.e. an increase in disclosure over appointments, and more honest behaviour by moderate as compared to heavier drinkers.

The resulting scenario is of substantial independent interest, and shows the potential utility of a simulation approach in domains where the process is obscured, here both because of the interest in concealment, and obvious ethical concerns. With this said, the lack of a strong quantitative evidence base against which to validate the behaviour of the model augers for caution in interpreting the results, and a necessary reminder that in this instance the model is primarily a tool for formalisation of the thought process (Epstein, 2008), rather than a machine for predicting.

A game theoretic approach to generating an abstract form of the problem gives a convenient, and well known framework to reason about the processes involved in the scenario. While scenarios may map to a number of games, exploring one candidate game still allows for a principled comparison between interpretations, and enforces explicit assumptions. But equilibrium is the *sine qua non* of game theory, which is concerned with the stable outcome of an infinite contest of second guesses. We wish to see the system in motion rather than just at rest, even if it does eventually settle to some stable point. Instead, we choose to focus on the behavioural processes driving a system in motion - a system *out* of equilibrium - to understand how these processes interact with the movement. Introducing decision theory takes a step down the ladder of abstraction from the mental chess of game theory. Dealing instead in the mechanics of decision making, and the calculus of choice, allows us to explore not only paths that arrive at the destinations we might consider in game theory, but also avenues not accessible where we constrain ourselves to a sometimes implausible degree of rationality.

This does not preclude a strategic dimension, since decision rules are to a great extent modular, and as demonstrated in this paper can be exchanged without altering the decision problem. In addition, rules are agnostic as to the source of information, suggesting room for multi-stage processes - for example, a more game theoretic, model of the opponent's mind, type approach could act as an information source for a decision rule. As a corollary, the decision problem agents attempt to answer can change, allowing behaviour in novel problems to be informed by beliefs derived under other conditions. This is also indicative of the broader benefits to ABM as an approach. Embedding these abstract rules in a simulated environment allows for mechanics which cannot be readily explored using purely analytic, or predictive approaches, for example, the social learning dynamic of the disclosure game model.

While there is no universal theory of human behaviour to sit at the centre of ABM as a method, a key motivation for decision rules is their claim to provide an account of decision making that is behaviourally and cognitively plausible. Their mooted capability in this regard is to some extent supported by work from neuroeconomics, which aims to empirically test theories of decision making (Rustichini, 2009). Many key aspects common to decision rules, for example the idea that a common currency is used by the brain to compare outcomes (Padoa-Schioppa and Assad, 2006, 2008), are supported by neurological findings. In addition, a single decision rule represents a parsimonious alternative to numerous case specific production rules.

Given these features, the application of decision, and game theory to ABM is an attractive approach to computational social science, where the locus of interest is process, and decision making. Taking a balance between models focused on replication of low level neurological mechanics, and those with a higher level emphasis where individual behaviours are abstracted away, yields a computationally tractable approach. Despite the relative simplicity, it nonetheless captures some of the nuance and sophistication of human decisions.

The remainder of this paper proceeds to review the substantive context of alcohol use and disclosure in the maternity context (section 3.2), then outlines the proposed approach to model development (section 3.3), and simulation experiments examining the ability of the model to replicate qualitative trends in disclosure (section 3.4), with selected results (section 3.5), followed by a discussion contrasting the decision models (section 3.6), and conclusions (section 3.7).

3.2 Alcohol, and Disclosure in the Maternity Setting

This section presents a brief overview of literature focusing on the impact of drinking behaviour in pregnancy, and factors affecting disclosure behaviour in the midwifery context.

3.2.1 Impact of Alcohol

Distinct from stigma attached to alcohol consumption in pregnancy, is the question of the real impact on woman and baby both in the antenatal period, and beyond. While the canonical example of alcohol linked disorders is Foetal Alcohol Syndrome (FAS), and others on that spectrum, heavy drinking during pregnancy has been mooted as a factor in a variety of negative health outcomes.

The impact of moderate alcohol consumption in pregnancy is more contested. For example, Andersen et al. (2012) examined moderate drinking in a large Danish cohort study, finding a significant increase in the risk of spontaneous abortion at low levels of consumption early in pregnancy. Savitz (2012) questioned the extent to which this can be interpreted as a causal connection, noting that there is a known relationship between absence of morning sickness, and spontaneous abortion, and suggesting that this may explain much of the difference in risk. Kesmodel et al. (2002) examined the relationship between alcohol consumption and still-birth, finding that increased consumption lead to an increase in risk to the baby, but in contrast to Andersen et al. (2012) this was significant at term.

Considering longer term negative outcomes, a meta-analysis by Latino-Martel et al. (2010) examined the potential for maternal alcohol consumption in pregnancy to feature as a risk factor for onset of childhood leukaemia, finding that any alcohol consumption was associated with an increased risk of childhood acute myeloid leukaemia (AML), but note the rarity of the condition as a limitation.

Huizink and Mulder (2006) reviewed literature looking at the impact of moderate consumption on neurodevelopmental and cognitive outcomes, concluding that maternal consumption can be a contributing factor to Attention Deficit Hyperactivity Disorder (ADHD), and impairments to learning and memory. They subsequently suggest that the underlying mechanism is not specific to alcohol consumption, but a more general phenomena arising from perturbations to foetal conditions (Huizink, 2009), but caution that methodological issues in many of the studies reviewed may undermine this hypothesis.

Contrary to this, a meta-study by Gray and Henderson (2006) found there was insufficient evidence to suggest any harm arising from moderate (under 1.5 UK units[2] per day) alcohol consumption. This ties to the current guidance from the NICE (National Institute for Health and Care Excellence, 2010b), advising that women should avoid drinking at all in at least the first three months of pregnancy, and no more than 1-2 units once or twice a week if they do. In giving this advice, NICE acknowledge that the risks to the foetus from alcohol are a somewhat contentious subject, concluding that the evidence of harm is inconclusive, but that this is not sufficient to rule out the risk of negative outcomes. This tension is reflected by earlier guidance from the Royal College of Obstetricians and Gynaecologists (RCOG) (Royal College of Obstetricians and Gynaecologists, 1996) suggesting no evidence of harm below 15 units per week, and subsequent criticism by Guerri et al. (1999), who suggest that this might be interpreted as legitimising binge drinking, while noting several studies indicating adverse affects linked to even a single drink per day (e.g. Day et al. (1990)). A subsequent RCOG statement (Fraser, 2006) revised the recommendations to incorporate newer findings, advising that there is no known safe threshold for drinking in pregnancy, and highlighting binge drinking as of particular concern.

There has recently been an increased interest in the impact of binge drinking, as a distinct pattern of consumption, with a wide variety of negative outcomes reported by Maier and West (2001), although a significant portion of their evidence base is drawn from animal studies which augers for caution in generalising findings to humans. Strandberg-larsen et al. (2008) explored links between binge drinking, and stillbirth, reporting a statistically significant increase in risk associated with more than three antenatal binge episodes. Sun et al. (2009) looked at seizure disorders in children whose mothers binged during pregnancy. They reported significantly greater risk of both neonatal seizures (ca. three-fold) and epilepsy (1.81-fold) associated with binge drinking between 11 and 16 weeks, but emphasised the exploratory nature of the results, and need for replication. In terms of neurodevelopmental outcomes, Streissguth et al. (1994) found a dose dependent association with scores on timed word, and arithmetic tests in fourteen year olds with a stronger association where bingeing occurred. A review by Henderson et al. (2007) cautiously

[2] 10ml, or 8g of pure alcohol.

supports the contention that binge drinking has a neurodevelopmental impact, but found no consistent support for adverse outcomes in pregnancy (e.g. stillbirth, miscarriage, etc.) and note a paucity of studies in the area. Meyer-Leu et al. (2011) considered the neonatal period, finding that both moderate and binge drinking were associated with an increased trend towards neonatal asphyxia. They also noted a large number of contradictory findings and raising methodological concerns about the studies reviewed by Henderson et al.. Barr et al. (2006) contend that binge drinking may also contribute to psychiatric issues in the later life of offspring, although in this case their findings are confined to individuals with FAS, which may in itself be a confounding factor, rather than indicating a directly causative relationship between antenatal binge drinking and subsequent psychiatric disorder in offspring.

Overall, there is a distinct lack of consensus on what, and how extensive, the effects of drinking on the immediate and long term health outcomes are for the child.

3.2.2 Disclosure

The issue of disclosure is central to the model presented here, in particular self-report by women of information that might disadvantage them, or be expected to do so in the immediate term. In general, the consensus is that alcohol self-reports have acceptable validity in the research context (Del Boca and Noll, 2000), but do not correspond perfectly to alternative methods. Del Boca and Darkes (2003) claim that the validity is generally accepted, and suggest that the current focus lies on what factors and processes underlie the discrepancies rather than questioning determining their existence. In this instance, the conjecture is that the information is in some way stigmatising; that, following Goffman (1990), disclosure equates to revelation of the mark. This is not immediately contentious, for example Gomberg (1988) identified stigma surrounding alcohol abusing women in particular, an issue also highlighted by Improving Access to Psychological Therapies (IAPT) guidance (Improving Access to Psychological Therapies et al., 2012), as well as a number of other studies relating response effects to perceived negative consequences (Langenbucher and Merrill, 2001; Del Boca and Noll, 2000; Blair et al., 1977). In the maternity context, Radcliffe (2011) identifies stigma pertaining to substance misusing women amongst staff, and suggests that this may represent a barrier to appropriate treatment; similarly, NICE guidance on pregnancy and complex social factors (National Institute for Health and Care Excellence, 2010c) recognises concern about the attitude of staff as a source of anxiety in pregnant women who misuse substances.

Stigma, or fear of a judgemental response on the part of the practitioner should not however be taken uncritically to explain inaccurate reporting by patients. While recent NICE public health guidance advocates routine alcohol misuse screening as a part of all practice, there is no specific policy for routine antenatal care beyond providing information on possible impacts of alcohol consumption (National Institute for Health and Care Excellence, 2010a,b). NICE guidance on pregnancy and complex social factors (National Institute for Health and Care Excellence, 2010c) does specifically address women who misuse alcohol, but presupposes knowledge of the problem through medical history, or via other services. Taken in concert with the potential for harm from even moderate alcohol use (section 3.2.1), this suggests that much of the onus is on the patient to volunteer information.

Where screening is used, Kaskutas and Graves (2000) note that the most basic method, i.e. number of standard drinks consumed, can lead to inaccurate estimates of consumption arising from inability to relate the concept of a standard drink, to actual consumption. This is compounded by the impact of memory effects on recall over a number of days (Stockwell et al., 2004), and a lack of consistency in the standard drink measure (Turner, 1990). Alternative screening tools, for example AUDIT, and T-ACE are available and have been shown to perform well in identifying problematic levels of drinking (Piccinelli and Tessari, 1997; Bradley et al., 1998; Russell, 1994; Russell and Martier, 1996), although the emphasis in these cases is on consumption at disordered levels.

Prior et al. (2003) considered a different health arena (mental health problems and general practitioners), with similar characteristics in terms of concealment of medically relevant information. The central finding in this case is that non-disclosure is not a result of stigma, but of mismatched ontologies surrounding mental illness. Work by Alvik et al. (2005), where the relationship between anonymity and reporting of alcohol consumption by pregnant women was investigated, found no significant relationship, suggesting that a fear of social judgement may not be a dominant factor. This draws an interesting contrast with a study by Alvik et al. (2006), which found that contemporaneous reports of consumption were significantly lower than those post-partum. Logistic regression results suggest that this trend is amplified by a number of factors, including level of alcohol consumption preceding conception, while anxiety about foetal well-being during pregnancy was associated with lower retrospective reports. Taken together with (Alvik et al., 2005), these results could be seen as conflicting, but may suggest self-stigmatisation (Watson et al., 2007), or reflect a lack of distinction between anonymity, and confidentiality (Malvin and Moskowitz, 1983).

In summary then, there is a consensus that alcohol consumption is generally under reported in the pregnant population, with some support for the idea that concern about social judgement associated with stigmatisation may be a contributing factor. Of particular interest in the wider context of this work, is the relationship between under reporting and consumption, i.e. that heavier drinking is associated with a greater tendency to understate intake.

3.2.3 Implications for Clinical Practice

Given that alcohol consumption is thought to be under reported some consideration must be given to the implications for midwifery practice, in terms of eliciting more accurate self-reporting. Phillips et al. (2007) present a qualitative account of factors influencing the disclosure of substance misuse to midwives, identifying particularly the need to build up a rapport, potentially over a number of appointments. This was related to continuity of care, seen as necessary by both midwives and women for building up a trust relationship, itself a key component of facilitating disclosure. Stevens and McCourt (2002) looked specifically at the process of transitioning to a caseloading model of care provision in one midwifery practice, reporting that both practitioners and women felt that this offered advantages in terms of long term relationship building. Relationship building was also highlighted by Kennedy et al. (2004) in a narrative investigation of midwifery practice, where the subjects interpreted the midwife-woman dynamic as about mutuality. Kennedy et al. suggest that this arises from a recognition that interactions in this context are about information exchange, with the knowledge base of the woman as significant as that of the midwife, rather than simply a one directional didactic relationship.

Hunter (2006) also focuses on much of midwifery as about relationship building, suggesting that there is an insufficiently recognised emotional labour component to practice. Observation and interview of a number of midwives as they practised suggested that many midwives effectively took a mother type role to their patients, with implications around the nature of information exchange that was able to take place. The emotional labour component was also reported by

Stevens and McCourt, who suggest that this is more evident under a caseloading system, particularly with challenging patients with complex needs. Todd et al. (1998) surveyed midwives working in a hospital environment, as well as those working in the community in a caseloading context, finding that community practice appeared to provide more job satisfaction, but was challenging to implement effectively because of limited resources. Community midwives suggested that larger team sizes, and smaller caseloads would contribute to a better realisation of the model. Farquhar et al. (2000) approached the same question from the perspective of women, also finding that faulty implementation hampered the expected benefits. They found that those cared for under a team scheme, with much higher continuity of care, reported that they had a better relationship with their midwives, but were not more satisfied in general with their care. In contrast, Biró et al. (2003) looked at a randomised control trial (RCT) of team midwifery care versus hospital care in Australia, finding a significantly higher level of satisfaction under the team model, the distinction in this case may lie in the different balance of team size to caseload size.

In terms of the impact of continuity of care on health, rather than experiential outcomes, research is relatively sparse. Marks et al. (2003) examined the impact of continuity of care on Postnatal Depression (PND), which has similar features to alcohol in that it carries an associated stigma that can act as a barrier to help seeking (Dennis and Chung-Lee, 2006). Based on the results of a RCT, they conclude that continuity of care is not protective, in the sense of reducing rates or impacting onset, but was very successful in supporting engagement with treatment. Echoing this, a 2009 Cochrane Review by Hatem et al. (2008) found no significant difference in incidence of PND, but reported benefits in terms of lower rates of episiotomies, anaesthesia, and shorter hospital stays, with higher satisfaction as found by Biró et al. (2003). While research in this area does not specifically pertain to disclosure, the general trend in results are suggestive when taken in concert with studies emphasising the importance of relationship building as key in fostering a disclosure friendly environment. Continuity of care is generally regarded as improving patient experience, and leading to better health outcomes in the wider medical arena (Van Walraven et al., 2010), but is clearly not a cost free endeavour, with particular concern arising from the emotional cost (Todd et al., 1998), and increased rates of 'burnout' in practitioners (Sandall, 1997).

3.3 Disclosure Game Model

In this section I sketch[3] the process of moving from a real world scenario to a minimal game which sufficiently captures reality, expressing the result as a decision problem representation, and translating this to a simulation model. I then outline four possible decision rules, and as an example of additional flexibility of process models and simulation in contrast to purely predictive or analytical approaches, extend the model to allow a simple form of social learning.

3.3.1 Modelling Approach

To model a scenario, I have taken the approach of first creating a formal game to represent it, capturing the key features as far as possible in the structure of that game. This game is in essence a conjecture about the real data generating process, which can be played out in simulation.

The appropriate game representative of the scenario of interest, which captures the desired strategic dynamics may not be immediately obvious. I would suggest that an iterative process is beneficial, beginning from the simplest possible game, and progressively augmenting it.

Transitioning from the resulting game, to a set of decision problems is a relatively simple task. I have treated the n player game as n one-player games (Ríos Insua et al., 2009), where the moves of other players are drawn from a probability distribution - nature, in game theoretic parlance. As with the game, the decision problem representation admits a degree of variation, and may need to be adjusted to reflect the decision rules that will be used.

These decision problems may then form the basis of an agent model, where agents use learning, and decision rules to play out the game. Simulation can then support features which cannot be readily represented within an analytic framework, for example, populations of heterogeneous players, individual and social learning, or network effects. In addition, the ability to observe the system in a state of flux rather than at equilibrium is desirable, since even where a social system reaches a stable state, the process by which we arrive at it is significant.

[3]A complete example of this for the alcohol misuse in pregnancy model is given in appendix A, with a schedule of simulation provided in appendix B

3.3.2 Scenario

Typically in the UK, women have 12 appointments with a midwife during the antenatal period, and in the majority of cases will encounter several different midwives (Redshaw and Henderson, 2014) in the course of their care. In the UK, and unlike most healthcare contexts, maternity notes are held by the patient, so midwives do not have extensive information prior to an appointment unless they have encountered the woman previously. Maternity notes are not generally linked to extra-departmental records, meaning that a history of alcohol related admissions to another service may remain unknown unless revealed by the woman.

According to NICE guidance (National Institute for Health and Care Excellence, 2010b,c) the issue of substance misuse should be raised at the initial booking appointment, followed by subsequent action if a concern is raised is at the discretion of the midwife. This may take the form of specific guidance to reduce intake, or if deemed necessary a referral to a specialist midwife and relevant interdisciplinary team. On alcohol consumption, policy regarding how to determine the level of consumption is at the time of writing generally at the level of local health authority, hospital trust, or according to the best judgement of the individual midwife, with no guidance provided by NICE. This commonly takes the form of average units per week, but may include Tolerance, Annoyance, Cut down, Eye-opener (T-ACE)[4] (Sokol et al., 1989) and similar measures.

Beyond the "booking" appointment, the onus is on women to raise concerns about their drinking behaviour, or the midwife to probe further if they feel it is warranted. In either case, once a concern has been raised the midwife must respond clinically, and inevitably personally, to the information.

In an ideal world, all interactions with healthcare providers would be immediately and fully disclosive, with no repercussions for the patient. However, alcohol misuse by women is known to attract stigma (Gomberg, 1988), and is a recognised barrier to appropriate treatment in the maternity context (National Institute for Health and Care Excellence, 2010c; Radcliffe, 2011).

[4]The T-ACE is a four question screening test for alcohol misuse intended specifically for use with pregnant women.

3.3.3 Disclosure Game

In order to translate the scenario sketched above into a more abstract, tractable form, I have cast it as a signalling game, and assume that women's disclosures (or not), are signals. I have also impose a discretisation on the continuum of alcohol use, and use three types of behaviour — light[5], moderate, or heavy. Correspondingly, they are limited in what signals they may send when claiming to be one of these three types.

Midwives are treated in a similar fashion, where their type corresponds to how negatively they regard a drinking pattern — non-judgemental, moderately judgemental, and harshly judgemental. The expression of this judgement is not a matter of choice on their part, and is assumed to have no impact on their clinical response, which is to either refer the woman for specialist treatment, or do nothing.

At the end of a game, each player receives a payoff dependent on the actions and types of both players. Because both women and midwives have an interest in the outcome of the pregnancy, and would prefer a healthy baby, the payoff has a common interest component. Hence, both players receive a payoff based on the outcome of pregnancy, but women bear a social cost dependent on the signal they sent and the midwife's reaction to it. Similarly, midwives pay a cost if they refer to a specialist, mirroring the organisational cost of non-routine care. Table 3.1 shows the three payoff matrices which together describe the game.

As an example, consider the challenge faced by an agent of the heavy drinking type. In order to get the best health outcome, they must be referred and would ideally achieve this without paying any social cost at all. The best move depends on the type, and beliefs of the midwife. For example, a particularly unlucky scenario might be for the midwife to not only be of a harshly judgemental disposition, but to believe that no women really need to be referred (i.e. that all women are light drinkers). Even a relatively weak belief in this possibility can make the honest signal look like an unwarranted risk.

To formally define the game, let $N = \{m, w\}$ be the set of players each with a private type $\theta_i \in \Theta$, and a set of types $\Theta = \{l, m, h\}$, with pure strategies $A_m = \{r, n\}, A_w = \{l, m, h\}$. Here,

[5] Or abstinent, the extent of alcohol consumption being such that it would generally be felt to pose essentially no risk.

Table 3.1: Payoff matrices

		Woman		
		Heavy	Moderate	Light
Midwife	Harsh	-2	-1	0
	Moderate	-1	0	0
	Nonjudgemental	0	0	0

(a) Social cost, X_s, for women, given their signal, and the midwife's type

		Woman		
		Heavy	Moderate	Light
Midwife	Refer	10	10	10
	Don't refer	-2	-1	10

(b) Health outcome, X_h, for women and midwives, given the midwife's action, and woman's type

		Woman		
		Heavy	Moderate	Light
Midwife	Refer	-9	-9	-9
	Don't refer	0	0	0

(c) Referral cost, X_c, for midwife, given their action and the woman's type

$\{l,m,h\}$ correspond to light, moderate, and heavy alcohol consumption for women, and non-judgemental, moderately judgemental, and harshly judgemental for midwives. Midwives' pure strategies $\{r,n\}$ are to refer, or do nothing, and those for women are to signal that they have one of the possible drinking patterns. Additionally define two utility functions -

$$u_w(s_w, s_m, \theta_w, \theta_m) = X_{s,s_w,\theta_m} + X_{h,\theta_w,s_m} \tag{3.1}$$

$$u_m(s_w, s_m, \theta_w) = X_{h,\theta_w,s_m} + X_{c,\theta_w,s_m}, \tag{3.2}$$

with X_c, X_h, and X_s being the payoff matrices as in table 3.1, s_w and s_m denoting a specific signal by a woman, and referral response by a midwife. Lastly let $p_w(l,m,h)$, $p_m(l,m,h)$ be distributions over types of women, and midwives respectively.

As noted, rather than solve the game, I have allowed populations of agents to play it, and hence stipulate further that women are drawn in order from a queue of n_w women (where $n_w = 1000$ in all simulations), and play against a midwife chosen at random from a population of n_m ($n_m = 100$). They play for a maximum of r_w rounds ($r_w = 12$ following the routine number of ante-natal appointments in the UK (National Institute for Health and Care Excellence, 2010b)) or until they are referred for specialist treatment, and a new player is drawn from the same distribution that produced the original players to replace them. If they are not referred, they rejoin the back of the queue after their appointment. In either case, they are informed of their payoff after each round and update their beliefs accordingly using one of the rules described in section 3.3.5.

Midwives play for r_m rounds ($r_m = 1000$ in all experiments), and conduct appointments in parallel, i.e. if there are 5 midwives, then five women are drawn from the queue and assigned at random to the midwives. Unlike women, midwives are only informed of their payoff if they choose to make a referral. Both groups of agents have perfect recall, and midwives are assumed to retrospectively update their observations if they make a referral after a number of appointments.

3.3.4 Social Learning

In reality, learning is not exclusively from personal experience, and social learning plays an important role. This social dynamic fits naturally into an agent framework, but is difficult to address without using an approach concerned with process, so I take advantage of this to show a naïve take on it here.

In the disclosure game model, this takes the form of having each agent recount their play history to their colleagues with some probability q. Individuals then incorporate shared information into their beliefs using weighted updates, e.g. for a midwife a shared observation of a low type signal contributes to their beliefs by w, and $0 \leq w \leq 1$ (i.e. $n_j = n_j + w$). Women share only when they have finished play, and provide their complete history of games, because they have accurate information about the outcomes. By the same rationale, midwives share only their history with the most recent woman they referred. Sharing occurs simultaneously for all players at the end of each round, and all memories are either shared immediately or discarded.[6] Accounts are shared with some probability, to all fellow players. For example, a heavy drinker finishes play having claimed to be a light drinker, without ever being referred, and their account is selected to be shared with some probability q_w.

Because of their differing problem representations, the simple payoff reasoners and their more complex counterparts incorporate this exogenous information differently. The simple payoff-based rule relies on a belief structure relating actions directly to rewards which is essentially model free. Because payoffs differ by the agent's private type, the information shared may not

[6]More precisely, memories of games remain, but it is assumed that only the most current information is relevant enough to be shared.

correspond to the experience of the listening agent in the same scenario. As a result, payoff reasoners have a belief bias towards the most common player type, and can believe in outcomes that are, for them, impossible.

A payoff-based agent, who is a light drinker, hears the account of the heavy drinker. They take the account as literally happening to them, and update their beliefs to include the possibility that there is a negative outcome attached to claiming to be a light drinker.

By contrast, representing the problem in terms of the probabilities of the individual lotteries imposes a model that abstracts the new information from payoffs, and allows the agent to discard implausible outcomes. This stronger assumption as to the static and known qualities of payoffs does however reduce the flexibility of the decision rule.

Returning to our example, a light drinker using this decision rule would follow the account through from their position in the game tree, correctly inferring that the outcome in their case would be positive.

3.3.5 Agent Models

While in principle a wide variety of agent models are possible, given that decision rules operate on essentially the same information, and produce the same output - a decision, I have limited myself here to four. The simplest is a lexicographic rule (1), in the spirit of a Fast and Frugal Heuristic (FFH) (Gigerenzer, 2004) which uses only information about payoffs given actions; this is followed by a Bayesian risk minimisation rule (2) using the same information; a second Bayesian risk rule (3) which uses information about the underlying lottery; and a two-stage Cumulative Prospect Theory (CPT) (Hau et al., 2008) agent (4) which is identical to 3, but uses the CPT decision rule from Tversky and Kahneman (1992). Hence, each successive decision model adds a layer of sophistication to the problem representation while retaining the same input-output characteristics.

Agents have perfect recall, midwives recognise women if they repeatedly encounter them, making use of new information for retrospective updates. However, all four agent models make decisions ‘as-if’ they were always facing a new “opponent”.

A simplifying assumption is made that all midwives have just qualified after receiving identical training. As a result, they have homogeneous beliefs about women and assume to some extent that they are honest. Women have heterogeneous beliefs, which correspond to experiencing k randomly chosen paths through the game, and following each path at least once.

3.3.5.1 Lexicographic Heuristic

The lexicographic heuristic (Luce, 1956; Fishburn, 1974) (algorithm 1) follows the form of that used in Hau et al. (2008), and assumes a simplified problem representation, where an action is a choice between combined lotteries. Functionally, the heuristic maintains a count of the number of times that each action was followed by a payoff, and chooses the action which most commonly has the best payoff, i.e. one reason decision making. Where there is no clear best action, but one or more is evidently worse, a choice must be made as to whether to discard the poorer action; in this case I have elected to retain it. This approach requires minimal computation, and does not assume that u_i is static, or known.

Women resolve this by approximating the utility function, as a function $f(s_w, \sigma)$ on their choice of signal and an unknown distribution σ, which maps to u_w – i.e. s_w is a choice between simple lotteries. The algorithm maintains a count, n, of the number of occurrences of each outcome given the choice from s_w.

Midwives solve a slightly different problem with more information, where s_w is known, and s_m is the lottery choice – $f(s_w, s_m, \sigma)$. This is resolved by maintaining a separate count for each signal (i.e. n_{s_w, s_m}), and otherwise following the same algorithm (1).

Algorithm 1 Psuedocode for the Lexicographic heuristic

$n \leftarrow 1, action \leftarrow none$
while $action = none$ **do**
  Calculate the nth most common outcome following each action.
  Sort actions by the value of the nth most common outcome.
  if clear winner **then**
    $action \leftarrow best$
  end if
  $n \leftarrow n+1$
end while
return action

3.3.5.2 Bayesian Payoff

The Bayesian payoff agent uses the same subset of information as the lexicographic method, but updates beliefs on the link between actions and payoffs using the Bayes rule, and attempts to choose the action which minimises risk.

Given the discrete nature of actions and payoffs, coupled with a desire for tractability of the simulation, the Dirichlet distribution is employed as a prior to represent these beliefs (Agresti and Hitchcock, 2005). The distribution is particularly convenient, in that to infer the probability of a signal implying a payoff is simply -

$$p(x = j|D, \alpha) = \frac{\alpha_j + n_j}{\sum_j(\alpha_j + n_j)} \tag{3.3}$$

Where n_j is simply the count of occurrences of signal-payoff pair j, and α_j is the pseudo-count of prior observations[7] for a pair j. Hence, the belief that a signal will lead to a payoff is the number of times that pairing has been observed (including the pseudo-count), over the total number of observations thus far. This makes computation of beliefs fast and simple, since all that must be maintained is a count of observations. As before, midwives follow a similar pattern but maintain n_{s_w} independent counts of pairings between referral choice and payoff, updating their beliefs about the relationship between the choice to refer and payoff given the signal they have received.

Agents then choose the strategy s_i to minimise risk R_i, which is simply defined as -

$$R_w(s_w) = \sum_{x \in X} -xp(x|s_w) \tag{3.4}$$

$$R_m(s_w, s_m) = \sum_{x \in X} -xp(x|s_w \wedge s_m), \tag{3.5}$$

where X is the set of payoffs, x, the agent has observed to follow s_i.

[7] Psuedo-counts are related to, but distinct from prior beliefs. Here, the pseudo-count is a hyperparameter of the prior Dirichlet distribution and is nothing more than a hypothetical count of prior observations. The outcome is a result of the relationship between α, and n, which is increasingly driven by the priors as for higher values of α.

3.3.5.3 Bayesian Risk Minimisation

The second Bayesian agent augments the reasoning of the simple payoff model, making the stronger assumption that the utility function is static, and known. Women maintain two sets of beliefs, corresponding respectively to p_m, and the probability of referral given signal choice. This leads to the risk function R_w, minimised with respect to s_w -

$$R_w(s_w, \theta_w) = \sum_{i \in A_m} \sum_{j \in \Theta} -u_w(s_w, i, \theta_w, j) p(j) p(i|s_w), \tag{3.6}$$

so that the risk of a signal is the sum of the products of all payoffs with the probabilities of their entailed midwife types and responses.

Midwives reasoning centres on determining the meaning of signals, since given the knowledge of what some signal s_w conveys about the true type of the sender, the payoff for an action is known. As such, their inference process is the same as for the simple Bayesian agent but over signal-type pairs, and they attempt to minimise the following risk function R_m, minimised with respect to s_m -

$$R_m(s_w, s_m) = \sum_{i \in \Theta} -u_m(s_w, s_m, i) p(i|s_w) \tag{3.7}$$

3.3.5.4 Descriptive Decision Theory

The most complex decision rule used is CPT, which attempts to reproduce a number of systematic deviations from rationality observed in humans. Rather than risk, ‘prospects’, the sequence of payoff-probability pairings in ascending order of payoff, associated with an action are used as decision criteria. While CPT has primarily been applied in the context of decisions from description, it has been modified to deal with decisions from experience by incorporating a first stage where probabilities are estimates from observations Fox and Tversky (1998). In this instance the Bayesian inference process fills the first stage role.

CPT uses transformed probabilities, under weighting small probabilities, and over weighting large ones. This is intended to reflect the observed behaviour of humans, where sufficiently high likelihoods are treated as certain, and in contrast, low probabilities as impossible. The correct weighting function is subject to some debate, but here I have used that of Tversky and Kahneman (1992), which treats probabilities differently for gains (3.8) and losses (3.9):

$$w^{+}(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{\frac{1}{\gamma}}} \tag{3.8}$$

$$w^{-}(p) = \frac{p^{\delta}}{(p^{\delta} + (1-p)^{\delta})^{\frac{1}{\delta}}}, \tag{3.9}$$

where p is the unweighted probability, and γ and δ are the weights for gain and loss probabilities respectively. Along similar lines, the values of losses and gains are transformed to reflect a tendency to regard a loss as more significant than a gain, such that -

$$v(u_i) = \begin{cases} f(u_i), & \text{if } u_i > 0 \\ 0, & \text{if } u_i = 0 \\ \lambda g(u_i), & \text{if } u_i < 0 \end{cases}, \tag{3.10}$$

where

$$f(u_i) = \begin{cases} u_i^{\alpha}, & \text{if } \alpha > 0 \\ \ln(u_i), & \text{if } \alpha = 0 \\ 1-(1+u_i)^{\alpha}, & \text{if } \alpha < 0 \end{cases} \tag{3.11}$$

$$g(u_i) = \begin{cases} -(-u_i)^{\beta}, & \text{if } \beta > 0 \\ -\ln(-u_i), & \text{if } \beta = 0 \\ (1-u_i)^{\beta}-1, & \text{if } \beta < 0 \end{cases}, \tag{3.12}$$

and α and β are respectively the power of a gain and a loss, and λ is a multiplier giving the aversion to loss.

Finally, the transformed probabilities are used to construct decision weights, π^+, π^- for losses and gains, where

$$\pi_n^+ = w^+(p_n) \tag{3.13}$$

$$\pi_{-m}^- = w^-(p_{-m}) \tag{3.14}$$

$$\pi_i^+ = w^+(p_i + \ldots + p_n) - w^+(p_{i+1} + \ldots + p_n), 0 \leq i \leq n-1 \tag{3.15}$$

$$\pi_i^- = w^-(p_{-m} + \ldots + p_i) - w^-(p_{-m} + \ldots + p_{i-1}), 1-m \leq i \leq 0 \tag{3.16}$$

The CPT value of a single outcome prospect $f = (u_i; p_i)$, is $v(u_i)\pi^+(p_i)$ if $u_i \geq 0$, and $v(u_i)\pi^-(p_i)$ otherwise. For any given action the CPT value, V is the sum of the value of the prospects of that action, as in the Bayesian risk model, and the agent chooses the option which maximises this quantity.

3.4 Method

This section provides details of simulation experiments conducted to examine the ability of the model to reproduce qualitative trends reported in the midwifery literature by Alvik et al. (2006), and Phillips et al. (2007); as well as a global sensitivity analysis and construction of statistical emulators to explore, and contrast the response surfaces of the four decision rules.

3.4.1 Qualitative Trends

Throughout this paper, parameters for the CPT model were as used in Tversky and Kahneman (1992) (table 3.2). While there has been significant work on determining appropriate parameterisation for the model (e.g. Neilson and Stowe (2002); Nilsson et al. (2011); Glöckner and

Pachur (2012), and particularly Byrnes et al. (1999) and Booij et al. (2009) addressing risk aversion and gender), a full exploration of the impact of these parameters, or heterogeneous values within populations is beyond the scope of this work. For simplicity, it was assumed that all three drinking types are equally prevalent within the population, although results derived from the Avon Longitudinal Study of Parents and Children suggest that the reality is far more positive[8] (Humphriss et al., 2013). The scenario was biased towards disclosure as the better option by presuming a distribution of midwives strongly skewed towards non-judgemental types, with beliefs initially favouring honesty. Payoffs were assumed as in table 3.1, which ensure that it is always strictly preferable to refer drinkers, and together with the initial belief that signals will be honest, not refer those claiming otherwise.

Table 3.2: Model parameters, $\alpha, \beta, \gamma, \delta$ and λ from (Tversky and Kahneman, 1992, pp. 311–312).

Name	Description	Value
n_w	Number of women	1000
n_m	Number of midwives	100
r_m	Number of appointments per midwife	1000
r_w	Maximum number of appointments per woman	12
Runs	Simulation runs	1000
$p_w(h)$	Proportion of heavy drinkers	$1/3$
$p_w(m)$	Proportion of moderate drinkers	$1/3$
$p_w(l)$	Proportion of light drinkers	$1/3$
$p_m(h)$	Proportion of harsh midwives	$5/100$
$p_m(m)$	Proportion of moderate midwives	$10/100$
$p_m(l)$	Proportion of non-judgemental midwives	$85/100$
q_w	Probability of women sharing	0.
w_w	Weight of shared information for women	0.
q_m	Probability of midwives sharing	0.
w_m	Weight of shared information for midwives	0.
$s_i[a_i] : s_i[a_{\neg i}]$	Pseudo-count favouring honesty[9]	10:1
γ	Probability weighting for gains	0.61
δ	Probability weighting for losses	0.69
α	Power for gains	0.88
β	Power for losses	0.88
λ	Loss aversion	2.25

Two key measures were used: the fraction of the subpopulation of women who had ever signalled honestly, and the proportion of the population who were referred for specialist treatment. Both

[8] In this study 95.5% of women in the sample reported consumption at, or below, NICE recommended safe levels.

[9] This is the ratio by which the Dirichlet hyperparameters favour assuming a signal is accurate. In the case of the Lexicographic rule, this equates to setting the starting count of payoffs following an action to ten, where $s_w = s_m$.

measures were taken after every round of play, and were taken relative to the agent's position in their sequence of appointments giving the probability of signalling honestly, or being referred having had a given number of appointments.

In addition to assessing the adequacy of the rules in capturing qualitative trends, I also examined the impact of simple social learning within the population of women (section 3.3.4) on the robustness of these trends. The original experiment was repeated for different values of q_w $\{0.25, 0.5, 0.75, 1 | q_w = w_w\}$, with 100 runs under each condition.

3.4.2 Global Sensitivity Analysis

In general, I have followed the example of Bijak et al. (2013) for global sensitivity analysis of stochastic agent-based models, although see Thiele et al. (2014) for a review of alternative techniques. For this purpose the Gaussian Emulation Machines for Sensitivity Analysis (GEM-SA) software (Kennedy, 2004) was used, which implements the Bayesian Analysis of Computer Code Outputs (BACCO) method developed by Oakley and O'Hagan 2002; 2004. This is a form of variance-based sensitivity analysis, which assumes that the model output is an unknown, smooth function of the inputs. The unknown function can then be approximated as a Gaussian Process, which is fitted to the training data using Bayes' Theorem and then serves as an emulator for the simulator. The meta-model is then able to provide an indication about the extent to which uncertainty in a parameter propagates to uncertainty in the output, and how sharply the output responds to change in each parameter.

Parameters for training were generated in R (R Core Team, 2014) using and appropriately transformed Latin Hypercube Sample (Carnell, 2012) over the space of parameters given in table 3.3, giving eleven free parameters which were treated as uniformly distributed in the range given. Given the limitation of 400 design points for the GEM-SA software, I produced exactly that many parameter combinations and collected results for 100 runs of each, with emulator quality assessed by leave-one-out cross validation. A fixed set of 100 random seeds was used[10], such that each parameter set was run once with each seed, for every decision rule.

[10]Fixed random seeds were used to allow simulation results to be reproducible, since the combination of a parameter set and a random seed yields a deterministic process.

Table 3.3: Parameter ranges.

Name	Description	Min	Max
$p_w(m)$	Proportion of moderate drinkers	0	1
$p_w(l)$	Proportion of light drinkers	0	1
$p_m(m)$	Proportion of moderate midwives	0	1
$p_m(l)$	Proportion of non-judgemental midwives	0	1
q_w	Probability of women sharing	0	1
w_w	Weight of shared information for women	0	1
q_m	Probability of midwives sharing	0	1
w_m	Weight of shared information for midwives	0	1
x_h	Health payoff for healthy delivery	1	100
x_r	Cost for referral	$-(x_h - 1)$	
$s_i[a_i] : s_i[a_{\neg i}]$	Pseudo-count favouring honesty	1:1	100:1

To capture the response characteristics for the model, I measured four outcome variables - (1) the interquartile range (IQR) of the average signal sent by each type of agent in a run, (2) the average signal of moderate drinking agents in a run, and (3, 4) the IQR of 1 & 2 between simulation runs. Together these four metrics give an indication of how far women are separable by their signalling behaviour (1), the behaviour of the at risk drinking groups[11] (2), and finally the variability of the system in response to changes to the parameters (3 & 4).

Measurements were taken at the end of 1000 rounds of play, and emulators were built against 400 sample points from the full set of simulation results (1 & 2), and IQR at each point (3 & 4) to assess both the overall trend, and the extent to which the parameters contribute to variance between runs.

Sixteen emulators were built, covering each of the four outputs on all decision models and used to conduct a probabilistic sensitivity analysis to assess the impact of parameters and interactions.

In addition to the sensitivity analysis, I also employed the resulting emulators to rapidly[12] explore the parameter space. While emulated results are subject to inaccuracy, they do provide an indication of the interest, and plausibility, of regions of the parameter space. Results for the outcomes of the interactions of $s_i[a_i] : s_i[a_{\neg i}]$ with x_h, and q_w with w_w are given in section 3.5.3.

[11]Under most conditions, the behaviour of heavy drinkers tracks closely with their moderate counterparts.

[12]Once constructed, the emulator has an analytical solution conditional on the roughness parameters (Oakley and O'Hagan, 2002), which obviates the need to use MCMC (Oakley and O'Hagan, 2004).

3.5 Results

3.5.1 Qualitative Trends

As shown in figure 3.2, all four decision rules were able to reproduce both qualitative trends towards more disclosure as women experience more appointments (Phillips et al., 2007), and a greater tendency towards under reporting of consumption by heavier drinkers (Alvik et al., 2006). Trends for all four rules are broadly similar, exhibiting a gradual increase across appointments which subsequently levels off. This levelling can in part be explained by the referral results shown in figure 3.1, which show that the majority of drinkers are referred, even with substantial concealment.

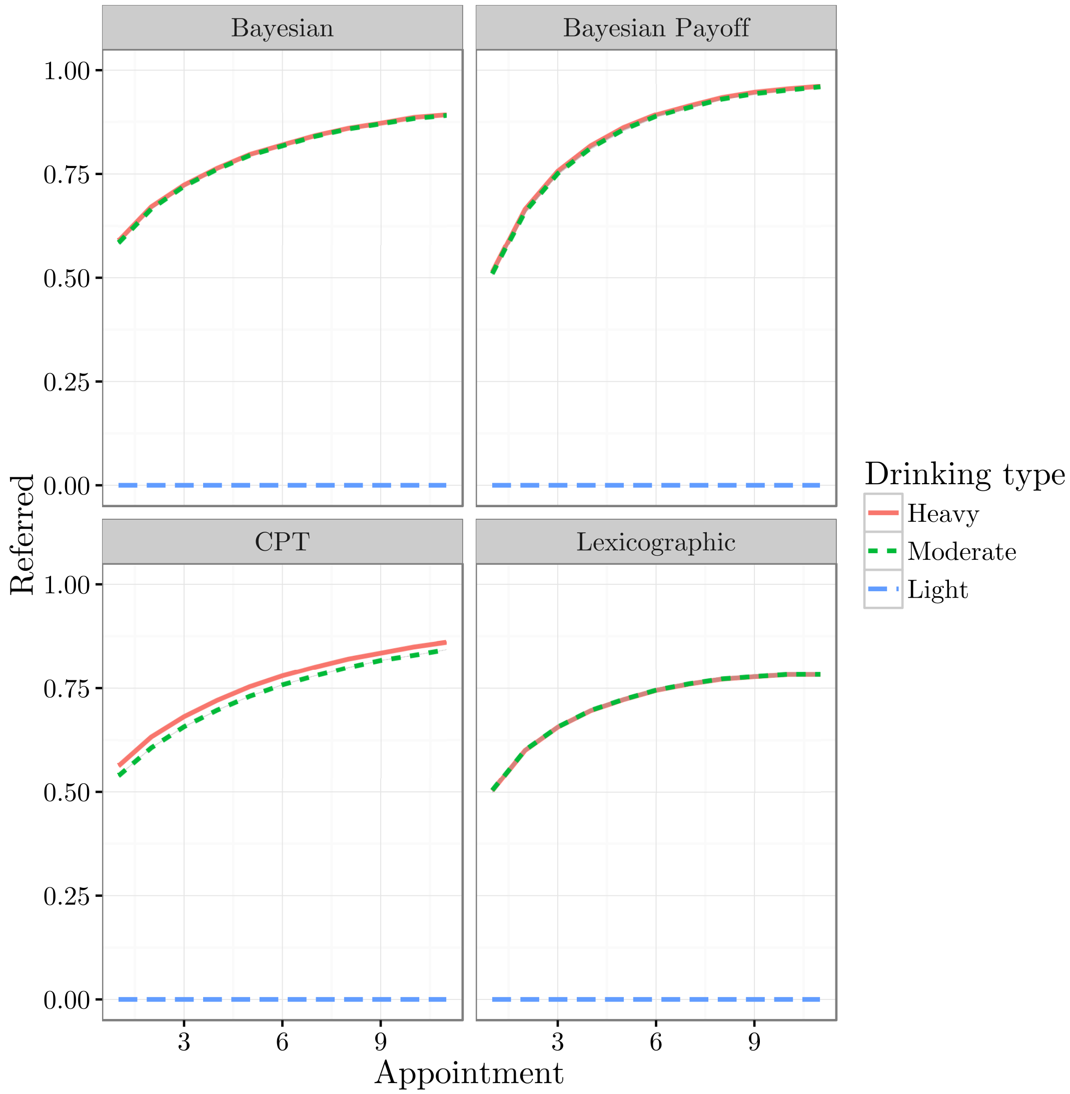


Figure 3.1: Average fraction of population referred by each appointment, after 1000 rounds, mean with 95% confidence limit over 1000 runs. Note that the large number of runs leads to very tight confidence intervals.

Referrals continue to occur, in the absence of honest signals, because drinkers are able to achieve a referral by masquerading as higher or lower types, dependent on how their initial beliefs are biased. Despite this the results suggest that a minority of risky drinkers will evade detection altogether, with no notable distinction between heavy and moderate types. Under these parameters, light drinkers always signal honestly and are never referred since there is no perceived advantage in doing so, and the evidence of deceptive signalling is insufficient to outweigh the biased priors of the midwives.

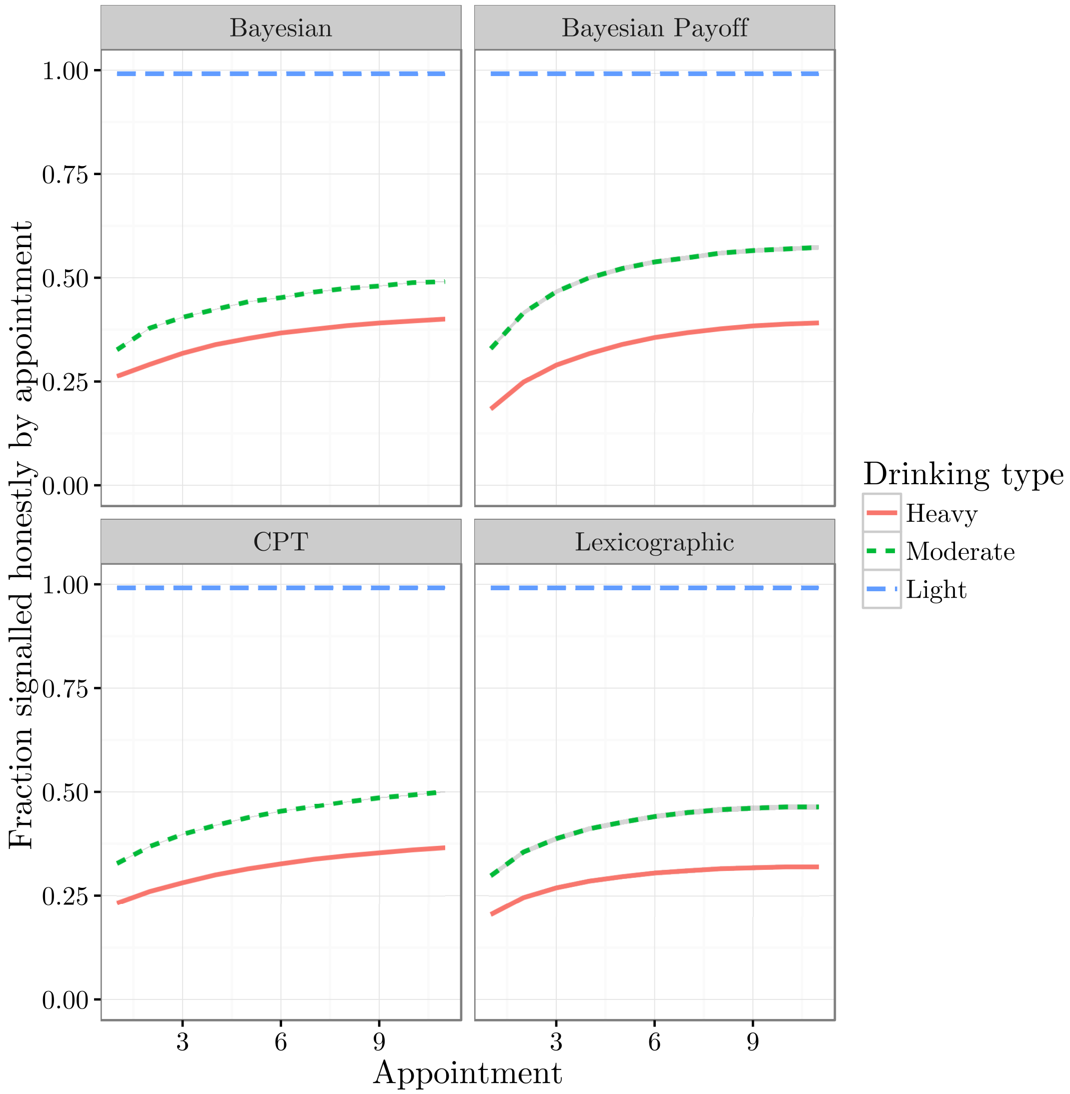


Figure 3.2: Average fraction of population ever signalled honestly by each appointment, after 1000 rounds, mean with 95% confidence limit over 1000 runs. Note that the large number of runs leads to very tight confidence intervals.

3.5.2 Social Learning

Introducing social learning amongst women leads the behaviour of the decision rules to diverge markedly, possible reasons for which I explore in section 3.6. Figure 3.3 shows the proportion of women who have signalled honestly at least once by their final appointment, under four sharing conditions.

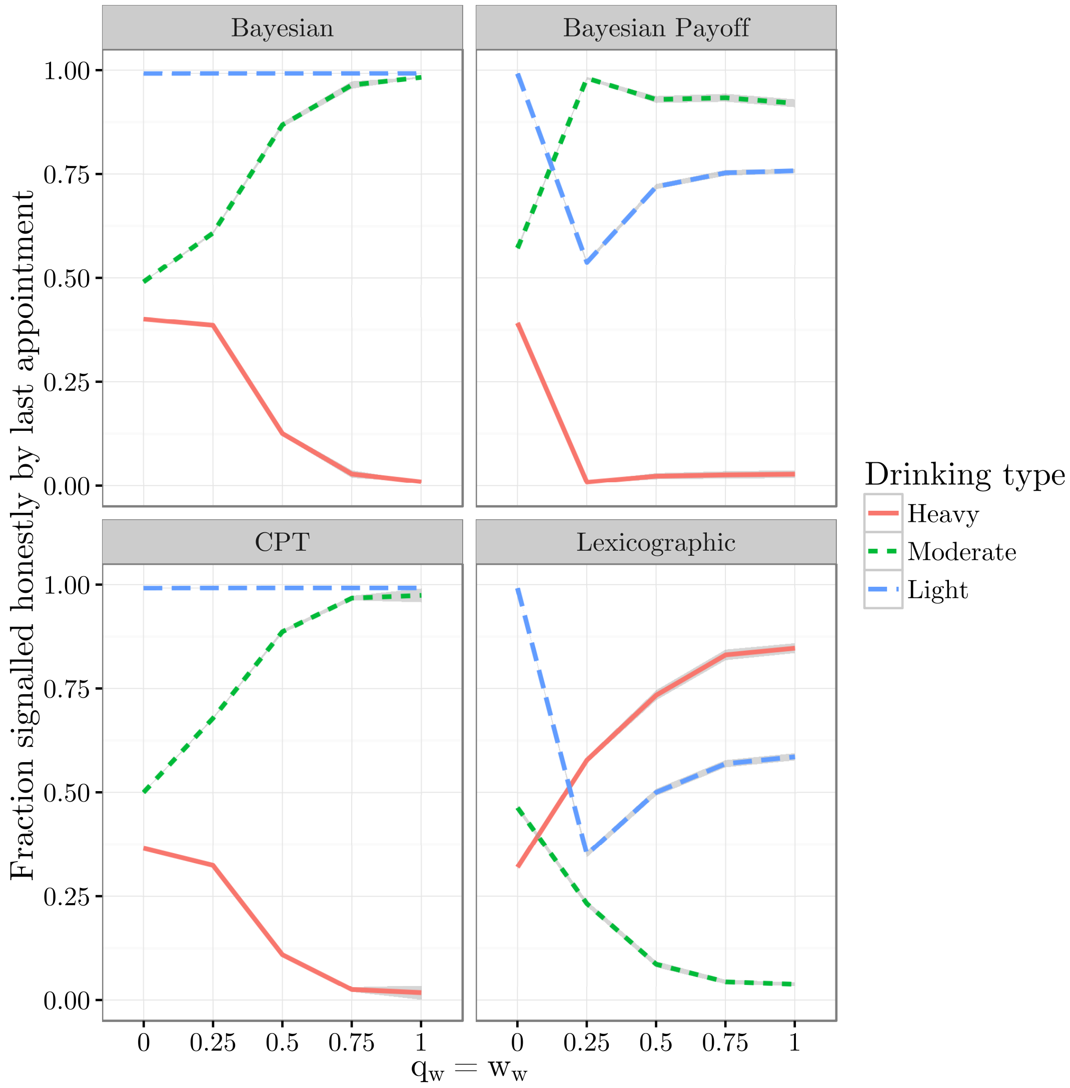


Figure 3.3: Impact of social learning on trends in the average fraction of population ever signalled honestly by their final appointment, after 1000 rounds, mean with 95% confidence interval over 100 runs

Aside from the lexicographic decision rule, the general tendency is towards less honest signalling by heavy drinkers, which is accompanied by a slight increase in referrals for the Bayesian, and CPT rules. For these decision models, this is because social learning exacerbates the existing tendency of heavy drinkers to impersonate moderate drinkers, who behave more honestly as heavy drinkers become less so. This arises because both classes of agent learn that the moderate signal is the lower risk option as it is both a reliable indicator of need, and does not attract strongly negative judgement. The reliability of the signal is self-reinforcing, since the more the agents use it and get referred, the more confident midwives become that it indicates need.

Particularly notable, is the decline in honest signalling by light drinkers visible in both heuristic type rules at the 0.25 level of q_w and w_w, which is associated with an increase in false positives. This arises because of the lack of homophily in social learning, as light drinkers become informed about negative outcomes associated with concealment, despite having nothing to conceal. The relatively high referral rates of drinkers heighten the effect further, because shared information becomes dominated by their experiences.

The relationship is not, however, entirely straightforward, in that increasing social learning leads to greater variance between runs. As a preliminary investigation of this, linear model was used to predict the between-runs interquartile range of the average signal sent by moderate drinkers. The predictors used were decision rule, and level of social learning, together with the interaction between the two. The regression results were significant ($F_{7,12} = 25, p < 2.9 \times 10^{-6}$) with $R^2 =$ 94%, and intercept 0.07. The only significant coefficients were for the interaction terms, which were 0.44 ($p < 0.05$) for the Bayesian payoff rule, and 0.69 ($p < 0.005$) for the lexicographic. This suggests that social learning, for the heuristic style decision rules introduces considerable uncertainty to the model, which is explored further in the sensitivity analysis below.

3.5.3 Sensitivity Analysis

In this section I present a brief overview of the sensitivity analysis, followed by selected results highlighting the global effect of changes to perceived payoffs and degree of bias towards honesty, as well as social learning within women. The full results for the sensitivity analysis covering all sixteen emulators are available in appendix D.

For the median signal choice of moderate drinkers, the results of the sensitivity analysis suggest that the proportion of light drinkers has a significant effect for all decision rules, accounting for 10%, 38%, 24%, and 5% of the variance in output for the Lexicographic, Bayesian Payoff, Bayesian, and CPT rules respectively. For the Lexicographic rule, the overwhelming majority of variance in signalling behaviour is reflective of the prevalence of stigmatisation by midwives (44% $p_m(m)$, 7% $p_l(m)$, and a further 15% for their interaction). The proportions of midwives are also key drivers in group separation, and the between run IQR of both measures for this rule.

Perhaps surprisingly, variance attributable to social learning between midwives is relatively low, with neither the weight nor probability accounting for more than 5% of variance in any measure. While there are small contributions to variance in interaction with other parameters (e.g. 4% to between groups IQR, for the interaction with the proportion of light drinkers under the Bayesian rule), this may suggest that the model is lacking in this area, which I touch on in section 3.7.

Figure 3.4 gives a qualitative picture of both emulator quality, and the divergent response surfaces of the decision rules in response to variations in social learning parameters. The figure shows the median signal by moderate drinkers versus the median IQR of the signal by all the drinking types. A higher IQR implies more separation in signalling behaviour between the groups, which in principle makes it less challenging to decide how to respond to a signal. The right hand side of the plot shows results from simulation for the four decision rules, at a variety of sharing parameters. The left hand side shows the output predicted by the emulator. Emulator fit is clearly imperfect, but overall behaviour is qualitatively similar, with both emulated and simulated plots demonstrating separation in outcome space for the decision rules.

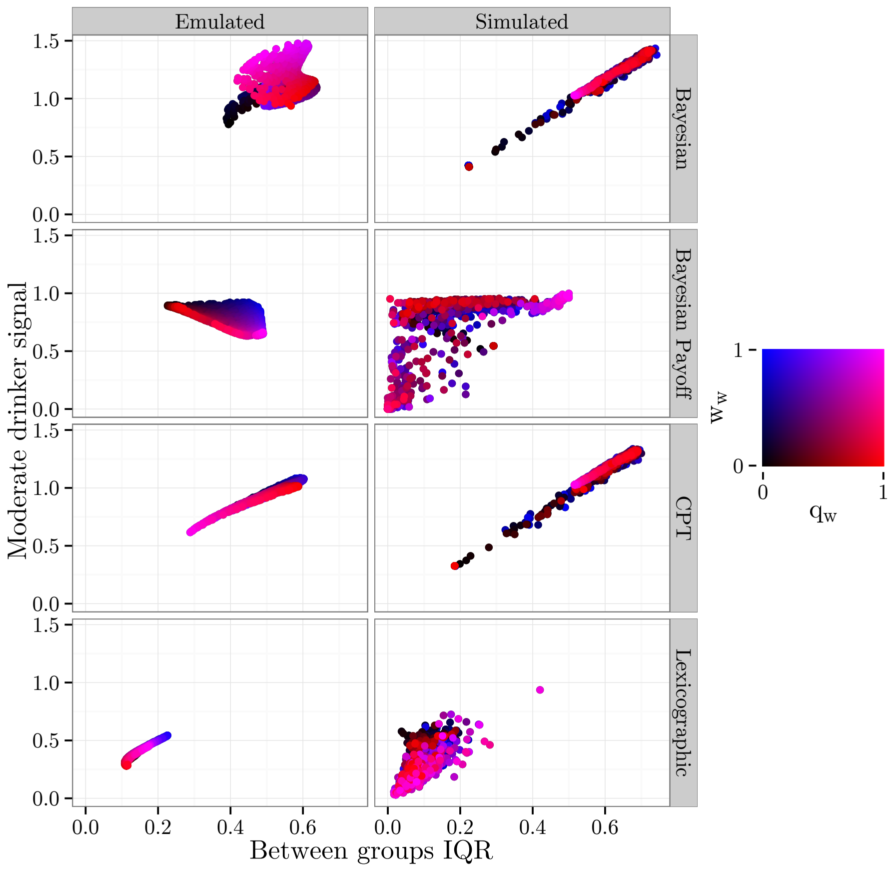


Figure 3.4: Median moderate drinker signal vs median between drinking type IQR for all decision rules, with signals coded as 0 = light, 1 = moderate, and 2 = heavy.

Following from the suggestive results for social learning introducing uncertainty (section 3.5.2), figure 3.5 shows emulated points covering the parameter space in high resolution. These plots reflect the increase in uncertainty of outcome shown for the heuristic type rules, which is especially severe for the Bayesian payoff rule. They also suggest that the Bayesian decision rule is less stable under conditions where the weight of shared information is substantially higher than the probability of sharing. This indicates that placing a high weight on information from limited sources leads to greater variability, i.e. what information is shared, matters.

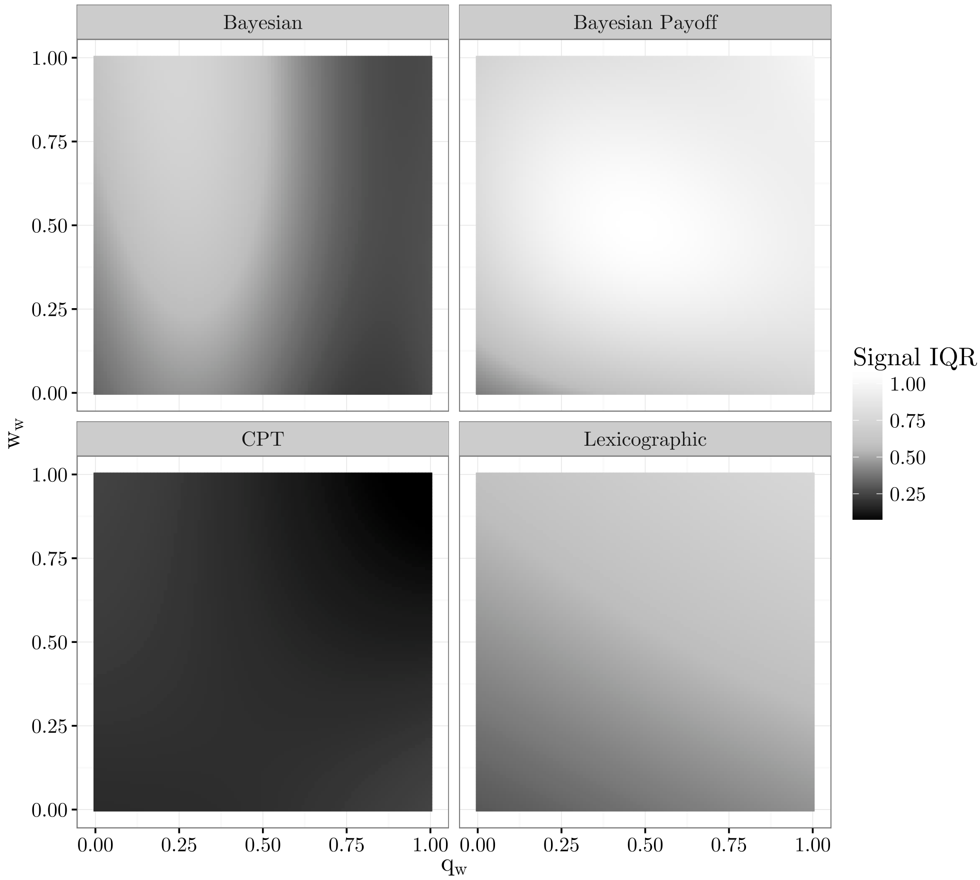


Figure 3.5: Emulated moderate drinker signal IQR in response to varying q_w, and w_w

For the CPT, and Bayesian decision models, the interaction of bias towards honesty, and distinction between payoffs has a significant, and non-linear effect on instability, and separability of groups. Figure 3.6 shows the effects, and also highlights the tendency towards poor separability of groups for both the heuristic type decision rules. The response surface of the Bayesian payoff rule is slightly more nuanced than the simple Lexicographic rule. Figure 3.6 shows better separation, close to partial pooling[13] at high payoff distinction, with relatively modest honesty

[13] Pooling occurs when signallers of different types 'pool' their signals, and one adopts the signals of another.

bias, which is reflected by the variance contributions of 11% and 8% respectively. For the more complex rules, the general tendency is towards less pooling for higher values of both, but with pockets where full pooling[14] occurs. The plots also suggest that the sensitivity of the CPT rule is marginally greater, which is supported by the significant contribution to variance of close to 15% for all measures of x_h.

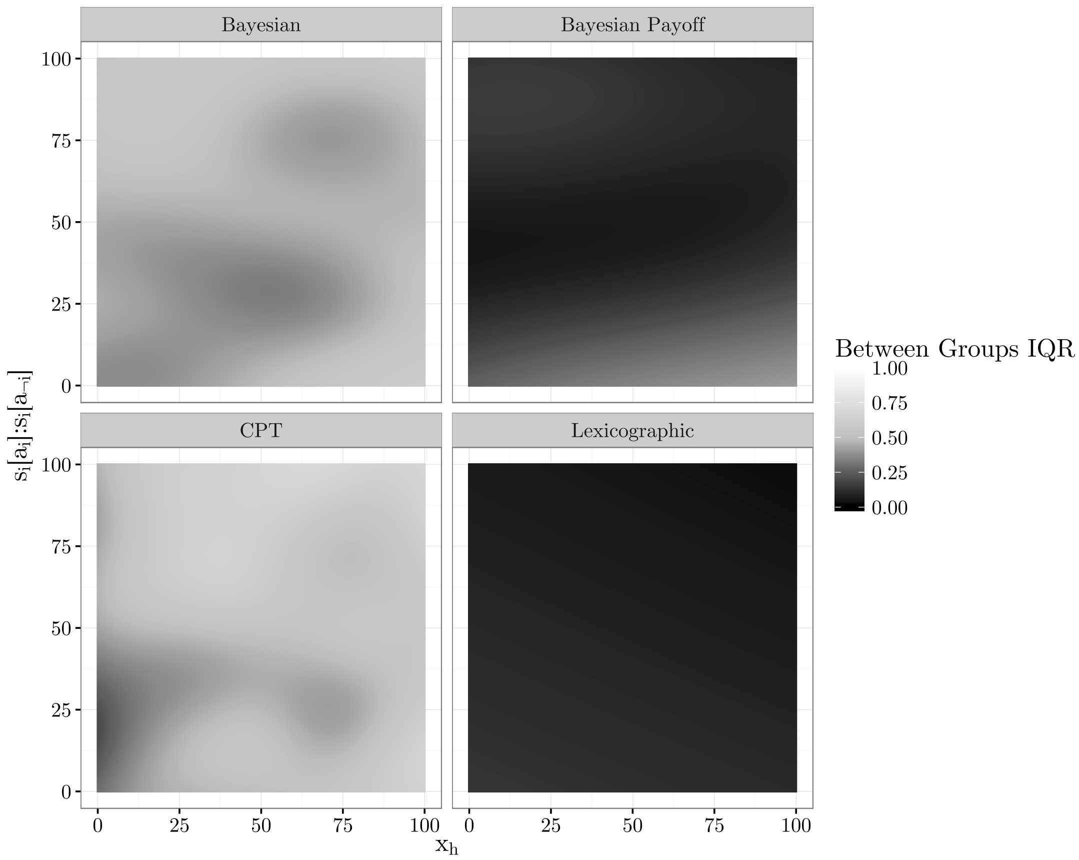


Figure 3.6: Emulated between groups IQR in response to varying $s_i[a_i] : s_i[a_{\neg i}]$, and x_h

For the CPT, and Bayesian decision models, the interaction of bias towards honesty, and distinction between payoffs has a significant, and non-linear effect on instability, and separability of groups. Figure 3.6 shows the effects, and also highlights the tendency towards poor separability of groups for both the heuristic type decision rules. The response surface of the Bayesian payoff rule is slightly more nuanced than the simple Lexicographic rule. Figure 3.6 shows better separation, close to partial pooling[15] at high payoff distinction, with relatively modest honesty bias, which is reflected by the variance contributions of 11% and 8% respectively. For the more

[14] Indicating that all signaller types are using a the same signal.
[15] Pooling occurs when signallers of different types 'pool' their signals, and one adopts the signals of another.

complex rules, the general tendency is towards less pooling for higher values of both, but with pockets where full pooling[16] occurs. The plots also suggest that the sensitivity of the CPT rule is marginally greater, which is supported by the significant contribution to variance of close to 15% for all measures of x_h.

3.6 Discussion

From a pragmatic perspective, the differing response characteristics of the classes of decision rules are substantial and significant, particularly when social learning is considered. There is a high level of uncertainty in the overall dynamics with the Lexicographic and Bayesian Payoff rules. This does not arise with the more complex rules, because they reframe information from others in the context of their own experiences, as what would happen to them in that situation. By contrast, the simpler rules treat the experiences as equivalent to their own, and since there is no mechanism of homophily, no way to listen only to accounts of agents similar to themselves, they can come to believe unreasonable things. Naturally, incorporating homophily, by, for example weighting shared information by the type of the sharer, would represent a trivial modification to the heuristic models. While to some extent this highlights the flexibility of the decision rule approach, it would of course sacrifice the parsimony of the model. This is an important consideration, given that part of the argument in favour of a decision theoretic approach lies in the minimal nature of the behavioural rules.

One of the notable features of the results is that the behaviour of rules within a class is very similar. To some extent this reflects poorly on the most complex rule, CPT, which diverges only minimally in behaviour from the Bayesian model. This might be to a degree anticipated since I have not elicited payoffs for obvious practical and ethical reasons, and they may be unrealistic, which limits the utility of the CPT approach. Additionally, work by Glöckner and Pachur (2012) has shown that there is considerable variation in individual parameters for the decision model, whereas I have let them remain homogeneous here. In the same vein, utility functions should arguably vary between individual agents, which could potentially be addressed by replacing the fixed payoffs used here with a distribution. With this said, the significant increase in complexity, which entails both additional parameters and increases to simulation time may auger for a middle

[16] Indicating that all signaller types are using a the same signal.

ground, particularly where elicitation of payoffs is impractical. This, together with the variability associated with the heuristic type decision rules speaks to a trade off between capturing reality, and replicating it.

Continuing the discussion of the issues raised by the representation of payoffs, the temporal aspect is significant, in that there is a timing difference in payoffs, since while the potential social pain of disclosure is immediate, the health outcome comes only later. In light of this, that there is a known impact of time on perceived utility Thaler (1981) suggests that incorporating some form of temporal discounting (e.g. exponential (Samuelson, 1937), or hyperbolic (Ainslie, 1991)), or a decision model which explicitly treats intertemporal choice, such as the CPT-like model of Loewenstein and Prelec (1992), is warranted.

As noted in section 3.4.2, the impact of social learning in midwives is surprisingly minimal, where it might be expected to play a more significant role in reality. A possible explanation for this lies in the implementation, which may place an excessive constraint on how much information midwives can share. The restriction to sharing only after a referral, together with the disparity in population sizes, and random allocation of appointments leads to midwives rarely having more than a single interaction with woman to pass on to their colleagues. Furthermore, because midwives are only informed of the true type if a referral occurs, they have an inherent myopia since until they have evidence of deception they will not refer, with said evidence difficult to obtain without a referral.

In reality it might be anticipated that midwives would not withhold judgement, and would pass on concerns about specific women to their colleagues, or that particularly dramatic stories would persist and be passed. This might be addressed by incorporating noisy type information (Feltovich et al., 2002), capturing the unintentional information transmitted during appointments, together with a relaxation of the assumptions about when information may be shared and more sophisticated model of information flow in general. This also highlights an advantage of the BACCO approach (which I describe in section 3.4.2), in diagnosing issues with model design by giving insight into parameters which are contributing inappropriately to variance in output. Coupled with the ability of emulators to rapidly explore parameter space, this clearly suggests that statistical emulation is a powerful tool to support simulation based approaches. As noted in section 3.5.3 the emulators here are indicative, but not definitive. Amongst the reasons discrepancy arises here are heteroskedasticity associated with social learning, the stochastic nature of the simulation, and a lack of precision given the large parameter range. The former issues could be addressed by a more comprehensive approach, which explicitly incorporates point variance, to setting the 'nugget' term, which weakens the presumed correlation of points very close together in parameter space. The latter could be improved through iterative fitting procedures, where the simulation is sampled more heavily in plausible regions of parameter space, a procedure not possible here given the dearth of data to evaluate plausibility. That the discrepancy exists is not prohibitive in this instance, since I am not using the emulator for prediction, only to achieve a broad strokes picture of the behaviour of the simulator.

3.7 Conclusion

The conclusions that can be drawn about the behaviours of real life women, and their midwives, are necessarily limited by the paucity of data available to validate the model. While qualitative trends offer some indication, they are limited in scope, and do not permit strong claims about the drivers of disclosure. As such, further work will focus on applying the model to domains where validation data is more available, which will support a more comprehensive evaluation of the model discrepancy. With this said, the trends reported by Alvik et al. (2006), and Phillips et al. (2007) are borne out by the model, and predictions from the two more complex rules suggest that encouraging information sharing between women may encourage disclosure, but at the expense of reducing accuracy. By contrast, if one takes the view that a Lexicographic model is a better approximation of real behaviour, then outcomes can best be influenced by controlling how far midwives punish their women socially. I would however suggest that there are better reasons than the outputs of a simulation for doing so.

More broadly, the results demonstrate the logistical feasibility, and its utility as a 'tool for thinking', of an agent model grounded in decision theory. The results also make clear that deciding the operationalisation of the decision making is of key significance. In the following chapter, I address the significant weakness of this model by applying it to older adult care, where more empirical data is available to assess plausibility, before addressing the equivalent lack of empirical grounding in the decision models, in Chapter five.

Chapter 4

The Risky Business of Asking For Help: An ABM of Unmet Need in Older Adults

A paper to be submitted to *Demography*.

Abstract

Decision making is at the core of many demographic processes, but is typically neglected in the analysis. I present an agent-based model with empirically grounded decision making, and use a substantive example of unmet need in older adults to show that such an approach is not only practical, but necessary.

I use data drawn from both large scale social surveys, and local authority expenditure reports to parameterise and calibrate four agent-based models. The four models utilise different models of decision making, of varying complexity. I find that only two of the models, both of which encode a representation of the decision problem, can be successfully calibrated.

To aid in replication and extension, the model has been implemented as a Python module, and is

freely available under the Mozilla Public License from https://github.com/greenape/risky-aging-model.

4.1 Introduction

Decision making is at the root of demography. The characteristics of populations are in part a result of the choices made by the individuals within them, and those of their antecedents. These decisions are made in a social context, and they influence the context they respond to. This means that not only is there a micro-macro link, but that it is a reflexive one. The implication then, is that to better understand how society behaves at the largest scales, we should consider how it operates at the smallest, and how each feeds change in the other.

This paper considers the provision of support for older adults, as an example of case where considering individual behaviour can improve our understanding of a macro phenomenon. That the existing need for social care is not being met is widely acknowledged, but gaining insight into why there is unmet need, and the processes that lead to this scenario is more challenging. A top-down approach to the problem can tell us the extent of unmet need, and identify the demographics of those not catered for (Vlachantoni et al., 2015). But this alone is insufficient without insight into the processes that drive individual experience of care, and in itself cannot fully answer why support is not forthcoming. Demographic characteristics may be the determinants of receipt of care, but in themselves are not explanations. Data and forecasts may, as suggested by Burch (2003b), constitute a body of theory, but an incomplete one.

In contrast, a bottom-up, process-oriented approach is hollow without a strong link to data. Such a link must exist not only from model outputs to reality, to validate the model, but in the parametrisation and mechanism of the model. If we construct a model that reproduces some phenomena that we have observed in reality, it remains meaningless unless the information that it operates on to do so can be derived from the empirical domain. A model that gives the right answer from the wrong information is as faulty as one which fails with the right information. Perhaps the more pernicious concern is one most readily associated with bottom up, simulation approaches. Like the Ptolemaic astronomers, we may comfortably construct an arbitrarily complex process that gives the right answers, but yields no true insight, or worse, misleads us.

The inescapable conclusion is that all aspects must be confronted with data, and informed by it. Theory and modelling alone are not enough.

I argue that both approaches are necessary, but neither is sufficient in isolation. Combining data-driven, empirically motivated process models with top-down statistical and forecasting methods can lead to an alloy stronger than either part. This melding strengthens both approaches, facilitating multilevel validation of models and enriching their construction, while guiding data collection and fortifying description and prediction with explanation.

This is not without substantial challenges; after all, these processes are driven by people, and there are no fundamental laws of human behaviour. But these are challenges demography is uniquely well equipped to address: the immense demand for multiscale data, both quantitative and qualitative, necessary to anchor hypothetical process to actual behaviour is the warp and weft of the discipline. And these challenges must be met: society is a restless, moving system, and to understand it we need to know not only where it is, and where it goes, but how and why.

In this chapter I expand on work by Gray et al. (2016) shown in the previous chapter, and present an agent-based model of unmet need in older adults, which draws on psychologically plausible, and empirically rooted models of decision making. The model design is informed by qualitative accounts of barriers to help-seeking, agents are animated by synthetic psychologies drawn from opinion surveys, and the overall dynamics of the model are validated against results from the English Longitudinal Survey of Ageing (ELSA).

I begin with the data that informs the design of the model, provides the calibration criteria, and where possible gives parametrisation of it. I then give the general form of the model, then, following George Box (Box, 1979), I give four realisations of it, all of which are wrong, but two of which I hope are useful. Next, as in the previous chapter I use statistical emulators to show that simple heuristic agents are unable to discriminate between those who do and do not require support. I then demonstrate that if agents employ a mental model of the process, the simulation is able to reproduce observed rates of receiving help. Finally, I reflect on the implications for representation of decision making, and explore which possible policy interventions could yield more desirable outcomes.

Table 4.1: Need for, and receipt of, support with washing and dressing by over 65s, in the ELSA (Marmot et al., 2015)

Year	Wave	Need support with (%)			Receive support with (%)		
		Washing	Dressing	Either/Both	Washing	Dressing	Either/Both
2006/7	3	8	8	25	41	33	44
2008/9	4	6	10	25	39	34	48
2010/11	5	5	9	22	40	29	45
2012/13	6	4	8	20	37	30	45

4.2 Unmet Need for Care

Quantitative information about the extent of the unmet need for care can be found in the ELSA. Vlachantoni et al. (2011) looked at need for support for two Activities of Daily Living (ADLs) (washing, and dressing) by over 65s, in the fourth (2008) wave of the survey. Their analysis revealed that while most surveyed (70.5%) felt they needed support of some kind, only a minority (25%) of older adults believed they needed assistance with washing and/or dressing. Of those who did require support with one or both of these activities less than half (47.5%) received help with them. I have also replicated their analysis for survey waves three, five, and six, which as shown in table 4.1 yielded similar results for the surrounding years. I have used the results of Vlachantoni et al. (2011)'s analysis as the basis for the population of agents in the simulations who may need assistance, with only a minority in need of support. This also forms the validation criteria of 47.5% of those in need receiving help.

To derive a representative value for the cost of providing support, H, I took the mean expenditure by English local authorities on providing one year of in-home care to an older adult, for the year 2008/9. This was drawn from Personal Social Services Expenditure reported by HSCIC (The NHS Information Centre Social Care Statistics, 2010), for 150 local authorities, and showed a mean expenditure of £7,881 (sd. £2,032).

Qualitative research, in turn suggests a partial explanation for not receiving help, is a failure to seek it. The reluctance of men to seek help is an oft mentioned phenomenon (Addis and Mahalik (2003) review this in several contexts), and Smith et al. (2007) found that this was a barrier to help-seeking in older men.

More generally, concern about reinforcing negative stereotypes about older adults as dependent ("stereotype threat") was identified as reducing help-seeking behaviour by Wakefield et al.

(2013). Although, since stereotype threat has also been found to affect memory in older adults (Barber and Mather, 2014), and more generally to reduce cognitive performance (Derks et al., 2008) this may represent a manifestation of the latter. A similar theme of a need to preserve the integrity of identity (the essence of stigma as discussed by Goffman (1990)), was reported by Day and Hitchings (2011), who found interviewees were reluctant to make use of assistive objects, for example blankets and hot water bottles, which were seen as symbolic of old age. These qualitative findings suggest that there is an unknown cost to seeking support, which I have incorporated into the model as 'stigma' (C).

Another key theme was identified by Walters et al. (2001), where interviewees attributed not seeking help to feelings of resignation, and low expectations about support being forthcoming, or making a positive difference if it were to be so. This highlights the need for a robust synthetic psychology founded in real data, which is as important as the synthetic population. To achieve this, I have drawn heavily on opinion surveys to generate realistic distributions of viewpoints for the population. The 2008 European Social Survey (ESS) (Norwegian Social Science Data Services, 2008) asked a number of questions specifically targeting perceptions of older adults, and I have used these results to fit a distribution of beliefs about the likelihood of being poorly treated for requesting support. The ESS used eight questions of the form 'most people in my country view those over 70...'[1], all of which were Likert type items. I restricted my analysis to responses by those aged over 65, and living in England.

These results have previously been examined using averaging by Hess and Dikken (2010), and I have followed their approach of reversing the polarity for questions on envy, contempt, and pity. However, given the relatively low internal consistency of the measures (Cronbach's alpha 0.65), and the desire for a standardized measure of perceived stigma, I have taken an Item Response Theory (IRT) approach and fitted a Graded Response Model (GRM) (Samejima, 1970) to the responses. The GRM gives a value for the latent trait the questions are intended to measure, without assuming that all questions are equally informative. Conversely, the value of the trait can then be used to predict how a respondent would answer the questions on which the GRM has been fitted.

Fitting a GRM to the ESS responses yields a distribution of traits similar to, but not identical

[1]Variables v70frnd ('most people in my country view those over 70 as friendly'), v70comp ('...as competent'), v70mrst ('...as having high moral standards'), v70resp ('...with respect'), v70envy ('...with envy'), v70pity ('...with pity'), v70resp ('...with respect'), and v70adm ('...with admiration').

with an averaged scale ($r = 0.76$). To generate a population with realistic beliefs about age related stigma, I then drew their prior beliefs about how likely others are to stigmatise them from a logistic distribution fitted to the resulting GRM ($\mu = 0.50, b = 0.063$).

In addition I have also used data from EuroBarometer 67.2 (European Commission, 2012), which asked participants about their expectation of receiving all the care they required when older[2]. In this case I have used the multinomial distribution of responses to appropriately bias agent's beliefs about how likely it is that they will be helped. Of those aged 65 and over, and living in England when asked "In the future do you think that you would be provided with the appropriate help and long-term care if you were to need it?" 26% strongly agreed, 49% agreed, 17% disagreed, and 7% strongly disagreed. [3]

Finally, I returned to the 2008 ESS and made use of three questions aimed at measuring social trust[4], to derive a distribution from which the prior beliefs of deciding agents were drawn. In this case Cronbach's alpha for the three questions was a more acceptable 0.75, indicating acceptable internal consistency, and fitting a GRM yielded scores which correlated closely ($r = 0.98$) with taking averages. As such, I have used a GRM for consistency purposes, and prior beliefs were drawn from a normal distribution ($\mu = 0.00, \sigma = 0.91$).

In this section, I have shown the qualitative basis for the game at the centre of the model, and provided a partial parametrisation of it. The next section describes the model which arises from the data.

4.3 A Model of Unmet Need

The model begins by thinking of help provision as a game for two players, belonging to separate populations. One player can choose whether to ask for help, and the other can choose whether to give it. Both players have the option of doing nothing, and both partially share in the game outcome.

[2]Variable v184.

[3]From 2009-2011 the Office of National Statistics (ONS) Opinions and Lifestyle Survey (OPN) also asked respondents about whether they expected to get all the help they needed, but the polytomous format of the EuroBarometer survey is preferable here.

[4]Variables ppltrst ('people are generally trustworthy'), pplfair ('...fair'), pplhp ('...helpful')

The asker may or may not need support, and initially this is known only to them, as in a signalling game (Spence, 1973). Their challenge is to weigh the cost of continuing in their current state, against what they perceive to be the risk of asking for help. The risks are drawn from qualitative reports (section 4.2) on barriers to help seeking, being partly concerned as to whether the deciding player will actually help, and partly a more general social risk which, following Goffman (1990), I have loosely termed stigma. Stigma in this sense might be a self applied damage to the ego, or could equally derive from the decider reacting negatively even if they do help. I have treated this as a property of the deciding agent, which the asker can observe only after encountering them[5].

It falls to the deciding player to infer whether the asker actually requires support, and to decide whether to incur the cost of providing it. Sharing part of the outcome aligns the players' motivations, since it is in the interests of both to provide the right level of assistance. Figure 4.1 shows the game tree for a single round of this game, where the paths outward from the centre are the possible sequences of moves within a game. Note that the game is structurally similar to that presented in Chapter three, with a reduced number of player types and actions and, as noted, in the simulations which follow only a single (benevolent) decider type. As in Chapter three, a complete schedule of simulation is provided in appendix E.

[5]In fact, in the simulations here there are no deciding agents with this characteristic, and the risk of stigmatisation is only perceived.

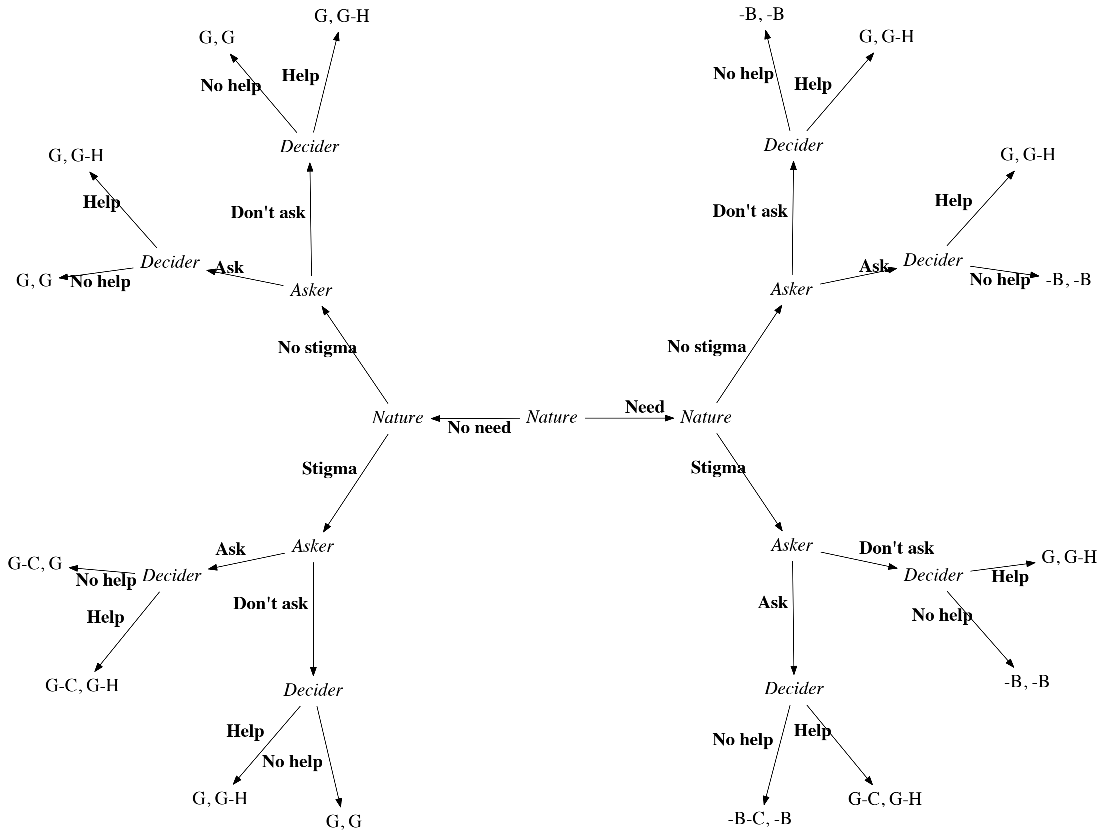


Figure 4.1: Game tree for a single round of play, showing all possible move sequences for the two players and their resulting payoffs. The tree also shows the possible moves by nature, which determine the characteristics of the players.

The general rules of play are that an asker plays twelve rounds of the game (they meet a decider once a month), and that after being given help they cease play. After playing a round, the asker is fully informed of the outcome and observes the decider's hidden type. Conversely, deciders are informed of the outcome only if they give support, but learn about the outcome of all their games with that player.

Individual games take place within a society, which consists of populations of askers, and deciders. At each round, every decider meets with one asker, and the two play the game together. If, after playing, the asker has played their allotted twelve rounds, or has been offered support, they then leave the game, and a new asker is added to the population to replace them. In addition to learning from their own experiences, both askers and deciders can also learn socially. Askers may, with some probability, report the games they have played when they finish, and deciders can give an account of the most recent case where they helped. This happens once, at the end of a round, and anything not shared is then forgotten.

Table 4.2: Outcome matrix, X. Following convention, the first and second entries in a cell are outcomes for the asker and decider respectively

			Asker			
	θ_x		h		n	
		s_x	$\top$	$\bot$	$\top$	$\bot$
Decider	m	$\bot$	$-B,-B$	$-B,-B$	G,G	G,G
		$\top$	$G,G-H$	$G,G-H$	$G,G-H$	$G,G-H$
	z	$\bot$	$-C-B,-B$	$-B,-B$	$G-C,G$	G,G
		$\top$	$G-C,G-H$	$G,G-H$	$G-C,G-H$	$G,G-H$

It is also convenient to give some notation for the game in terms of player types, outcomes, and moves[6]. Hence, define a set of players consisting of askers, and deciders as $N = \{a,d\}$, where each player has a private type $\theta_{x\in N} \in \Theta_{x\in N}$. Player types for askers shall be $\Theta_a = \{h,n\}$ where h indicates need, and equivalently $\Theta_d = \{z,m\}$ for deciders where a player with type z is stigmatising. Players have actions $A = \{\top,\bot\}$, corresponding to asking/giving help ($\top$), or not ($\bot$). The outcome of a game shall be denoted $S = (s_a, s_d, \theta_a, \theta_d)$, with s_a and s_d the moves chosen by the asker and decider, and θ_a, θ_d their private types.

The set of outcomes for askers is defined as $L_a = \{-B, G-C, G, -C-B\}$, and $L_d = \{-B, G, G-H\}$ for deciders, and with the full outcome matrix, X, indexed by S as given in table 4.2. If an asker in need of help does not receive it, both players pay a cost, B. If an asker does not need, or receives help, both players receive a benefit, G. The cost of stigma to an asker is C, and the cost to a decider of giving help is H.

Finally, I define utility functions u_a, u_d for askers and deciders as in equation (4.1). Taking an example game, where the asker is in need of help, and the decider stigmatises, if the asker does ask for help, and the decider supplies it then the outcome S is $(\top,\top,h,z)$, and hence the outcomes are $u_a(\top,\top,h,z) = X_{h,h,h,z,1} = G-C$, and $u_d(\top,\top,h,z) = X_{\top,\top,h,z,2} = G-H$.

$$
\begin{aligned}
u_a(s_a, s_d, \theta_a, \theta_d) &= X_{\theta_d,\theta_a,s_d,s_a,1} \\
u_d(s_a, s_d, \theta_a, \theta_d) &= X_{\theta_d,\theta_a,s_d,s_a,2}
\end{aligned}
\tag{4.1}
$$

Each agent also has a set of traits, which inform their basic attitude to play. Askers have traits governing their expectations of stigma, and of receiving help. Deciders have only one trait:

[6]In signalling games moves by the asker are also known as signals, and those of the decider as responses.

trustfulness. When an agent is generated, their traits are drawn from a distribution, previously derived in section 4.2.

Translating from the distributions of attitudes derived in section 4.2 to the behaviour and learning of agents warrants further explanation. For all four decision rules in this paper, the prior beliefs of an individual agent have three qualities: valence, relative magnitude of a specific belief, and the weight given to new information. For example an agent might believe they are twice as likely to be stigmatised as not, but be very open to contradictory evidence. For the Graded Response Models, the sign of the latent trait is interpreted as giving valence, and the value of it as relative magnitude. I have also introduced two parameters governing the weight of prior information, (W_s for askers, and W_r for deciders), because determining how entrenched the beliefs are would require longitudinal individual data.

As in Chapter three, all four decision rules use prior beliefs parametrised by pseudo-counts (denoted α), which in this case correspond to a hypothetical count of previous experiences playing the game. This is framed as $n(W+1)$ hypothetical experiences occurring in a $n:1$ ratio. As an example, consider an asker with a stigma trait of -4. This is interpreted as their having had $4(W_s+1)$ hypothetical encounters, with the negative sign indicating that 4 were experienced as stigmatising.

A related difficulty arises with the multinomial distribution over expectations of receiving help, which I resolve in a similar fashion by operationalising {strongly yes, yes, no, strongly no} as {3, 2, -2, -3}. Hence, if the asker answered ‘strongly yes’, they have received help $3W_s$ times, and not received it W_s times.

Having defined a game, the next step is to play it. In the following section, I describe four decision rules of escalating complexity, and show how each can realise the model.

4.4 Models of Decision Making

In this section I describe three approaches to modelling decision making – heuristic, normative, and descriptive, and four rules which epitomise those approaches. As in Chapter three, these are the lexicographic heuristic, a simple payoff-based Bayesian rule, a more complex Bayesian rule, and a Cumulative Prospect Theory (CPT) model.

4.4.1 Heuristic Decision Making

Perhaps the simplest possible approach to modelling decision making behaviour, is to take the approach of Todd et al. (2005), and assume that individuals use a simple rule of thumb - a heuristic - to make choices. Heuristics have the advantage of being fast, and from a computational perspective very cheap. In fact, it has been suggested that if we take an ecological perspective, these characteristics would make heuristic decision making rational (Gigerenzer and Goldstein, 1996).

The simplest heuristic that will suit my purposes in this environment is the lexicographic heuristic (Luce, 1956; Fishburn, 1974) introduced in Chapter three in algorithm (1), which boils down to choosing the option with the better typical outcome, i.e. of the most likely outcomes for each option, choose the most appealing. This is not guaranteed to lead to choosing the best option, but often does so, and compensates with tractability and robustness (Gigerenzer, 2008). Learning, in this context consists of keeping a running count of outcomes observed, incrementing as necessary. Social learning is implemented very simply through the same process, by multiplying the increment to the appropriate count by Q, which is the weight ascribed to outside information.

Note that there are two implications which follow from this. First, if two outcomes have the same value, agents are not able to differentiate between them. Second, askers can make decisions based on outcomes which they are not eligible for, based on social information.

As an example, consider two agents: Alice, and Bob, who will take the roles of decider and asker, respectively. Both Alice, and Bob have a uniform prior over the outcomes, as if they had played each of the paths through the game available to them exactly once. I will also assume that the outcomes for the decider are related thus $G > G - H > -B$, and similarly for the asker $G > G - C > -B > -B - C$.

Alice first plays a round with an agent who claims to not need any support. Alice must choose randomly here because to her knowledge, all outcomes are equally likely. She is aware of one outcome for giving help ($G - H$), and two possible outcomes for not giving help (G, and $-B$).

This means that there are two ways to compare her choices, depending on how the outcomes for not giving help are ordered. If she orders them $\{-B, G\}$, then helping is the better option. Conversely, if she compares G first, doing nothing is preferable. The result is that there is an even chance that she will offer help without being asked.

On this occasion, Alice chooses to provide help, and the other player turns out to not need it, $S = (\perp, \top, n, m)$, and Alice receives a benefit of $G - H$, which does not change her decision making, since there is only one possible outcome from giving help. Alice now plays another round of the game, with a new counterpart. This time, the other player does ask for help, and as before, Alice chooses her action randomly[7], and does not help. This time, the other player really did need help, hence $S = (\top, \perp, h, m)$, so both players pay a cost of $-B$, and Alice now considers this the most likely outcome of not helping someone who asks for help.

Bob needs help to get dressed in the morning, and whether or not he will ask Alice for help when they meet is uncertain. The uncertainty arises because there is no clear order of precedence for Bob to use to compare outcomes, since all are equally likely. As a result, he picks his comparison points at random, and will ask for help with probability $\frac{5}{16}$. To understand why this is so, consider that Bob is aware of four possible outcomes if he asks for help, and two if he does not. Because all the outcomes are, as far as Bob is aware, equally likely, there are 24 ways he can order the four outcomes of asking for help, and two in which he can order the two outcomes of not asking. This means there are 48 equally probable ways he can compare his actions. In 15 of those comparisons, asking for help is the better option, hence the probability that he does so is $\frac{15}{48}$, or $\frac{5}{16}$.

If we assume that when Bob meets Alice, he does not ask, then whether Alice helps is once more a matter of chance. If she uses G as the comparison, then she will not help Bob, but for $-B$ she will. In this example, Bob is lucky and Alice decides that given the most common outcome of not helping is worse than for giving support, she will help him.

Bob now believes that if he does not ask for help, the most likely outcome is that he will get it, and leaves the game. Alice's count of $G - H$ when giving help despite not being asked grows by one, but she effectively learns nothing, since $G - H$ is the only possible outcome whenever she

[7] While Alice has learned from playing one game, she has gained no new information about what outcomes follow from her actions when the other player does ask for help.

gives help. As a result, the next time she encounters the same situation, her decision will again depend on how she orders the outcomes for giving help.

4.4.2 A Normative and Cardinal Decision Model

The most obvious concern about the lexicographic heuristic model is that it only considers ordinality, and uses very little information in reaching a decision. Introducing the concept of risk, in the form of Bayesian risk, resolves both of these concerns by considering cardinality and using all available information to decide. Rather than choose the option with the best typical outcome, the Bayesian risk rule chooses the action with the lowest risk.

Learning now takes the form of Bayesian inference, and uses the same information as the lexicographic heuristic; namely counts of observed outcomes. The additional step required is the calculation of the posterior probability of each outcome. Because actions and outcomes are discrete, I have used a Dirichlet prior, which yields the general form of the posterior probability function given in equation (3.3). Agents maintain several distributions, with one α and n for each signal in the case of askers, and each combination of a signal and response for the deciders.

Askers evaluate the risk of taking an action (R_a), as the weighted sum of the observed outcomes (V) of it, as in equation (4.2), which is the equivalent of (3.4) in Chapter 3.

$$R_a(s_a) = \sum_{x \in V} -xp(x|s_a) \tag{4.2}$$

Deciders consider the risk of a response, given the action of the decider (R_d). As with askers, this takes the form of the weighted sum of outcomes which they have observed to follow a response to an asker's action (equation 4.3), the equivalent of (3.5) in Chapter 3.

$$R_d(s_a, s_d) = \sum_{x \in V} -xp(x|s_a \wedge s_d), \tag{4.3}$$

To illustrate, let us consider how this new rule would alter the encounter between Alice and Bob. At this point, it will be helpful to assign values to B, G, C, and H, since unlike the heuristic model, cardinality matters. For this and other examples then, let $G = 2$, and B, C, H all equal to 1. Alice has, as before played two rounds of the game, and Bob still chooses not to ask for help. However, this is now a certainty, rather than dependent on how he orders the outcomes he uses for comparison. As Bob is still uninformed, in the sense that he has seen one of every possible combination of player, signal, and response, he has pseudo-counts as in (4.4), where $V_\top$, and $V_\perp$ denote the subsets of the outcome set L_a Bob has observed to follow asking, and not asking for help[8].

$$\begin{aligned} V_\perp = \{-1, 2\}, V_\top = \{-2, -1, 1, 2\} \\ \alpha_{\perp,-1} = 4, \alpha_{\perp,2} = 4 \\ \alpha_{\top,-2} = 2, \alpha_{\top,-1} = 2, \alpha_{\top,1} = 2, \alpha_{\top,2} = 2 \end{aligned} \tag{4.4}$$

When Bob calculates R_a for his two possible actions he believes equally in all possible outcomes, i.e. the posterior probability for the two outcomes of action $\perp$ is $\frac{1}{2}$, and $\frac{1}{4}$ for the four outcomes of $\top$. Since the outcomes for asking for help are symmetrical, and equiprobable, they cancel each other out, and $R_a(\top) = 0$, while $R_a(\perp) = -(2(\frac{1}{2})) + -(-\frac{1}{2}) = -1 + \frac{1}{2} = -\frac{1}{2}$. Because the Bayesian risk rule uses all available information to make a decision there is no ambiguity in this case, and Bob must not ask for help.

Alice's behaviour also changes, and she will now always give help. To see why this is the case, consider Alice's prior beliefs. Alice is initially uninformed, in the sense that she has seen one of every possible combination of player, signal, and response, which leads to pseudo-counts for her possible responses to not being asked for help as in equation (4.5), where as before $V_\top$ is set of those outcomes she has observed to follow giving help, and $V_\perp$ where not assisting.

$$\begin{aligned} V_\top = \{1\}, V_\perp = \{-1, 2\} \\ \alpha_{\perp,\top,1} = 2, \alpha_{\perp,\perp,-1} = 1, \alpha_{\perp,\perp,2} = 1 \end{aligned} \tag{4.5}$$

When Alice receives Bob's signal of $\perp$, she evaluates R_d for both her possible responses, $\perp$ and $\top$. Since $R_d(\perp,\perp) = -(-1p(-1|\perp\wedge\perp)) + -(2p(2|\perp\wedge\perp)) = -0.5 > R_d(\perp,\top) = -(1p(1|\perp\wedge$

[8] Note that while in this example $V_\top$ and $V_\perp$ are identical to the corresponding sets of outcomes possible for Bob following these actions, this need not hold in general.

$\top)) = -1$, she must conclude that helping is the better choice. After observing that this was indeed the right choice, Alice updates her beliefs by letting $n_{\perp,\top,1} = n_{\perp,\top,1} + 1$ which does not alter her decision making since there are no other outcomes for giving help.

4.4.3 Bayesian Model with Mental Representation

Introducing cardinality allows an agent to use more information to make decisions, but the lack of a structural representation of the problem players are attempting to resolve is rather limiting. Adding a mental model of the problem will allow agents to make more effective use of information. This also has an introspective appeal if we consider our own decision making process.

To transition from reasoning in terms of a direct link from action to consequence is not an overwhelmingly complex task, because there is already a well-defined model of the problem. For the askers, the decision making can be decomposed into two sets of beliefs, about whether the other player will be stigmatising, and if they will receive help if they ask.

The risk function for askers is modified as in Chapter three (equation 3.6), giving equation (4.6). Naturally, this alters the learning process somewhat, because agents must maintain an additional belief structure corresponding to the mixture of stigmatising vs. accepting deciders in the population. This increase in complexity is offset by the entailed simplification of beliefs on outcomes of actions, to possible responses to actions, because in this game there are only two responses a decider can give.

$$R_a(s_a, \theta_a) = \sum_{i \in A_d} \sum_{j \in \Theta} -u_a(s_a, i, \theta_a, j) p(j) p(i|s_a), \tag{4.6}$$

As in Chapter three, deciding agents can be treated in the same way. Their problem becomes to determine the true need status of the asker, given the action they have chosen. Their risk function is ostensibly similar to the action-consequence Bayesian model (section 4.4.2), but evaluated over the link between signal, and ground truth. This then becomes equation (4.7), which parallels equation (3.7) in Chapter three.

$$R_d(s_a, s_d, \theta_d) = \sum_{i \in \Theta} -u_d(s_a, s_d, i, \theta_d) p(i|s_a) \tag{4.7}$$

In both cases, this replaces inference about outcomes, with inference about actions, and the use of a mental model permits players to use more information. Askers are also aware of their own need state, i.e. if they are on the left or right branch of the game tree (figure 4.1). As a result, rather than taking the reported outcome of another player as identical for them, it instead informs them on probable paths from their current state. A side effect of this which I do not explore in this paper is that the separation of reward from inference opens the scope for heterogeneous utility functions.

Incorporating the mental representation does not change Bob's behaviour, and he still never asks for help. However Alice's behaviour does alter, and whether or not she helps is again a 50-50 chance. If she does choose to help, she is certain to assist without being asked in the next round.

The reason this behaviour differs from the action-outcome decision rule lies in the representation of beliefs. Equation (4.8) shows Alice's α values for a signal of $\perp$, with the relevant observations after her initial two games.

$$\begin{aligned} &\alpha_{n,\perp} = 1, \alpha_{n,\top} = 1 \\ &n_{n,\perp} = 1, n_{n,\top} = 0 \end{aligned} \tag{4.8}$$

When Alice receives Bob's signal, her belief that he needs help is $p(h|\perp) = \frac{\alpha_{\perp,h} + n_{\perp,h}}{\sum_j (\alpha_{\perp,j} + n_{\perp,j})} = \frac{1+0}{(1+0)+(1+1)} = \frac{1}{3}$. Hence her comparison is between, $R_d(n,\top) = -\frac{1}{3} + -\frac{2}{3}$, and $R_d(n,\perp) = -2(\frac{2}{3}) + \frac{1}{3}$, i.e. both options have a risk of -1. Assuming she does help, she observes that Bob did in fact need help, and $n_{n,\top}$ increases to 1 which raises $p(h|\perp)$ to $\frac{1}{2}$. This results in a revised risk for inaction ($R_d(n,\perp) = -\frac{1}{2}$), and because the risk of giving help remains at -1, assisting is now the safer option.

4.4.4 A Positive Approach to Decision Modelling: Cumulative Prospect Theory

There is significant caveat to the use of normative models of human decision making: real human behaviour is not always rational. Simon has discussed this extensively (1956; 1959; 1996; 2000) in the context of bounded rationality, which is the conceptual underpinning to the heuristic approaches I have explored in section 4.4.1. This feeds into the wider dichotomy between positive, and normative approaches to modelling of decision making which is, borrowing from Hume (1739), of an is-ought nature. This is to say that where normative approaches attempt to answer how decisions ought to be made, the positive approach sets out to reflect how people actually do so.

Of the positive school, the best known approach is that taken by Tversky and Kahneman (1992), with their CPT decision rule. CPT is similar to Bayesian risk minimisation, or expected utility maximisation, but distorts probabilities and outcomes to account for a number of observed deviations from strictly rational behaviour, notably loss aversion, and over and under weighting of probabilities. The original formulation of CPT treats only decisions where odds are provided, but as shown by Fox and Tversky (1998), it can also form part of a two-stage model to resolve decisions from experience. As in Chapter three, I have implemented this by exchanging the Bayesian risk minimisation rule of equations (4.6), and (4.7) for the CPT equivalents, shown previously in equations (3.8-3.16).

The precise form, and parameters for the CPT rule are subject to some debate, and may vary both with age (Kovalchik et al., 2005; Booij et al., 2009; Mikels and Reed, 2009; Albert and Duffy, 2012), and the presence or absence of stereotype threat (Carr and Steele, 2010). However, given the lack of a strong consensus, I have retained those of Tversky and Kahneman's original formulation (table 3.2, (Tversky and Kahneman, 1992, pp. 311–312)).

On our final return to Alice and Bob, Alice's behaviour changes, and she now always helps Bob. However, despite needing support Bob's behaviour remains the same and he continues not to seek it. Here, the representations remain identical and the primary reason that Alice makes a different decision is the curved utility function of CPT. The potential -1 loss is magnified to an effective value of $\lambda(-(1^{\beta})) = -2.25$, and the possible best outcome of 2 is reduced to $2^{\alpha} = 1.84$.

Having outlined the game, and players, and the underpinnings of the model in local authority and survey data, the following section proceeds to describe simulations undertaken to calibrate the four variations against real world data.

4.5 Simulations

As discussed in section 4.2, it is desirable to, as far as possible, support the inputs to the model empirically, using real world data. However, as is often the case with simulation, there are many parameters and data sources are not available for all of them. To assess how far uncertainty about the true values of the parameters leads to uncertainty in the output, I have performed a sensitivity analysis. A wide variety of techniques for the sensitivity analysis of simulations, and agent-based models (Thiele et al., 2014; ten Broeke et al., 2016; Grow, 2016), have been suggested, with several under loose categorisation of reduced models reviewed by Heard et al. (2015). Here, as in Chapter three, I have again followed the Bayesian Analysis of Computer Code Outputs (BACCO) approach (Kennedy and O'Hagan, 2001; Oakley and O'Hagan, 2002, 2004), which has seen some use in both the sensitivity analysis (Bijak et al., 2013; Hilton and Bijak, 2016), and calibration (De Mulder et al., 2015) of Agent Based Models (ABMs) to perform a variance based sensitivity analysis on the uncertain parameters, using the GEM-SA package (Kennedy, 2004).

This approach entailed fitting a statistical emulator of the simulator for each decision rule. Because the fitted statistical emulator has an analytical solution, this also has the secondary advantage of facilitating very rapid exploration of parameter space. As a consequence, I have been able to fit values for the unknown parameters that allow the model to reproduce the observed empirical data.

To construct the emulators, the simulator was run at 400 design points selected using Latin hypercube sampling of the parameter space defined in table 4.3 for each decision strategy[9]. Because the simulator is stochastic, emulators were fitted to the mean over 25 simulation runs at each input point, using leave one out cross-validation.

I captured three key measures: the proportion of that those who do and do not need help that received it, and the mutual information (MI) of the askers' signals, and their need for help. This last measure, MI, gives an indication of how much information askers' signals tend to reveal about their true state. Since the three simulator outputs are bounded between zero and one[10],

[9] γ,δ,α,β, λ, relative population sizes, and number of rounds in a simulation were as given in table 3.2

[10] Two are proportions, and the MI can convey at most a single bit of information in this game.

Table 4.3: Simulation parameters

Name	Description	Value(s)
G	Benefit of a good outcome	$0\ldots 10{,}000$
B	Cost for a bad outcome	$0\ldots 10{,}000$
C	Cost of stigma	$0\ldots 10{,}000$
P_d	Probability of a decider's experiences being shared	$0\ldots 0.1$
P_a	Probability of an asker's experiences being shared	$0\ldots 0.1$
Q_d	Weight of shared social information for deciders	$0\ldots 0.1$
Q_a	Weight of shared social information for askers	$0\ldots 0.1$
W_d	Weight of private prior information for deciders	$1\ldots 100$
W_a	Weight of private prior information for askers	$1\ldots 100$
H	Cost of providing support	7,881
p_h	Probability of an asker needing support	0.25
Logist(μ, b)	Distribution of pseudo-counts for stigma	Logist$(0.5, 0.06)$
$\mathcal{N}(\mu, \sigma^2)$	Distribution of pseudo-counts for trust	$\mathcal{N}(0, 0.83)$
MN(3, 2, -2, -3)	Distribution of pseudo-counts for support	MN$(0.26, 0.49, 0.17, 0.07)$

but the emulator is unbounded, a transformation was necessary. I have used an arcsine transform because it largely preserves the features of the underlying distributions.

4.6 Results

In this section I present sensitivity analyses which characterise the behaviour of the model with the four decision rules. These sensitivity analyses are based on statistical emulators, which I then use to fit the model to the level of unmet need found by Vlachantoni et al. (2011). This allows me to show that, for the two decision rules which incorporate a mental representation, the model is able to replicate observed levels of unmet need. I then combine the fitted values with the statistical emulators, and show how interventions targeted at the perceived value of receiving help, and information sharing would be predicted to change the level of unmet need for support.

Figure 4.2 shows the parameter sensitivity for all decision rules, and parameters, with complete details on all emulators produced provided in appendix F. The lexicographic heuristic (first row) behaves as step function, driven by the expected values of good, and bad outcomes (columns G, and B). The behaviour of the system flips between giving no assistance under any circumstances, and help given uniformly, without consideration as to whether it was needed. While this is obscured in the figure by the interpolating nature of the emulator, the plot does show that the critical point was almost exclusively dependent on the balance between the benefit of a good outcome, and the cost of a bad one.

The Bayesian rule (second row) is marginally more complex than the lexicographic heuristic, and while the sensitivity results for good and bad outcomes appear similar, the smooth transition between poles shown in figure 4.2 is no longer a result of interpolation. Instead, there was a gradual increase in helping behaviour which was, as before, largely a function of the balance between good and bad outcomes. As a result, it was possible to find parameters where on average, a person in need had an even chance of receiving help. The model still fails, however, because, as with the lexicographic rule, there was no discrimination between those who were in need of help and those who were not.

Both the Bayesian with mental representation, and the CPT models (third and fourth rows) also show an increasing level of support with higher rewards, but unlike the simpler rules, they also show differentiation in support provision. The rules behave similarly across the majority of parameters, with the differential effect tending to be less pronounced with the CPT rule. The models differ markedly in the expected response to increasing the perceived value of a good outcome. Where the normative model predicted that this lead to better differentiation

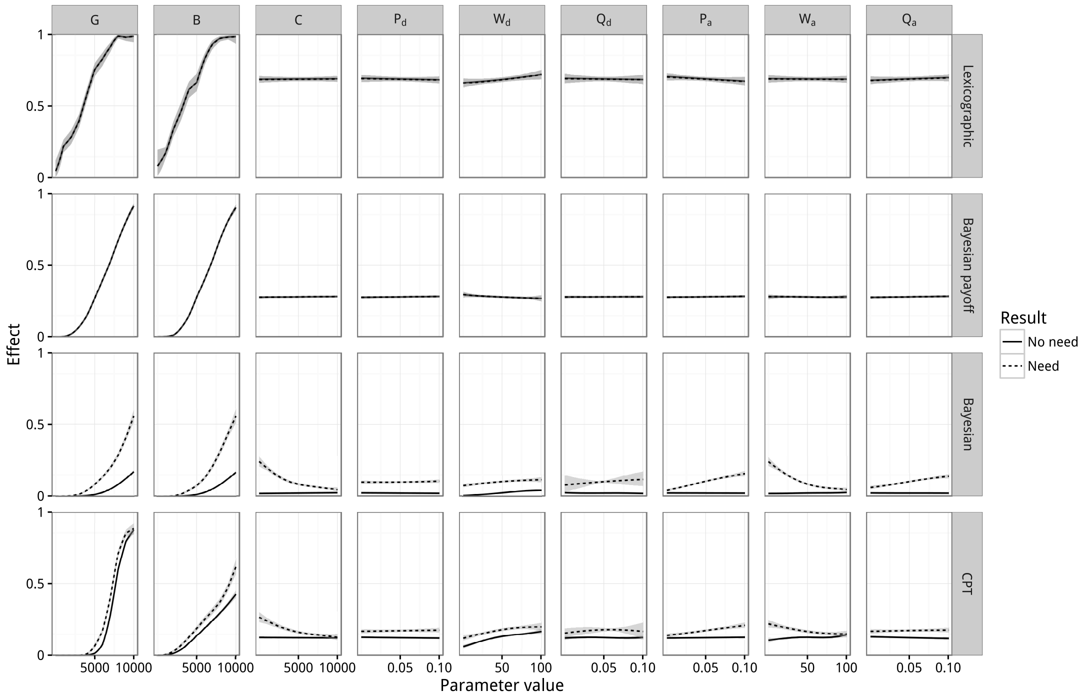


Figure 4.2: Parameter sensitivity for all decision rules. Average effect on the proportion of those who do, and do not need support that received it in response to varying a single parameter. Shown with 95% confidence interval, back-transformed from arcsine square root transform.

between need groups, under the CPT model this did not occur, and a similar although less extreme distinction can be seen for worse negative outcomes.

In both cases, I was able to go beyond sensitivity analysis and fit parameters against the 47.5% rate of support from Vlachantoni et al. (2011)'s analysis, using differential evolution (Storn and Price, 1997). Because the proportion of those who do not need help but receive it anyway is unknown, but may be reasonably supposed to be small, this required fitting on two fronts. Using the Euclidean point distance between $(p(+|h), p(+|n))$ and (0.475, 0.00) as the measure of fit allowed me to incorporate both of these criteria. Once a candidate set of parameters (table 4.4) was found using the emulators, I then tested them using the simulator. For the Bayesian rule with mental representation, this yielded $p(+|h) = 0.45$, $p(+|n) = 0.01$ with standard deviations of 0.046, and 0.003 over 100 simulation runs.[11] I was also able to successfully fit parameters (table 4.4) for the CPT decision rule, and simulating a candidate parameter set gave $\mu = (0.44, 0.01)$, with standard deviations of $(0.024, 0.015)$.

[11] Had the discrepancy between emulated and simulated values been large, the simulated result could then have been added to the emulator to improve accuracy.

Table 4.4: Fitted parameters for the CPT, and Bayesian with mental representation, decision rules

Name	Description	CPT	Bayesian
G	Benefit for a good outcome (£)	4,886.03	7,032.90
B	Cost for a bad outcome (£)	6,978.77	4,662.17
C	Cost of stigma (£)	3,500.59	860.41
P_d	Probability of a decider's experiences beings shared	0.09	0.06
P_a	Probability of an asker's experiences beings shared	0.07	0.05
Q_d	Weight of shared social information for deciders	0.02	0.06
Q_a	Weight of shared social information for askers	0.06	0.02
W_d	Weight of private prior information for deciders	28.22	15.62
W_a	Weight of private prior information for askers	17.61	63.16

In both cases, giving help represents a loss because the cost of giving help (H), based on the average across English local authorities, is $£7,881$. The loss is greater for the CPT decision rule, where $G - H = £ - 2994.97$, compared to $£ - 848.10$ for the Bayesian model. This difference may in part be because the CPT agents are risk averse under gains, and risk seeking under losses.

As in Chapter three, having fitted parameters to the simulation model, I then used statistical emulation to explore how the model would be expected to respond to perturbations in them. This is attractive if we consider the problem of how to focus policy on improving real world outcomes. As a demonstration, I show this for the sample points shown in table (4.4) in figures 4.3 and 4.4.

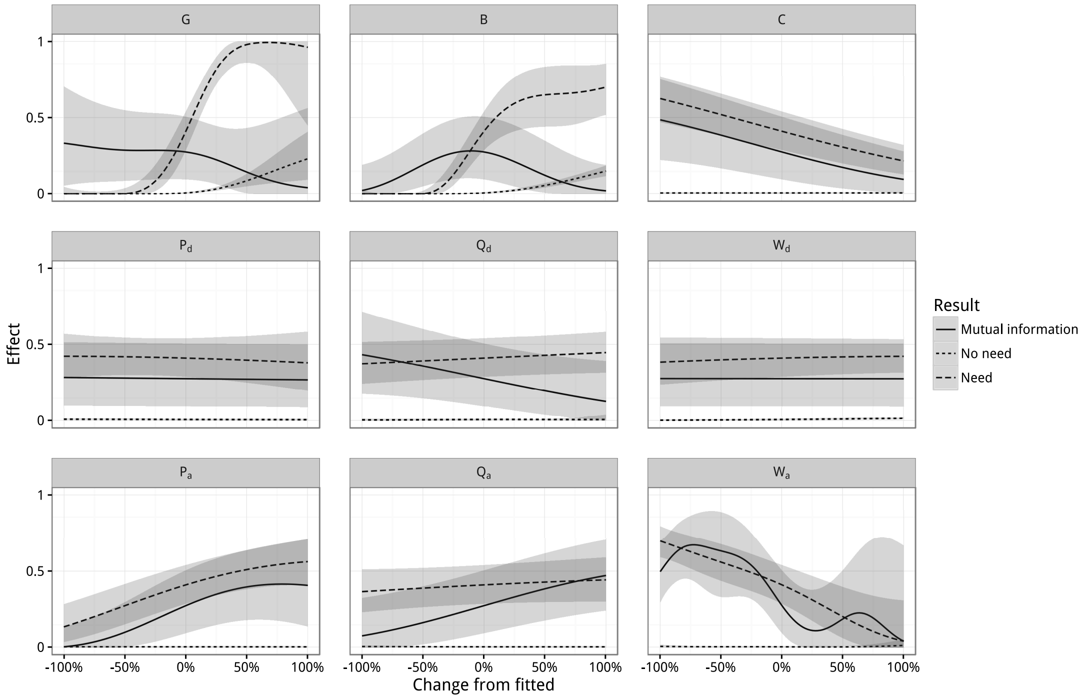


Figure 4.3: Mean expected change in outputs with 95% confidence interval back-transformed from arcsine square root transform, in response to varying one parameter by a percentage of the fitted value while fixing the others, for Bayesian decision rule.

Figure 4.3 shows the expected change in outputs when varying a single parameter from the fitted value, for the Bayesian rule with mental representation. Considering the expected change in outputs suggests which aspects of the system can be most profitably targeted to increase the number of those in need who receive support. Taking a normative perspective, it appears that the best target for intervention would be to increase the perceived value of a good outcome. A positive approach, as shown in figure 4.4, gives a contrasting perspective - greater perceived value does increase the receipt of help, but does so universally. This is also reflected by the trend in mutual information for both rules, which drops off rapidly for higher perceived values, indicating that communication begins to break down, eventually rendering it impossible for deciders to reliably distinguish those who need help.

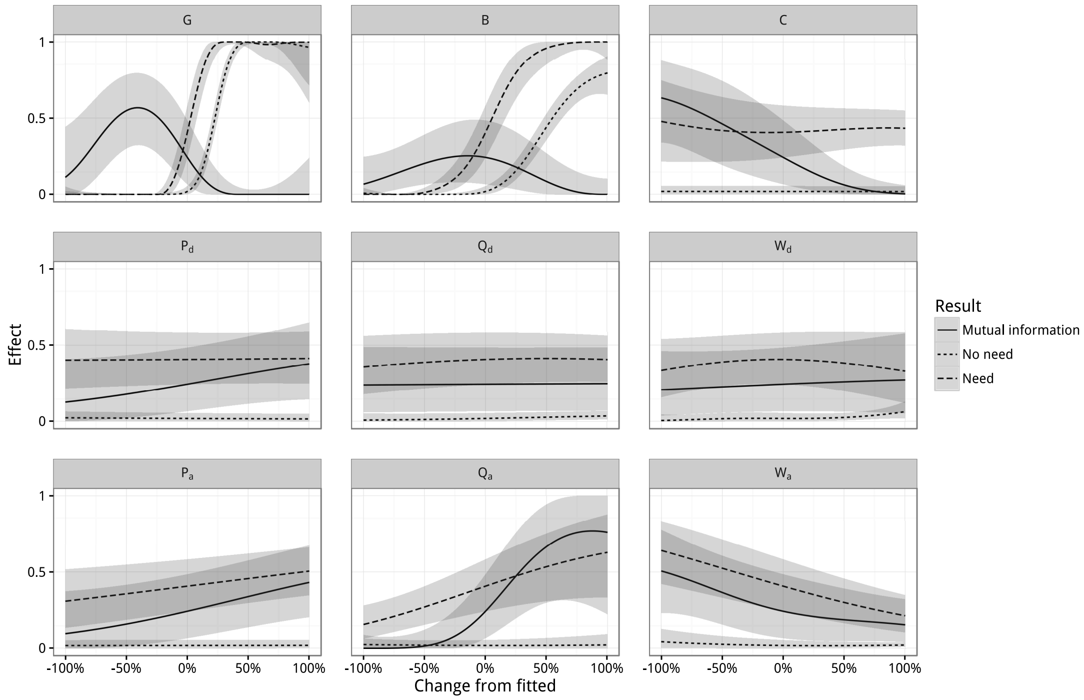


Figure 4.4: Mean expected change in outputs with 95% confidence interval back-transformed from arcsine square root transform, in response to varying one parameter by a percentage of the fitted value while fixing the others, for CPT decision rule.

Because as yet, there is no clear right answer to which decision rule is 'better' in this domain, I suggest an alternative to drawing on a single model for prediction. We should instead look for commonalities in system behaviour, which here were most evident in the social learning behaviour of askers, the cost of stigma (C), and the weight of askers' prior beliefs (W_a). Both the normative, and positive models predicted that reducing the perceived cost of stigma, or weaker priors for signallers would lead to more help given to those in need. In terms of actionable interventions however, the analysis indicated that increasing the probability of information sharing between askers (W_a) would achieve similar gains.

4.7 Discussion

Adopting an agent-based approach focused squarely on individual decision making has allowed me to explore possible explanations for an opaque, large-scale process. Although the model is, inevitably, incomplete, it does demonstrate a plausible link between qualitative explanations for a lack of help-seeking behaviour and observations at the population level. This is in contrast to the results of the simulations described in Chapter three, where the paucity of data limits the claims that can reasonably be made about the explanatory power of the model.

There are clear implications vis à vis the modelling of decision making, perhaps the most critical of which is that how the choice-making process is characterised matters a great deal at the macro scale. Unlike the midwives and women model shown in Chapter three, the failure of the simple action-consequence decision rules to adequately reproduce large-scale behaviour is indicative of this, and more specifically suggests that there is a certain minimum degree of sophistication necessary in an agent.

In itself, the apparent need for an agent to hold a mental model of the task they undergo raises a more challenging issue. Here, I have equipped the agents with a simple mental model, the structure of which is a remarkable match with the simple decision task. The reality is vastly more complex, and in the general space of human interactions it would seem grotesquely naive to imagine that every person reasons using an identical hypothetical process, the structure of which aligns perfectly with that of the decision task they face. As demonstrated by the substitutability of decision models, the general methodology that I have applied makes no demands of, and places no limitations on, the mental model used. As such, a potentially valuable line of future enquiry might be heterogeneity in agent's mental models of process. Along similar lines, while the agents have a model of process, they do not have a model of mind, and treat their opponents as a stochastic part of the process. More sophisticated approaches are possible, with several more game theoretic approaches centred on modelling the mind of the opponent discussed in recent work by Ríos Insua et al. (2015a; 2015b).

Naturally, all of the models that I have developed here are to varying extents unsatisfying and incomplete, although, as I have demonstrated, including a mental representation makes them more useful. The model of social learning and information flow is excessively simplistic, as is the homogeneous nature of agents in the deciding role. These aspects could be addressed though, by the inclusion of social networks into the model, which I see as especially important given that the majority of care is delivered informally, and by networks, rather than individuals (Wenger, 1991).

Another criticism is the static nature of the simulation. While time in some sense passes, agents are indifferent to it and unchanged by it. In reality it would be expected that the help required by askers varies over time, and past need for help influences future need for help, dependent on whether any help was received. Introducing an appropriate statistical model of transition probabilities would enhance the realism of the model, and potentially link well with a more sophisticated social network model. As in Chapter three, there is also a very clear sense in which the decisions considered here are intertemporal in nature, for example, the trade off of an immediate social cost against ongoing suffering. Similar remedies to those previously suggested may be applicable since it has been shown that people discount health, as well as money (Chapman and Elstein, 1995; Chapman, 1996).

Turning for a moment to the domain of the model, combining statistical emulation with simulation models inhabited by, and confronted with, data, has yielded powerful insights into how choices made by individuals can manifest in distressingly low levels of support. The simulations show that this can arise even when those that need help have a powerful incentive to ask for it and those that can offer help have a powerful incentive to provide it. In contrast to the model described in Chapter three, the infusion of the model with data has also allowed us, cautiously, to make policy recommendations, based on similarities in the responses of the normative and positive models. Of the possible policy interventions, I suggest that encouraging information sharing amongst older adults is most achievable. This is not without qualification however, bearing in mind concerns over the impact of stereotype threat raised in section 4.2, and the potential that counter-productive dependency stereotypes might be evoked in older adults who hear that other older adults have asked for, and received, support (Adams-Price and Morse, 2009).

The broader case to be made is that integrating macro-scale forecasting approaches with process models can be mutually beneficial. While statistical modelling is the more mature predictive tool, it is weak from the standpoint of theory and understanding. By contrast, agent-based models, and simulation more generally, excel as an explanatory tool, but their utility for forecasting in the social sciences is an as yet unknown quantity. Combining these approaches has the potential to strengthen both, leading to powerful theoretical tools, which support prediction and understanding at multiple scales. When considering the role of decision modelling in this, there is also a clear interdisciplinary interest: demography has a unique capability to evaluate the plausibility of psychological models, which focus on the local and individual, at the scale of populations. Put simply, if a model of decision making works, it would be expected that a simulation that incorporates it should exhibit plausible large-scale behaviour. In this sense then, explicitly modelling the micro-macro link can contribute to our understanding at both ends of the scale. This highlights a key deficiency in both this model, and that presented in Chapter three, in that they might be said to affirm the consequent – with the infusion of data, I am able to show that the model demonstrates plausible behaviour at the macro-scale, but this does not imply that the models of the decision making are sound.

In the following chapter, I address the micro scale, and use controlled experiments to assess the extent to which that latter assumption – that the models of decision making work – can be supported.

Chapter 5

Decision Making Under Paired Gambles from Experience

Abstract

In this paper I report experiments designed to validate the choice behaviour of agents in an agent-based model. I examine human decision making about paired gambles from experience, where the gambles have a common feature. I report results for eight decision problems undertaken by 20 participants, and contrast the predictive ability of four models of decision making. I then estimate parameters to maximise the fit where possible, and find that while the best performance is offered by decision models with a representation of the problem, they do not offer a significant advantage over heuristic methods. I discuss the implications of this, in the context of the original agent-based model, and for agents more generally.

5.1 Introduction

As agent-based modelling grows in popularity, the foundations become more critical. The mechanism by which agents make decisions is the crux of agency, yet there is substantial disagreement about what a decision model should look like (Klabunde and Willekens, 2016). To some extent, this stems from the relative weakness of validation methods for agent-based modelling, which arises in part from their complexity. An Agent-Based Model (ABM) is a restricted version of reality, built of approximations of those aspects the modeller believes necessary to produce the phenomena of interest. This means that any given ABM is a system of many models, which may or may not be valid in isolation. Decision making is perhaps the most obvious example of this, seeking as it does to reduce the complexity of human behaviour to a manageable, minimal model. This is important in demography, where individual decision making is a secondary focus, but underlies most phenomena of interest.

I argue that one route to validation of complex ABMs is to complete the circuit, by having actual human participants enact the model, with the expectation that their behaviour in this constrained environment can be predicted by the ABM.

In this paper, I take the ABM introduced in Chapter four, and experimentally validate the decision models applied in it. To do this, I have created a decision problem analogous to the one solved by the agents, and recruited participants to solve variations on it.

The following sections proceed to describe the agent-based model, and then to create an decision problem which follows the structure of the one solved by the agents (section 5.2). I then outline the decision models used in the ABM (section 5.3), and use them to generate variations of the decision problem which produce contrasting predictions by the models (section 5.4). I then report the design and results of an experiment where participants solve the variations (section 5.5), and contrast the predictive performance of the four decision models with fitted variants, and additional alternative models (section 5.6). Finally, I discuss the implications for the original ABM, and agent-modelling more generally (section 5.7).

5.2 Decision Problem

In the ABM described in Chapter four, which I wish to validate, two populations of agents, representing older adults and care providers, play a series of games. Some of the older adults require support in their day-to-day lives, and a game consists of their choosing whether to ask for help, and the provider choosing whether to give support. Older adults are also concerned with avoiding the potential stigma of requesting help, and believe that some portion of the provider population will react negatively to them if they are asked for help. In addition, they may not believe that they will receive help, even if they do ask for it. In the model, older adults play a limited number of such games with randomly chosen providers, leaving the game after they have played a predefined number of games or received help. After each game, they observe the response of the provider, whether they are the type to stigmatise, and learn from the experience.

Providers are faced with a scenario where some of the population needs help, but not all will request it. They may also be uncertain as to whether all requests for help are genuine. Providing help, and failing to provide it where needed, both exact a cost, the latter being the same as the cost paid by the older adult who went unhelped.

In this paper, I concern myself with the problem faced by the older adults, which in isolation from the model as a whole can be described as in figure 5.1.

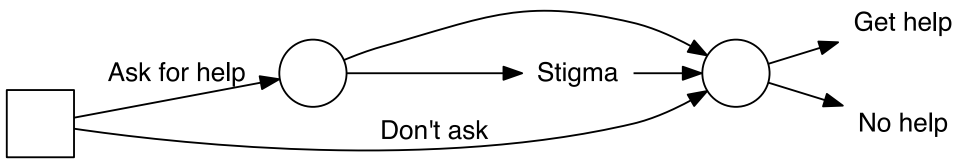


Figure 5.1: Decision problem experienced by agents in need of support. Circular nodes represent chance events (from the perspective of the agent), and square nodes signify a decision point.

5.2.1 Three Coins

We can imagine a simple game of chance, analogous to the decision to ask, or not ask for help, where the player may choose to toss one of two coins. We will label the coins A and C. Both coins offer the same payoffs for heads and tails, but have differing biases.

A slightly more complex scenario arises if we add a third coin, B, and require the player to always toss it in addition to one of A or C. We will also say that coin B, while always tossed and observed, will only give a payoff if partnered with coin A[1]. This third coin is equivalent to the risk of being stigmatised when asking for help. The resulting game is shown in figure

[1]Or equivalently that the payoff for coin B is a function of both the face, and the action, which happens to be zero when the action is C.

5.2, which as can be seen by comparison with figure 5.1, is structurally identical with the older adults' decision problem, differing only in the details of the payoffs.

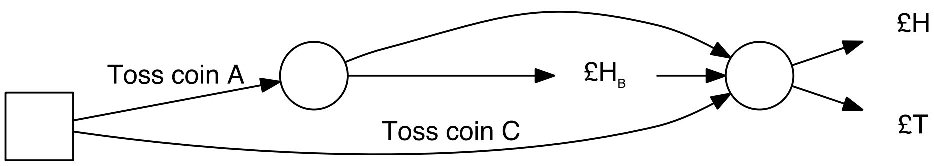


Figure 5.2: The three coins decision problem, structurally identical to that shown in figure 5.1.

In the original model, the costs and payoffs are fitted to replicate the proportion of older adults who receive support when they require it, and the bias of coins A and C is a function of the experience of providers. Here, I will allow the costs and payoffs to differ between decision problems, and hold the bias fixed within them.

5.3 Decision Models

5.3.1 Model-Free Approach

I consider two decision approaches which are model-free, in the sense that they pay attention only to the payoffs they receive and do not consider the structure of a problem. For example, in the coin flipping game, if there are multiple outcomes that lead to the same payoff, these approaches treat them as equivalent.

As mentioned in Chapter 3, the first model is the Lexicographic heuristic, which makes choices by comparing the most common payoff of each option. In the event of a tie, the procedure is repeated for the second most common payoff, until the tie is broken or all payoffs are explored (Luce, 1956; Fishburn, 1974)[2]. The heuristic learns by keeping a count of the payoffs which it has observed.

[2]The behaviour of the Lexicographic heuristic in the event of all options being equal is typically unspecified, but here I allow it to choose randomly if this arises.

Building on this foundation, I introduce a simple model of Bayesian Risk, which I term Payoff Bayes (PB) (Gray et al., 2016). This model, as with the Lexicographic heuristic, considers only payoffs, but applies Bayesian updating to them and chooses the action which minimises the Bayes risk (DeGroot, 1962), as in equation (4.2. In this instance, lacking any good reason to believe a priori that any payoff is more likely than another, the hyperparameters to the Dirichlet distribution shall be uniform such that $\alpha_{a,j} = 1$, i.e. Bayes/Laplace uninformative prior (Agresti and Hitchcock, 2005).

5.3.2 Model-Based Approach

I also introduce two models which encode a mental representation of the decision problem. In the coin flipping game, this means that the models pay attention to the faces of the coins, rather than solely the payoffs.

The first model extends the PB model to differentiate between a payoff and an outcome by introducing,

$$R(a) = \sum_{i \in V_{a,1}} \dots \sum_{j \in V_{a,n}} -u(i, \dots, j) \cdot p(i \wedge \dots \wedge j | a), \tag{5.1}$$

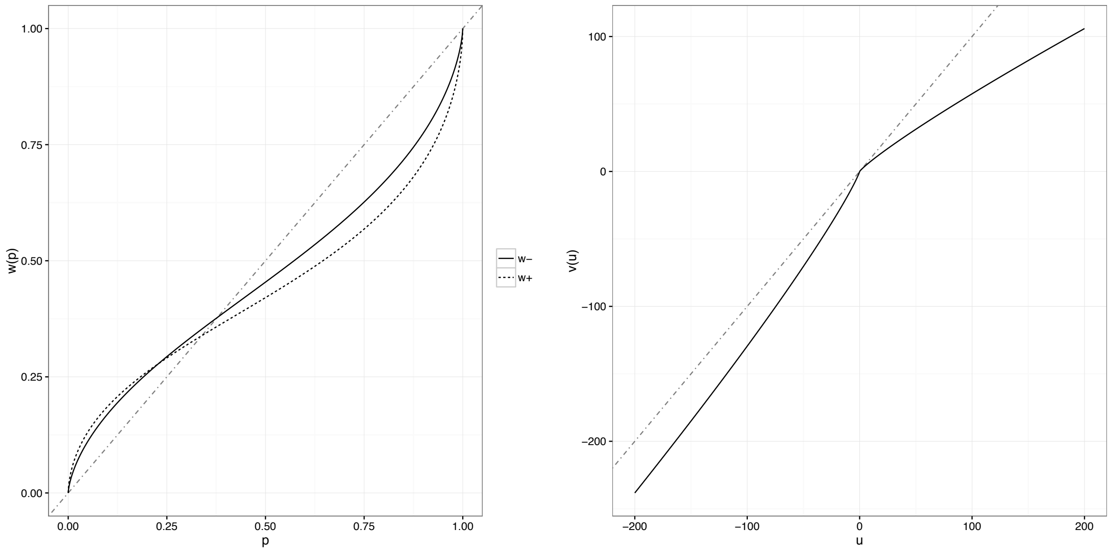
where $V_{a,n}$ is the nth set of outcomes which may come from action a, $u(i, \dots, j)$ is the payoff if the outcomes occur together, and $p(i \wedge \dots \wedge j|a)$ is the probability of this happening.

Assuming $i \wedge \dots \wedge j$ are independent events, $p(i \wedge \dots \wedge j|a)$ can be equivalently decomposed to $p(i|a) \cdot \dots \cdot p(j|a)$, and in the case of the coin toss game $R(a)$ can be written as,

$$R(a) = \sum_{i \in V_a} \sum_{j \in V_B} -u(i, j) \cdot p(i|a) \cdot p(j), \tag{5.2}$$

where $V = \{Heads, Tails\}$, and $a \in \{A, C\}$, because coin B is always tossed, making it equivalent to equation (4.6) in Chapter four. This also implies that the model evaluates the face probabilities for all three coins separately. As a result, and unlike the model-free models, information is gained about the likelihood of payoffs when tossing coin A, from tossing coin C, and vice versa, by virtue of observing coin B. As with the PB model, the pseudo-counts for all coins are equal at one.

Finally, I will introduce the two-stage Cumulative Prospect Theory (CPT) model (Tversky and Kahneman, 1992) described in chapters three, and four, and given fully in equations (3.8-3.16). The CPT model adds considerable flexibility, and the ability to model individual idiosyncrasies in choice behaviour, by capturing differing degrees of loss aversion, and approaches to probability. Figure 5.3a shows how the model distorts probability, understating the likelihood of common events, and overemphasising rare ones, an effect which is more pronounced for negative outcomes. Figure 5.3b shows how CPT affects the value of an outcome, such that a loss weighs more heavily than an equivalent gain.



(a) Asymmetric weighting of probability for losses and gains.

(b) Asymmetric weighting of value for certain losses and gains.

Figure 5.3: Illustration of distortions to probability and value under the CPT rule with $\delta = 0.69$, $\gamma = 0.61$, $\lambda = 2.25$, $\eta = 0.88$, and $\beta = 0.88$. In both cases, the untransformed relationship is indicated by a dashed grey line.

This flexibility comes at the expense of an increase in complexity and challenges of parameterisation. I evade the latter at this stage by applying the parameters originally found by Tversky and Kahneman (1992).

Having introduced four models of decision making, in the following section I report the design and results of an experiment to determine which best predicts human behaviour in the three coins game.

5.4 Experiments

5.4.1 Participants

Twenty members (11 male, nine female) of the University of Southampton participated. Their ages ranged from 25 to 43, with a median age of 28.5 years (IQR=7.75). All participants were provided with a copy of the participant information sheet (appendix H.2), as approved by the University of Southampton ethics board (approval number #20700).

5.4.2 Decision Problems

The decision problems were created by exploiting the nature of simulation, in order to target the efforts at validation. To maximise the information gain, I wished to design problems where the agent models would give contrasting predictions from one another. In practice, this means identifying combinations of payoffs, and coin probabilities where a single agent model predicts with a high degree of certainty that the final choice, irrespective of sampling sequence, will be to flip coins A and B, where the others are adamant that the choice will be B and C. To achieve this, I defined the absolute difference of mean predicted agent choice for a single rule vs. all others, or between two pairs of rules as the measure for problem quality. I was then able to use this as a fitness criteria to apply differential evolution[3] (Storn and Price, 1997), to evolve decision problems which satisfied this criteria by treating the probabilities of heads or tails for the three coins, and the corresponding payoffs as a genome to be iteratively recombined and mutated. Parameter sets were evaluated over a simulated population of 1000 agents, who made 10 random samples of the coin pairs before making a final choice. All agents used the same parameters, with $\alpha = 1$, and parametrisation of the CPT model as in (Tversky and Kahneman, 1992, pp. 311–312), i.e. $\eta = 0.88$[4],$\beta = 0.88, \lambda = 2.25, \delta = 0.61$, and $\gamma = 0.69$.

[3] Although any global optimisation algorithm would be similarly effective.

[4] I have used η here in place of α in the original formulation to avoid confusion with the prior distribution hyperparameter.

Table 5.1: Summary of decision problems. $p(H)$ shows the probability of obtaining Heads when flipping each coin, Payoffs shows the payoff received for the coin faces of A and C (H, T), and for Heads when coin B is flipped with coin A (H_B). Contrast prediction shows the final coin pair choice predicted by a model or models, where the remaining models predict participants will choose the other alternative.

		$p(H)$			Payoffs			
Decision Problem	A	B	C	H	T	H_B	Contrast Prediction	Model(s)
1	0.990	0.040	0.010	-67	-64	58	BC	Lexicographic
2	0.010	0.010	0.997	4	-13	65	AB	PB
3	0.990	0.996	0.020	57	18	-20	BC	CPT
4	0.002	0.999	0.997	-80	29	-87	BC	Bayesian/CPT
5	0.850	0.100	0.100	-96	-95	67	BC	Lexicographic
6	0.102	0.500	0.103	-21	-17	1	BC	PB
7	0.880	0.530	0.100	95	22	-52	BC	CPT
8	1.000	1.000	0.002	-76	-64	10	AB	Bayesian/CPT

5.4.3 Design and Procedure

Experiment participants answered eight decision problems (shown in table 5.2), which were variations on the basic form described in section 5.2.1. Each decision problem consisted of a choice between flipping two of three coins (A, B, and C), where the participant could choose to flip either A and B, or B and C. Coins A and C offered the same pair of payoffs for heads and tails, but with differing bias. Coin B offered a payoff for heads which was only received when it was flipped with coin A. Participants were given 10 sampling rounds, before making a final choice of which coin pair to flip.

Decision problems were presented in a random order on a computer screen, which displayed the payoffs of the individual coins, with two buttons (AB, and BC) which when clicked displayed the outcome of the tossed coin pair as coin faces, payoffs for the individual coins, and the total payoff, for 5 seconds. The final choice was indicated by colouring the buttons red, and displaying a banner reminding the participant that this was their final decision for this problem.

After completing all questions, the outcome of a single final choice was selected at random, and participants were paid at £0.10 per (positive) point received, in addition to £12.00 for participating.

The full experimental protocol is provided in appendix G, in addition to the complete ethics application as approved by the University of Southampton ethics board (appendix H).

5.5 Results

Table 5.2 shows the sampling, and final choice behaviour of participants for each question. Particularly noteworthy is that the contrast predictions are universally incorrect. I have also used the individual sampling data to train the four models, by pairing each participant with an agent who experienced the same sampling sequence. Figure 5.4 shows the accuracy of the four models when predicting final choices based on the participant's experience of play. The Lexicographic model, the simplest of the four, offers the best performance, predicting 77.5% of cases correctly. The CPT model performs very poorly, and actually does worse than could be achieved by assuming the same choice for every case.

Post-session questioning of participants suggests a possible explanation for this result, with many reporting that in several of the problems the coin flips felt 'certain'. Because the remaining three models take a Bayesian approach to the evidence using a Bayes/Laplace uniform prior with an α hyperparameter of one, they are considerably slower to follow the weight of evidence. The worse than random performance of the CPT model may also suggest that the parameters used by Tversky and Kahneman (1992) are not representative of the participants. In section 5.6 I address this by comparing performance when using other reference priors, fitting the models, and introducing alternative Logistic, heuristic, and expected utility models.

Table 5.2: Summary of sampling and final choice behaviour for the eight decision problems.

Decision Problem	Sampling AB (in %)	Choice AB (in %)	Minority Prediction
1	55	45	BC
2	47	0	AB
3	52	80	BC
4	51	90	BC
5	66	60	BC
6	57	55	BC
7	56	85	BC
8	47	0	AB

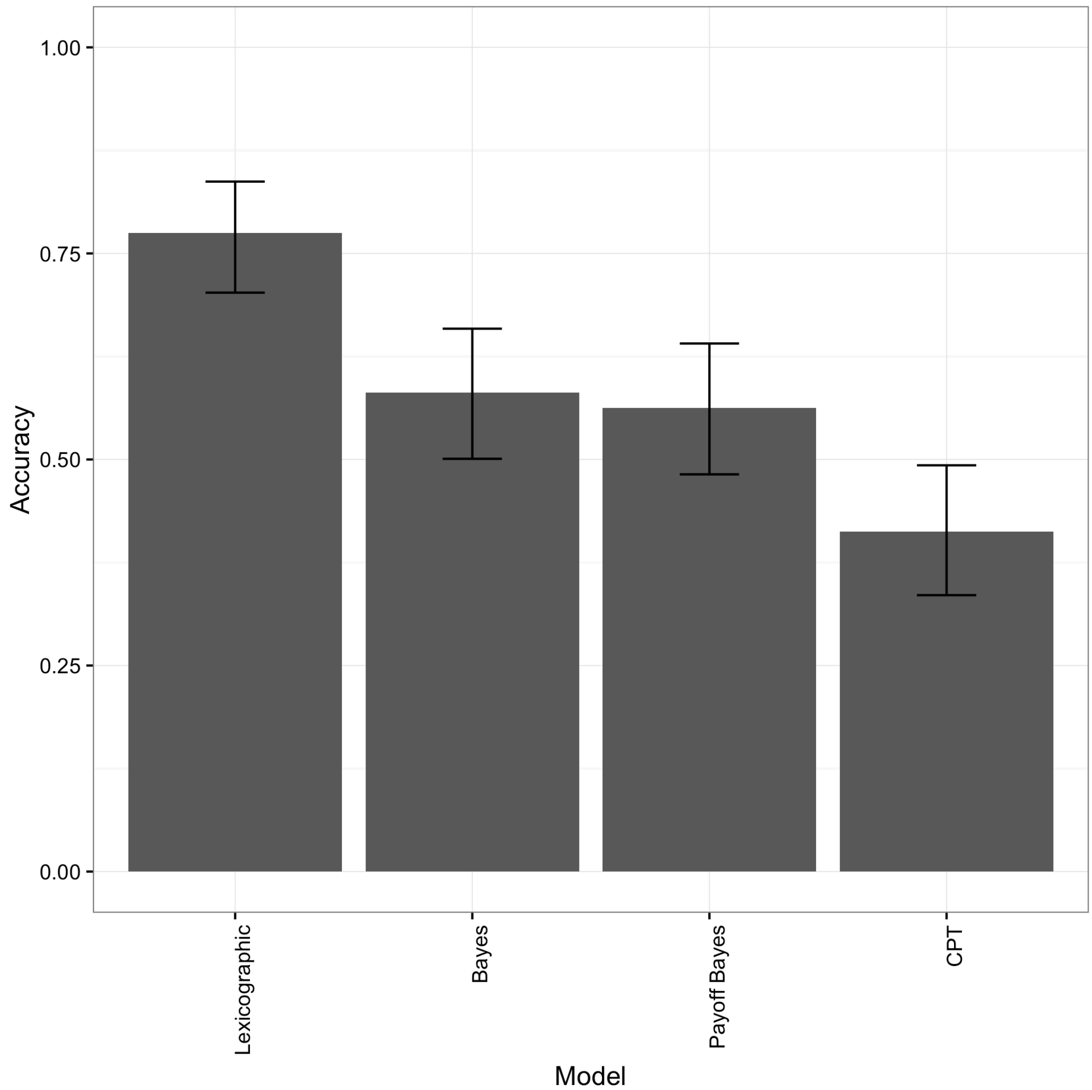


Figure 5.4: Accuracy with 95% confidence interval for the Lexicographic, Bayesian, PB, and CPT models when trained on individual play data.

5.6 Model Fitting

In this section I show that proper parametrisation of those models which permit it (the PB, Bayesian, and CPT models) can improve their accuracy, I also compare them with expected utility, heuristic, and Logistic regression based alternatives.

Specifically, I include two heuristics shown to perform well in predicting decisions from experience by Hau et al. (2008). The Maximax (MM) heuristic, which chooses the option with the highest observed payoff, and the Natural Mean (NM) heuristic, which selects the option with the highest mean payoff[5]. I also include expected utility (EU), which chooses the action which maximises equation 5.3.

$$E(a) = \begin{cases} H \cdot p(H|A) + T \cdot (1 - p(H|A)) + H_B \cdot p(H|B), & if\, a = A \\ H \cdot p(H|C) + T \cdot (1 - p(H|C), & if\, a = C \end{cases} \tag{5.3}$$

I include expected utility (EU) models based both on the true probability distributions for reference, and a version which learns from experience by estimating $p(H|x)$ as the fraction of heads observed for that coin. The EU model is widely used, but is also known to fail to predict human decision making in a number of circumstances (Tversky and Kahneman, 1974). As a further point of comparison, I also include the equivalent CPT model, also based on the true probability distribution and without learning.

For the three decision rules which use Bayesian inference, i.e. the PB, Bayesian, and CPT models, I have tested the impact of different prior distributions which reflect the observation that these models learn too slowly. I introduce two variants on each, which use a Jeffreys prior ($\alpha = 0.5$), and Haldane prior ($\alpha = 0$) (Agresti and Hitchcock, 2005). In addition, I have fitted parameters for the PB, Bayesian, and CPT models. In all three cases, I fitted parameters for individual participants using randomised parameter search with 3-fold cross validation. Because an individual's play history consists of only eight games, this used the whole data set for fitting, and as a result is likely to be over fitted. I have also used the same approach to fit parameters based on the aggregate dataset, with 10-fold cross validation, and 5% of the games held back.

Figure 5.5 shows the accuracy for all models, with the CPT and Bayesian models fitted at the individual level clearly offering the best predictive performance, likely as a result of overfitting. However, the Bayesian model still offers good performance when fitted to the aggregate data, while the CPT model suffers considerably, although since the intent of the model is to capture individual variability, this is perhaps unsurprising.

That the individually fitted variants of both are the most accurate supports the contention that a representation of problem structure is beneficial. Counter to this, however, is the strong performance of the NM heuristic, already known to perform well on decisions from experience

[5]As with the lexicographic heuristic, behaviour in the case of a tie is unspecified, but here I allow the model to choose randomly.

(Hau et al., 2008), and, with an appropriate prior, the PB model. These models do not represent problem structure and, excepting the PB model, pay no mind to probability. They are not significantly less accurate than the CPT model fitted to the aggregate data, and the NM heuristic does not significantly under perform the best aggregate model based on a comparison of area under the curve (AUC)[6] (AUC$= 0.867, Z = -1.296, p = 0.098$) if not constrained to binary prediction. The MM heuristic performs slightly less well than the NM heuristic (the reverse was found by Hau et al. (2008)), which may suggest that it understates the amount of information participants use in making a decision. An alternative perspective would be that people employ heuristics suited to the task at hand (Gigerenzer and Goldstein, 1996), and that the MM heuristic is less convenient here because of presence of the third coin.

I have also fitted two Logistic regression models (coefficients for both are shown in table 5.4), commonly used in demography as agent decision models (Klabunde and Willekens, 2016). These predict based on the same information as the EU model, i.e. observed proportion of heads for the three coins and the corresponding payoffs. I include models with (Logit$^{(x)}$) and without (Logit) interaction terms introduced by step-wise AIC. Introducing interaction terms yields good accuracy, but at the expense of a somewhat convoluted model which is bound to the structure of the problem, and hence does not generalise readily. Like the NM heuristic, both logit models perform better if probabilistic predictions are permitted. In the probabilistic case, Logit$^{(x)}$ is the second best-performing model, with an AUC of 0.93 not significantly different from that of CPT$^{(3)}$ (0.94).

[6]This metric is the area under the true positive rate-false positive rate curve. When considering probabilistic classification, the AUC gives the probability that given two randomly chosen final choices, one for AB and the other for BC, the classifier will rank the AB choice higher, assuming AB is encoded as one, and BC as zero.

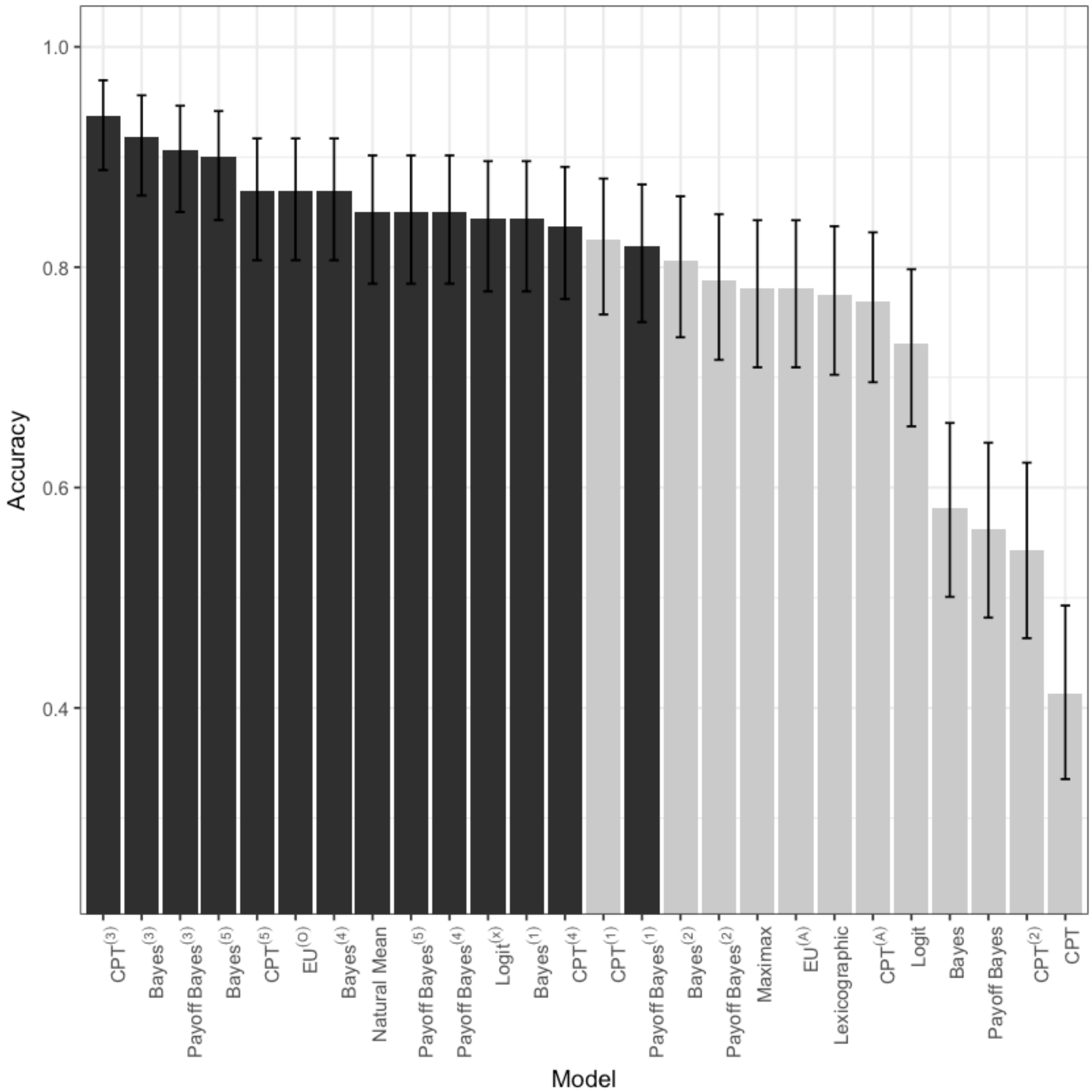


Figure 5.5: Accuracy with 95% confidence interval for binarised predictions of all decision models, over all participants and gambles. Superscripts indicate (1) Haldane prior, (2) Jeffreys prior, (3) Individually fitted, (4) Median of (3), (5) Randomised 10-fold cross validation, (x) Logit model with interaction terms, (A) True probability distribution, (O) Probability distribution observed by the individual. Models shown by light grey bars perform significantly worse[7]than the cross-validated $\text{CPT}^{(5)}$ model (McNemar's exact test, $p < 0.05$)

Figure 5.5 also shows that the value of α, in models applying Bayesian inference, affects performance significantly. One way to interpret the α parameter to the models applying Bayes rule is as a learning rate which approaches zero as α approaches infinity. With a uniform $\alpha = 1$ these models learn slowly, which may underlie their poor performance. As shown in table 5.3,

[7] $\text{CPT}^{(1)}$, despite having a marginally higher accuracy than $\text{PB}^{(1)}$, is strictly worse than $\text{CPT}^{(5)}$.

Table 5.3: Best parameters for Bayesian, PB, and CPT decision models based on median of individually fitted parameters, and randomised 10-fold cross validation.

Parameter	Bayes$^{(5)}$	PB$^{(5)}$	CPT$^{(5)}$	Bayes$^{(4)}$	PB$^{(4)}$	CPT$^{(4)}$
α	0.11	0.12	0.01	0.13	0.12	0.02
η			0.84			0.87
β			0.49			0.88
λ			5.44			2.25
δ			0.81			0.70
γ			0.59			0.61

Table 5.4: Coefficients for the logistic regression models.

	Logit		Logit$^{(x)}$	
Parameter	Estimate	$p-value$	Estimate	$p-value$
$p(H\|A)$	1.06	>0.05	0.47	0.26
$p(H\|B)$	-2.59	0.09	3.20	0.16
$p(H\|C)$	-2.60	>0.05		
H	-0.02	0.01	-0.37	>0.001
T	0.03	0.05	0.65	>0.001
H_B	-0.02	0.03	0.02	0.63
$H:H_B$			-0.003	>0.01
$p(H\|B):H_B$			0.27	>0.01
$p(H\|A):H$			0.20	>0.01
$p(H\|A):T$			-0.20	>0.05

the best fit weight is between the Haldane and Jeffreys priors, with a smaller weight, and hence faster learning, producing better performance.

The potential non-uniformity of α can also have significant effects, as demonstrated by the advantage of the EU model with observed frequencies, over the Bayes model with a Haldane prior. Superficially, the two should behave identically, however in scenarios where a player only samples from a single coin pair, the EU model will assume that the unsampled coin is certain to land tails, whereas the Bayesian model allows the probability to be zero and estimates the payoff accordingly.

While even the simple learning model I have used here introduces additional requirements in terms of parametrisation, it does offer some benefit, as can be seen when comparing the performance of CPT$^{(A)}$, with the learning alternatives. The variant based on the true probabilities unequivocally outperforms both the $\alpha = 1$, and Jeffreys prior versions. However, it is a less accurate predictor of behaviour than the model using the Haldane prior, which parallels the relative performance of the EU rules.

These results suggest that both learning and a mental representation of structure are beneficial, although not greatly so in the latter case. In addition, while the most accurate predictions can be obtained by attending to individual variation, this may be outweighed by the more parsimonious requirements of heuristic methods in terms of data and computation.

5.7 Discussion

In terms of the originating ABM, I can make some tentative recommendations, although it should be acknowledged that I have examined only half of the decision problem here, and that unlike the ABM, sampling is "free" because payoffs are observed but not actually received. Additionally, the risk of being removed from the game (by receiving help, as in Chapter four), is not present here . Most critically, the weight of priors may be significantly too great, and are likely to dominate learning. I would also suggest that the addition of a natural mean-based agent-model may be beneficial.

Considering agent-based modelling more generally, much depends on the requirements of the modeller. Both heuristic methods, and Logistic regression perform well in empirical terms, particularly where stochastic choices are acceptable, or even desirable. Logistic regression remains a particularly attractive option where there exists data to fit a model, and where there a viable mapping can be found from that data to the environment of the ABM.

There are some clear limitations to this research, of which the greatest is the small sample size, which of necessity make this only a pilot study. A second key limitation is that the structure of the problem is provided from description, which elides the question of how a mental representation is constructed. Hau et al. (2008) suggest, in essence, that in decisions from experience, people develop faulty models of the true probabilities. A similar model discrepancy is likely if they must also learn the structure of the problem. Models which incorporate this aspect may offer greater explanatory, and predictive power.

The strong performance of the NM, PB, and EU models suggests that the small advantage offered by more complex models may be outweighed by the challenges of parametrisation. However, where parametrisation is achievable, there is a gain to be made from a more complex model. This is still not without compromise however, if we bear in mind the additional computational demands associated with a more sophisticated model. This more prosaic constraint is perhaps overlooked, but becomes increasingly pertinent as the scale of simulation increases, and hence is relevant in demographic modelling in particular, where large populations are often a necessary feature.

An alternative possibility, is that the mathematical nature of the decision problems invites the use of simpler mental accounting like strategies. This can be related to the granting of a structural representation to the decision models, and suggests that participants may be using a simper mental representation, more akin to that of the payoff focused decision models. This may be heightened by the close to certain outcomes in several of the problems. These arise because, with the priors used in generating the decision problems, the potential for an outcome carries substantial weight even when it is never observed to occur. Whether they perform well in more subjective, less fungible environments is a question worthy of further study.

The other critical trade off is in fidelity of model. Much of the purpose of agent-based modelling hinges not on the ability of these models to offer predictions, but on their power to explain and describe (Epstein, 2008). In a sense, the majority of the decision models I have considered here fail in this regard. The logistic regression models are perhaps the best example. They offer, in the best case, almost unrivalled predictive capability. However, they are fundamentally unilluminating about the actual process of decision making, beyond informing us about the relative importance of the inputs. More troublingly, the model does not generalise beyond the specific problem. The more human focused alternatives generalise more readily, and a single model is better able to predict behaviour under a variety of problems. It is also important to note that all the models face challenges where it is not clear how to value outcomes. This is least problematic for the Lexicographic heuristic, which depends only on the existence of an ordinal comparison. It is considerably more problematic where a cardinal valuation is necessary, as is the case for the majority of the models considered. As is often the case, this can be ameliorated with more data by, for example, eliciting utility equivalents. Neither option is perfect, since dispensing altogether with a sense of the relative magnitude of outcomes is unrealistic, whereas the alternative increases already significant data demands and may fail where it is very challenging to translate an outcome to, effectively, money.

In this paper I have used discrete choice experiments to validate the behaviour of agents in an ABM. I have shown that while the decision models used can perform well in predicting human decisions, the parametrisation used in the original simulations is likely to be flawed. More broadly, I have outlined a method which may usefully be replicated to exploit the advantages of simulation in designing experiments to validate individual components of a complex model. The necessity of this validation is demonstrated by the ability of the ABM to reproduce aggregate patterns of care supply to older adults. This apparent validity at the macro scale is based on faulty foundations. This is illustrative of the need for multilevel approaches to the validation of complex models in general, and as suggested by both Courgeau et al. (2016), and Conte et al. (2012) the need to construct models from empirically valid components.

Chapter 6

Summary and Conclusion

6.1 Summary of Results

The primary contribution of this work is in showing the necessity for, and a route to, calibration and validation of agent-based models at both macro, and individual scales. In addition, it develops a methodology for the design of agent-based models which emphasises simulation at a bridge between simplified analytical approaches, and intractable reality, while supporting the modularity, and experimental validation of the components of the simulation. This work pays particular attention to the modelling of decision making, demonstrating that how this is implemented in the model makes a significant difference to the model's behaviour, and advocating for agent-based models which support a modular approach to decision modelling in the absence of a universally agreed upon model of choice behaviour. This approach has been elucidated over the course of three papers, which begin by demonstrating the approach, and showing that the resulting model is able to reproduce stylised facts, but is significantly lacking in that it cannot be validated against empirical data. The second paper applies the approach to a substantive problem, and calibrates the model against real data. The approach facilitates empirical calibration at the macro-scale, but is missing any assurance that a critical component - the decision models used by agents, is valid. Finally, the third paper addresses the validation of the individual level of the model using controlled laboratory experiments on individual decision making.

6.2 Limitations

The crux of the limitations of this work is the challenge inherent in validating, in any meaningful way, an agent-based model. While models can be successfully calibrated, and I have done so here, a calibrated model is not *eo ipso* a valid one. This applies particularly to the older adult simulation model (chapter 4), which is successfully calibrated, but as demonstrated by the experimental results obtained in chapter 5 could not comfortably be said to be valid.

There are more prosaic limitations, for example there is insufficient data to calibrate or validate the alcohol misuse disclosure model (chapter 3), beyond a cursory qualitative comparison. While this is clearly a limitation, it does also represent an opportunity, if considered from the mindset that part of the role of agent-based modelling is as a tool to guide empirical research (Epstein, 2008).

The small sample size for the coin tossing experiment (chapter 5) is also a limitation, although the concordance with the results found by Hau et al. (2008) ameliorates this concern to some extent. The experiment also addresses less than half of the problem, since the simulation results in chapter 4 arise as the result of interaction with agents resolving a different challenge. This highlights part of the challenge in validating complex models: there is no limitation on the complexity we can express in simulation, but considerable constraint on what can be empirically tested. A more extensive validation would be desirable here, evaluating both the other side of the decision problem, and how the inclusion of interaction might affect behaviour. Arguably, the more principled solution to this is that proposed by Courgeau et al. (2016), namely that the direction of research should be to first validate the individual components before building models from them.

More broadly, I have made two key underpinning assumptions that are not tested in this work. The first is that emotional outcomes are evaluated in a way comparable to tangible outcomes. This not unreasonable from an economics perspective, given, for example, the link between outcomes and risk aversion in relation to drug treatments found by Eraker and Sox (1981), and the suggestion by Kahneman and Tversky (1984) that Prospect Theory is applicable in this domain. While I also discuss a potential linkage of mechanism for social and tangible outcomes in section 2.3, this remains untested.

The second, is that it is reasonable to dispense with theory of mind in a social decision making context. This is a less plausible, if pragmatic, choice. I have assumed that inference over actions is an acceptable approximation in this case, although as shown by Frey (2014), human behaviour in games is affected by their beliefs about the agency of other players. This also speaks to the gap between models of decision making, and models of behaviour. The former is a relatively simple act of processing on some variant of information about how likely various outcomes are. The steps that lead from perception to the package of information that is decided upon, is a more challenging proposition.

Both these assumptions are an example of simplifications of reality. Such simplifications are necessary, in that for a model to be of use it must be more tractable than the thing to be modelled. Equally necessary however is an awareness of the extent to which such simplifications are appropriate, and the importance of making such assumptions explicit in framing the model. In both cases, what is more implicit is that if the superstructure of the model is plausible, it should remain so were the decision model to be substituted for a more sound, or comprehensive one. As it stands however, the models presented are only sound in the context of the assumptions made, which represents a not inconsiderable caveat, albeit one fundamental to any simulation approach.

6.3 Conclusion

It would be desirable to point to a single decision model as the right choice. Regrettably, this is not possible, because all of the models I have examined are indisputably wrong, but in the right context, useful. There is a pragmatic trade-off that the modeller must consider in the context of their specific modelling problem. This encompasses the degree to which fidelity to real human decisions in outcome, and process, is necessary, balanced against the expense of the model in terms of the data required to parametrise it, and the computational tractability of the simulation. Simple logistic regression performs well in terms of tractability, and with sufficient data can do very well at prediction, with the added advantage that it is a familiar, and well tested technique. In fact, on these metrics, any of a number of machine learning methods will do admirably as the decision rule in an agent-based model. But this is unsatisfying, because these models tell us nothing about the process, and while the techniques are general, a fitted model is not. For example, if, after fitting the logistic regression models used in chapter 5, the game was modified to add a payoff, the model would need to be refitted. In contrast, both the heuristic and inferential models would find this less problematic, even if this occurred during play. In most modelling scenarios the possible decisions are of course fixed in advance, but where this is not the case, such a lack of flexibility is troubling. The former objection, that the decision model should inform us about process, is arguably more salient because this is the key advantage of ABMs over approaches which are for now more capable from a prediction and forecasting perspective. A further concern is the extent to which we can comfortably claim that a fitted prediction model is falsifiable, since the strength of the approach rests on the ability to fit essentially any data rather than making any general predictions about individual behaviour.

The other extreme is occupied by the CPT model, which is the most computationally complex, and has considerable challenges to effective parametrisation which, if not addressed, make it a weak predictor. It is however flexible and can be used in a dynamic environment; it also offers the best performance in predicting individual choices in the coin tossing experiment. In addition, it is strongly opinionated on a theoretical level about how people actually make decisions, which makes it considerably more falsifiable. However, the parametrisation is a salient issue, in part because of the substantial demand this places on the modeller for data, and because in a complicated ABM, it adds substantially to the complexity, and computational expense.

The alternatives are single, or zero parameter models, for example the Bayes or PB models, and the heuristic approaches. These are similarly opinionated on the decision process, and carry a significant advantage in terms of the parsimony of their parametrisation. The models based on Bayesian inference are, as I have discussed in section 2.3, subject to a degree of criticism as to their suitability as models of learning. This concern is scarcely less likely to be levelled at the heuristic approaches, which offer considerably less flexibility in this respect.

The results of the agent-based simulations in chapters 3, and 4 suggest that a model of problem structure is necessary in resolving problems more complex than binary choice. However, the experimental results in chapter 5 would seem to directly contradict this, with decision models which ignore structure performing well in predicting participants' choices. While decision models which incorporate structure perform better, they are not significantly better. In fact, the outcome is more nuanced, since the well performing models do have a notion of structure, albeit one with considerable discrepancies from the real decision problem. More generally, the answer to the question of which decision model is best has been best expressed by Conte (2000), who, in arguing for a necessary level of intelligence in agents, suggests that such models should be 'as simple as suitable'. In this respect, my work here has contributed to the understanding of what a suitable level of simplicity is in the context of the modelling problem, and how to evaluate this.

In fact all four of the models I have discussed here share a common advantage, in that they are falsifiable in isolation. That is to say, because they are not specific to the scenario and model in which they are used, they can be tested in a controlled environment. By comparison, a decision model tied inextricably to the supra-model in which it is applied can only be tested in the context of that model. This argues strongly for the application of general models of decision making improving the validity of agent-based models. Additionally, it suggests that an important consideration in the modelling process is to design with validation at multiple levels in mind, which is a considerable advantage of a game theoretic framing of the model.

Evaluation of ABMs more generally has also been considered in this thesis, with considerable energy dedicated to sensitivity analysis of both the alcohol misuse model (primarily in section 3.5.3, but see also appendix D), and the older adult care model (section 4.6). This is critical not only in prying open what otherwise runs the risk of rapidly becoming a black box to facilitate insight into what aspects of the model drive behaviour, but as shown in the older adult care model, in supporting efforts at calibration. Both are significant, if we bear in mind the purpose of a model from the generative perspective, which is in the most limited sense to prove by the existence of the model that the phenomena observed can be produced by the subset of reality captured in the model. Applying sensitivity analysis with this in mind not only gives us a route to understanding better how significant the components embodied are in generating the phenomena, but also whether they are necessary. That is to say, could we explain as much, with a simpler model. This is not a complete solution, because sensitivity analysis considers the impact of parameters, as distinct from processes. This differentiation is significant, because while sensitivity analysis can identify that a particular parameter has very little impact on the behaviour of the model, it cannot do similarly for the process fed by the parameter. A process might be unnecessarily complex, or implausible yet still have considerable influence on model behaviour, or be entirely redundant, without this being obvious from analysis. This of course highlights the need for careful consideration in constructing the model, but also for validating the parts of a model rather than just the whole.

The other benefit is, from the pragmatic mindset that the modeller must always be within touching distance of, equally important. Calibration of complex models presents substantial challenges, because of the computational demands of simulation, and the necessity of substantial exploration of the parameter space to a sound calibration. On this basis, approaches to sensitivity analysis which substitute a metamodel for the computational model, while placing us at a further remove from reality, pay huge dividends in terms of rapid calibration. While in both papers, Gaussian Processes Emulators have filled this role, there are alternative options which offer similar benefits.

I have advocated for an approach to modelling which begins by creating a game to represent the hypothesis about which aspects of reality are salient to the phenomena, then setting the process in motion using simulation, and calibrating at the macroscale, before examining the validity of the model at other levels. There is a sense in which this might seem to be an inverted approach to the problem, and that modelling should proceed by assembling upwards from parts known to be valid. There is merit to this suggestion, and ideally this would be exactly the case. However, at the present moment there is no collection of parts known to be sound. This means that rather than approaching a linear process backwards, we are merely entering at a particular point in what should properly be framed as a cyclical process. This cycle passes from empirical observation of phenomena, to analysis of the drivers of it, through simulation, and returns to observation by way of experimental work informed by the simulation (Courgeau et al., 2016). At each turn through, the process gains from the previous step as methodological approaches are enhanced, which is fundamentally what I have set out to achieve here.

6.4 Future Work

In part, the question I have set out if not to answer, then to contribute to the answer of, is whether agent-based models can ever be a useful predictive tool. Or failing that, what are agent-based models useful for. Inevitably, I am limited in the conclusions I can draw on this matter. There was a hope that simulation might achieve a measure of freedom from the tyranny of the "beast" of data collection (Silverman et al., 2011). As we have seen, particularly highlighted in chapter 5, there is a very real risk that the need to parametrise simulations leaves 'the beast' insatiable. Whether this is an inescapable limitation to agent-based modelling as a predictive technique remains debatable. However, there is a sense in which the kinds of data demanded for effective agent-based modelling represent an opportunity, rather than a crisis. The atom of the ABM is the agent, and to split it requires a multidisciplinary approach in which demography, and as demonstrated by Chapter five, psychology, should play critical roles.

In terms of the future development of decision models for agents, I have shown that the mental representation of problem structure can be important in understanding how people make decisions. Models which incorporate the process of constructing this mental representation are an interesting future direction. This would also necessitate greater emphasis on the balance between exploration of the problem space, and resolution of the problem, a meta-decision problem which none of the models considered in this thesis address. Such models have the potential to contribute to the theory of decision making by bridging the gap between perception and decision, and also to more powerful agent-based models.

Appendix A

Disclosure Game Model Development

This appendix provides a more in depth exploration of the model development process, beginning by deriving a game to serve as the basis for the model, and decision problems.

A game, in the game theoretic sense, can be any interaction where the result for one person is dependent on the actions of another. In this scenario, the result for the woman would seem dependent on whether the midwife chooses to refer her for specialist support (although naturally the reality can only be thought of in terms of risk mitigation), and conversely, the right choice for the midwife is somewhat contingent on what the woman is willing to tell them.

A very simple way to represent this would be a game with two players, who both have two possible moves - ask for help, or not; and refer, or not (fig A.1). Since both parties are invested in the outcome of the pregnancy, we might allow them to share the same payoff if everything ends well.

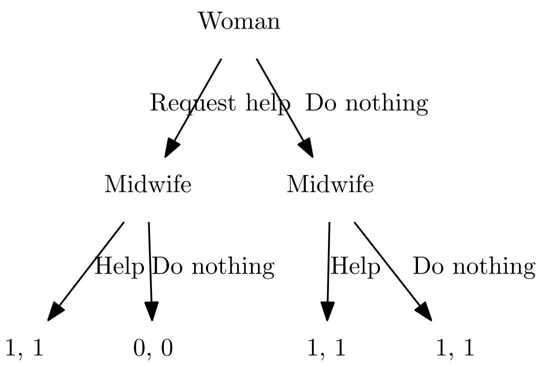


Figure A.1: A very simple two-player game. The only time things in this very restricted world obviously end poorly, is if the woman asks for help but does not get any. This implies that a rational player would always refer if asked for help, and is indifferent otherwise - in other words, there are three possible Nash equilibriums[1].

The first complication, is that there should be differentiation between referring, and doing nothing because specialist treatment incurs a cost. We can modify the payoffs to reflect this, by reducing the midwife's payoffs when they refer. If the cost of referring is less than the value of a good outcome, then the effect of this is to make the only rational choice when not asked for help to do nothing.

This simple game is however not very informative, and clearly neglects much of the nuance of the scenario. The wider difficulty here is that the real outcome depends on an attribute of one of the players, rather than their actions. In this case, we would expect the right choices to depend on the alcohol consumption of the woman, rather than entirely on what she has claimed about it. To reflect this, we would need different variations on the same game to reflect this attribute.

To resolve this, we can do exactly that, and cast it as a signalling game (fig A.2), with three types of player, corresponding to categories of drinking behaviour (light, moderate, and heavy). Each of these types of player, will play a different game. This also introduces a third player, who we will call nature. Nature takes the first move, and decides the type of the woman according to some probability distribution; in this case we will allow the probability of types to be uniform. This changes the dynamics of play substantially, since the midwife can no longer be certain of which game they are playing, and hence which move yields the best outcome. We must also amend the moves, and payoffs slightly. The woman now claims to be one of the types, and may send a signal to say that, for example, she is a heavy drinker. We will also modify the

[1] A Nash equilibrium is a solution to a game between two or more players, where no player can gain from changing their move.

common payoffs to allow light drinkers to get the best outcome no matter what, and moderate and heavy types to get the best outcome only if referred. We can also differentiate between the consequences of not getting help for these types by letting heavy drinkers have a very negative outcome, and moderate drinkers a slight one.

At this point, the game becomes challenging to analyse from a Nash equilibrium perspective (there are several hundred). But, having raised the issue of stigma, we would also like incorporate this in the game. A possible approach to this is similar to the drinking behaviour of the women, and lets midwives have a type as well, corresponding to how judgemental they are when receiving signals: non-judgemental, moderately judgemental, and harshly judgemental. The expression of this judgement is not a matter of choice on their part, and is assumed to have no impact on their clinical response. Nature now has an additional move, to choose the type of the midwife, and we add costs for sending moderate and heavy signals. A heavy signal to a harshly judgemental midwife adds a heavy cost, and a moderate cost from a moderate midwife. The resulting game might reasonably be said to be intractable.

At this juncture, we do not gain much further from the game representation, and instead separate it into multiple decision problems. This can be achieved by treating the moves of the other players as a chance node, and omitting moves by nature that are known to the player. For women, there are two such nodes, corresponding to the move by nature determining the type of midwife they play against, and the midwife's action. Midwives have simpler problem with only a single chance node, because the woman's move is known to them. Figure A.3 shows the structure of the resulting decision problems. Note that there are in fact three distinct decision problems for the three types of woman, since the move by nature determining their type is known to them.

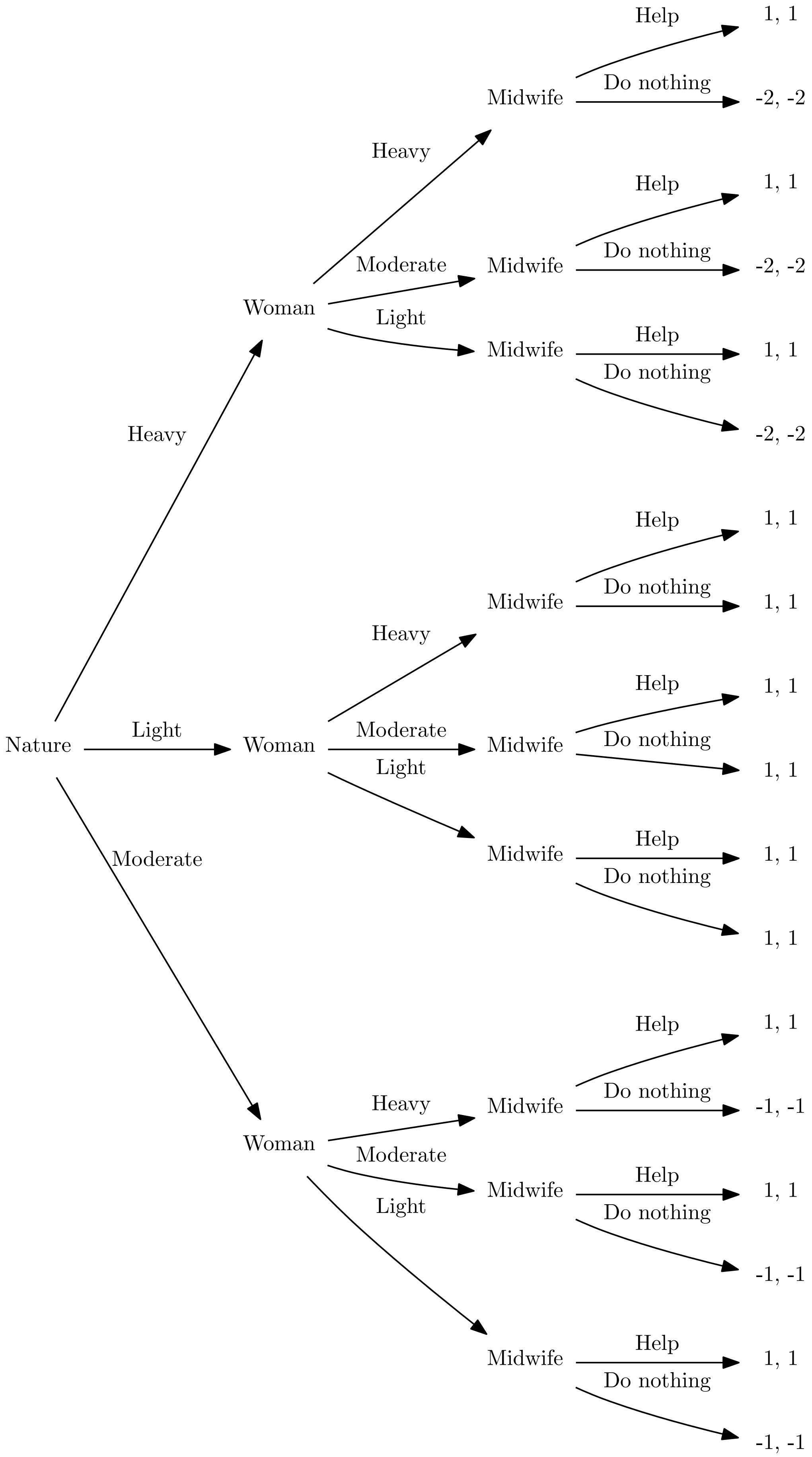
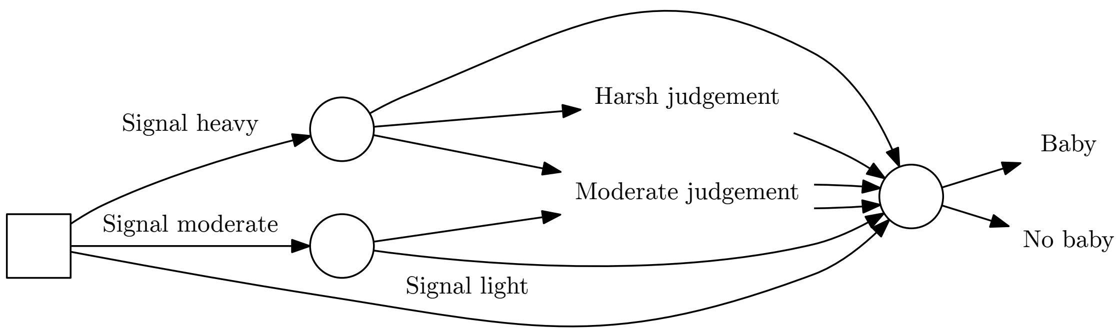
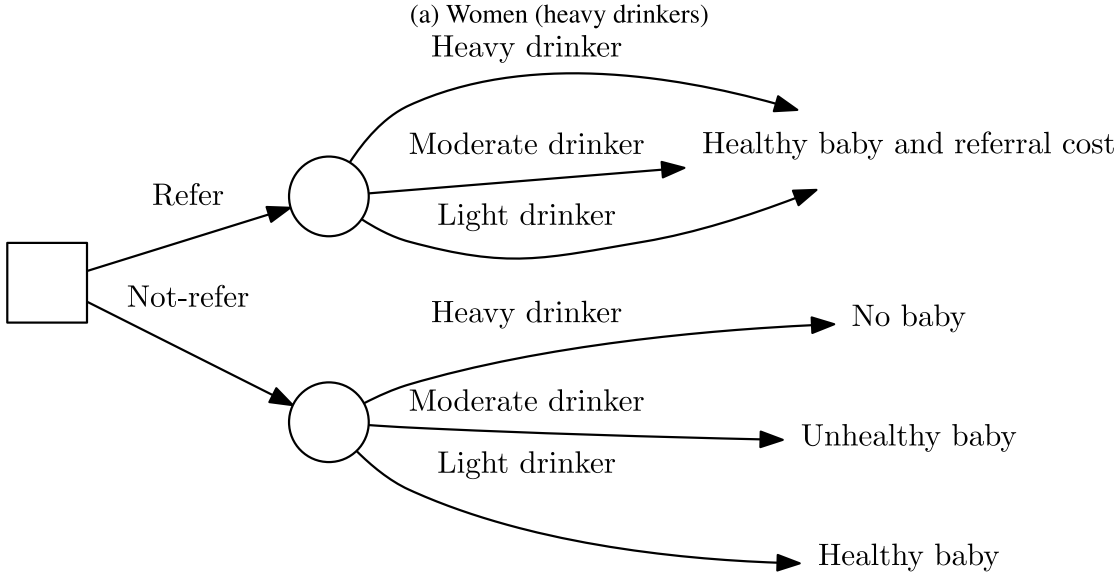


Figure A.2: A less simple two player signalling game.



(a) Women (heavy drinkers)



(b) Midwives

Figure A.3: Influence diagrams, showing the game broken into two decision problems. Squares indicate a decision node, while circles are (from the perspective of the agent) chance nodes

The precise structure of the decision problem is to some extent dependent on the decision rule in use, for example the Lexicographic heuristic rule is concerned only with a direct relationship between action and consequence. However, the literal translation from game to decision problem for women yields two chance nodes. As a result, solving this using the heuristic approach requires that the nodes be combined. By the same token, an arbitrarily complex problem could be resolved by rules without this limitation. This is significant, in that the decision problem is an individual agent's model of the situation, which might not be expected to correspond perfectly with the true sequence of events.

From this position, simulating play, and augmenting the basic conjecture is easily achievable, since together the game, and the decision rules specify the basis for a simulation model. In the disclosure game case, we make additional stipulations on how many games agents play, order of play, the circumstances under which agents observe true types, and the structure of agent populations amongst others.

Appendix B

Disclosure Game Simulation Schedule

This section gives the step by step process for a single run of the disclosure game simulation.

1. Generate 1000 women, and place them in a queue
2. Generate 100 midwives.
3. For each round of the game
   (a) Take 100 women from the queue
   (b) Pair each one with a random midwife
   (c) For each pair
      i. The woman sends a signal
      ii. The midwife refers or not based on the signal
      iii. The woman is informed of her payoff, the midwife's type, and whether she is referred
      iv. The woman updates her beliefs
      v. The midwife stores the game in their memory
      vi. If the woman is referred
         A. The midwife is informed of the woman's true type
         B. The midwife retrospectively updates their beliefs using the true type, and memories of any games with this woman

C. The midwife is now eligible to share their memories of games with this woman

(d) Women who have not been referred or had their baby, join the back of the queue

(e) New women are generated to replace those referred, or delivered

(f) The new women are added to the back of the queue

(g) For each referred or birthed woman

  i. With probability p, her memory of games is shared with the active women

  ii. She is removed from simulation

(h) The active women update their beliefs

(i) For each midwife with information to share

  i. With probability p, their memory of games with the referred woman is shared

  ii. The memory is no longer eligible to be shared

(j) The midwives update their beliefs

Appendix C

Disclosure Game Agent Examples

This section provides a worked example for the learning and decision process of each agent model, focusing on the behaviour of the signalling agent.

C.1 Lexicographic Heuristic

As an example, take a light drinker who has played three rounds with a succession of particularly judgemental midwives, signalling honestly in two and claiming to be a moderate drinker in one. The most common outcome of the honest signal was a payoff of 10, which is clearly preferable to the 9 gained by claiming to be moderate. On that basis, they choose to signal honestly.

C.2 Bayesian Payoff

If we take our light drinker from the lexicographic case, and assume that they began with an uninformative prior the 6 possible signal-payoffs pairings are then $[(l,10),(m,10),(h,10),(m,9),(h,9),(h,8)]$, with $\alpha_i = 1$ for all i. After playing the three rounds, $n_{l,10} = 2$, and $n_{m,9} = 1$.

The agent then evaluates R_w for each signal, e.g. for the light signal:

$$
\begin{aligned}
X &= \{10\} \\
R_w(l) &= \sum_{x \in X} -xp(x|l) = -10p(10|l) \\
R_w(l) &= -10(\frac{\alpha_{l,10} + n_{l,10}}{\sum_j(\alpha_j + n_j)}) = -10(\frac{1+2}{1+2}) \\
R_w(l) &= -10(\frac{3}{3}) = -10
\end{aligned}
$$

and by the same method, $R_w(m) = -9\frac{1}{3}$, and $R_w(h) = -9$, concluding that signalling honestly is the best move.

C.3 Bayesian Risk Minimisation

Returning to our example agent, under this model the type of the midwife becomes salient, hence $n_h = 3$, and $n_{l,n} = 2$, $n_{m,n} = 1$. Their prior beliefs remain uninformative, i.e. $\alpha_j = 1, j \in \{l,m,h\}$, $\alpha_{i,j} = 1, i \in \{r,n\}, j \in \{l,m,h\}$. As before, the agent evaluates R_w for the three signals, and the process for the light signal is given below.

$$
\begin{aligned}
R_w(l,l) &= \sum_{i \in A_m} \sum_{j \in \Theta} -u_w(l,i,l,j)p(j)p(i|l) \\
R_w(l,l) &= -u_w(l,r,l,l)p(l)p(r|l) - u_w(l,n,l,l)p(l)p(n|l) \\
&\quad - u_w(l,r,l,m)p(m)p(r|l) - u_w(l,n,l,m)p(m)p(n|l) \\
&\quad - u_w(l,r,l,h)p(h)p(r|l) - u_w(l,n,l,h)p(h)p(n|l) \\
u_w(l,i,l,j) &= 10 \\
R_w(l,l) &= -10p(l)p(r|l) - 10p(l)p(n|l) - 10p(m)p(r|l) - 10p(m)p(n|l) \\
&\quad - 10p(h)p(r|l) - 10p(h)p(n|l) \\
p(l) &= \frac{1+0}{1+1+1+3} = \frac{1}{6} \\
p(m) &= \frac{1+0}{1+1+1+3} = \frac{1}{6} \\
p(h) &= \frac{1+3}{1+1+1+3} = \frac{2}{3} \\
p(r|l) &= \frac{1+0}{1+1+2} = \frac{1}{4} \\
p(n|l) &= \frac{1+2}{1+1+2} = \frac{3}{4} \\
R_w(l,l) &= -10 \cdot \frac{1}{6} \cdot \frac{1}{4} - 10 \cdot \frac{1}{6} \cdot \frac{3}{4} - 10 \cdot \frac{1}{6} \cdot \frac{1}{4} - 10 \cdot \frac{1}{6} \cdot \frac{3}{4} - 10 \cdot \frac{2}{3} \cdot \frac{1}{4} - 10 \cdot \frac{2}{3} \cdot \frac{3}{4} \\
&= -10
\end{aligned}
$$

and similarly for moderate ($R_w(m,l) = -9\frac{1}{3}$), and heavy ($R_w(h,l) = -8\frac{1}{2}$) signals, once again concluding that honesty is the better option.

C.4 Descriptive Decision Theory

Once again, we return to the light drinker example. The inferential aspects are identical with the more complex Bayesian risk minimisation algorithm, hence $p(j)p(i|l)$, and $u_w(l,i,l,j)$ remain the same, but the agent additionally calculates $v(u_w(l,i,l,j))w^+(p(j))w^+(p(i|l))$. For the CPT parameters, the values are those originally given by Tversky and Kahneman (1992) and used in the actual simulations which are given in table C.1.

Table C.1: CPT parameters.

Name	Description	Value
γ	Probability weighting for gains	0.61
δ	Probability weighting for losses	0.69
α	Power for gains	0.88
β	Power for losses	0.88
λ	Loss aversion	2.25

$$\begin{aligned}
\alpha &= 0.88 \\
\gamma &= 0.61 \\
p(l) &= \frac{1}{6} \\
p(m) &= \frac{1}{6} \\
p(h) &= \frac{2}{3} \\
p(r|l) &= \frac{1}{4} \\
p(n|l) &= \frac{3}{4} \\
u_w(l,i,l,j) &= 10 \\
f^+ &= (10;\frac{1}{24},10;\frac{1}{8},10;\frac{1}{24},10;\frac{1}{8},10;\frac{1}{6},10;\frac{1}{2}) \\
f^+ &= f, f^- \equiv 0 \\
n &= 5 \\
v(u_w) &= f(u_w) = u_w^\alpha \\
v(u_w) &= 10^{0.88} = 7.59 \\
\pi_0^+ &= w^+(\frac{1}{24}+\frac{1}{8}+\frac{1}{24}+\frac{1}{8}+\frac{1}{6}+\frac{1}{2}) - w^+(\frac{1}{8}+\frac{1}{24}+\frac{1}{8}+\frac{1}{6}+\frac{1}{2}) \\
&= w^+(1) - w^+(\frac{23}{24}) \\
&= 0.19 \\
\pi_1^+ &= w^+(\frac{1}{8}+\frac{1}{24}+\frac{1}{8}+\frac{1}{6}+\frac{1}{2}) - w^+(\frac{1}{24}+\frac{1}{8}+\frac{1}{6}+\frac{1}{2}) \\
&= w^+(\frac{23}{24}) - w^+(\frac{5}{6})
\end{aligned}$$

$$
\begin{aligned}
&= 0.17 \\
\pi_2^+ &= w^+(\frac{1}{24}+\frac{1}{8}+\frac{1}{6}+\frac{1}{2}) - w^+(\frac{1}{8}+\frac{1}{6}+\frac{1}{2}) = w^+(\frac{5}{6}) - w^+(\frac{19}{24}) \\
&= 0.04 \\
\pi_3^+ &= w^+(\frac{1}{8}+\frac{1}{6}+\frac{1}{2}) - w^+(\frac{1}{6}+\frac{1}{2}) = w^+(\frac{19}{24}) - w^+(\frac{2}{3}) \\
&= 0.09 \\
\pi_4^+ &= w^+(\frac{1}{6}+\frac{1}{2}) - w^+(\frac{1}{2}) = w^+(\frac{2}{3}) - w^+(\frac{1}{2}) \\
&= 0.09 \\
\pi_5^+ &= w^+(\frac{1}{2}) \\
&= 0.42 \\
V(f) &= V(f^+) + V(f^-) = V(f^+) + 0 \\
V(f^+) &= \sum_i^n \pi_i^+(f^+) v_i^+(f^+) = 7.59
\end{aligned}
$$

And as before, following the same process for moderate, and heavy signals which yields respectively 7.14, and 6.22, the agent chooses the higher valued action and sends an honest signal.

Appendix D

Disclosure Game Model Sensitivity Analysis

This section provides complete variance based sensitivity analysis results for the disclosure game model. Each subsection gives results for one simulation output under all four decision rules, with tables providing the percentage of overall variance attributable to the individual parameters, emulator quality statistics, and the five most important interaction contributions to variance in the output.

D.1 Median Moderate Drinker Signalling

Table D.1: Median moderate drinker signalling parameter sensitivity

Parameter	Description	Lexicographic	Bayesian Payoff	Bayesian	CPT
$p_w(m)$	Proportion of moderate drinkers	0.367	1.145	0.801	0.614
$p_w(l)$	Proportion of light drinkers	10.080	37.750	23.968	5.137
$p_m(m)$	Proportion of moderate midwives	6.715	13.017	0.894	1.485
$p_m(l)$	Proportion of non-judgemental midwives	43.942	1.655	1.602	2.618
q_w	Probability of women sharing	0.198	5.527	4.460	1.159
w_w	Weight of shared information for women	0.355	13.025	2.716	0.888
q_m	Probability of midwives sharing	0.145	0.667	0.368	0.157
w_m	Weight of shared information for midwives	0.118	0.376	0.176	0.200
x_h	Health payoff for healthy delivery	0.457	9.618	1.912	15.355
$s_i[a_i]:s_i[a_{\neg i}]$	Pseudo-count favouring honesty	0.140	7.537	10.427	7.795
Total	All parameters and two way interactions	86.777	96.527	85.529	74.123

Table D.2: Median moderate drinker signalling emulator statistics

Rule	σ^2	Nugget σ^2	μ	Total output variance	Code uncertainty	RMSSE
Lexicographic	0.834	0.131	0.817	0.012	0.252	1.746
Bayesian Payoff	1.667	0.475	0.662	0.003	0.181	3.12
Bayesian	3.352	0.534	1.160	0.001	0.068	2.423
CPT	1.503	0.331	1.241	0.002	0.101	1.842

Table D.3: Top five interaction terms for median moderate drinker signalling

Parameter	Variance			
	CPT	Bayesian	Lexicographic	PB
$x_h{*}s_i[a_i] : s_i[a_{\neg i}]$	20.814			1.856
$p_w(l){*}s_i[a_i] : s_i[a_{\neg i}]$	5.698	17.270		
$p_w(l){*}x_h$	2.895			
$s_i[a_i] : s_i[a_{\neg i}]{*}w_w$	2.799	3.084		
$s_i[a_i] : s_i[a_{\neg i}]{*}q_m$	1.686	3.538		0.997
$p_w(l){*}q_m$		6.054		1.231
$q_m{*}w_w$		3.814		0.929
$p_m(l){*}p_m(m)$			15.331	
$p_m(m){*}p_w(l)$			3.682	
$p_m(l){*}p_w(l)$			3.581	
$p_m(m){*}q_m$			0.349	
$p_m(l){*}q_m$			0.279	
$p_w(l){*}w_w$				4.045

D.2 Median Between Groups IQR

Table D.4: Median between groups IQR parameter sensitivity

Parameter	Description	Lexicographic	Bayesian Payoff	Bayesian	CPT
$p_w(m)$	Proportion of moderate drinkers	0.327	0.688	0.457	0.586
$p_w(l)$	Proportion of light drinkers	11.223	20.123	11.046	4.081
$p_m(m)$	Proportion of moderate midwives	36.630	1.160	0.364	1.945
$p_m(l)$	Proportion of non-judgemental midwives	6.228	4.487	0.0964	2.627
q_w	Probability of women sharing	0.498	0.235	2.537	1.812
w_w	Weight of shared information for women	1.018	2.307	1.889	0.740
q_m	Probability of midwives sharing	0.158	0.343	0.387	0.156
w_m	Weight of shared information for midwives	0.076	0.973	0.125	0.213
x_h	Health payoff for healthy delivery	0.317	10.960	3.305	16.493
$s_i[a_i] : s_i[a_{\neg i}]$	Pseudo-count favouring honesty	1.107	8.411	2.890	6.729
Total	All parameters and two way interactions	81.702	83.693	47.449	71.032

Table D.5: Median between groups IQR emulator statistics

Rule	σ^2	Nugget σ^2	μ	Total output variance	Code uncertainty	RMSSE
Lexicographic	0.930	0.240	0.249	0.002	0.040	1.832
Bayesian Payoff	1.242	0.417	0.232	0.001	0.0034	2.308
Bayesian	1.254	0.131	0.644	0.000	0.019	1.167
CPT	1.190	0.313	0.659	0.000	0.024	1.701

Table D.6: Top five interaction terms for median between groups IQR

Parameter	Variance			
	CPT	Bayesian	Lexicographic	PB
$x_h * s_i[a_i] : s_i[a_{\neg i}]$	19.551	2.680		2.360
$p_w(l) * s_i[a_i] : s_i[a_{\neg i}]$	3.838	2.943		12.883
$s_i[a_i] : s_i[a_{\neg i}] * w_w$	2.450			
$p_w(l) * x_h$	2.337			2.447
$s_i[a_i] : s_i[a_{\neg i}] * q_m$	2.046	4.284		
$p_w(l) * q_m$		3.866		
$q_m * w_w$		2.282		
$p_m(l) * p_m(m)$			12.046	
$p_m(m) * p_w(l)$			5.054	1.919
$p_m(l) * p_w(l)$			3.005	
$p_m(l) * w_w$			0.819	
$p_m(m) * w_w$			0.757	
$p_w(l) * w_w$				5.667

D.3 Median Moderate Drinker Signalling IQR

Table D.7: IQR of median moderate drinker signalling parameter sensitivity

Parameter	Description	Lexicographic	Bayesian Payoff	Bayesian	CPT
$p_w(m)$	Proportion of moderate drinkers	0.428	9.828	3.816	2.068
$p_w(l)$	Proportion of light drinkers	8.369	13.791	4.045	2.400
$p_m(m)$	Proportion of moderate midwives	13.416	0.712	0.676	0.583
$p_m(l)$	Proportion of non-judgemental midwives	21.079	0.648	0.659	0.373
q_w	Probability of women sharing	2.307	3.481	0.891	0.600
w_w	Weight of shared information for women	6.021	6.009	0.562	0.937
q_m	Probability of midwives sharing	0.315	1.829	0.114	0.117
w_m	Weight of shared information for midwives	1.652	1.354	0.260	0.0.139
x_h	Health payoff for healthy delivery	0.253	0.612	4.889	15.146
$s_i[a_i]:s_i[a_{\neg i}]$	Pseudo-count favouring honesty	0.504	3.096	19.863	25.999
Total	All parameters and two way interactions	84.9968	77.413	57.125	83.322

Table D.8: IQR of median between groups IQR emulator statistics

Rule	σ^2	Nugget σ^2	μ	Total output variance	Code uncertainty	RMSSE
Lexicographic	1.425	0.436	0.549	0.008	0.114	2.719
Bayesian Payoff	1.223	0.496	0.747	0.012	0.207	2.034
Bayesian	1.065	0.000	0.230	0.002	0.088	1.015
CPT	0.874	0.213	0.233	0.001	0.066	1.806

Table D.9: Top five interaction terms for IQR of median moderate drinker signalling.

Parameter	Variance			
	CPT	Bayesian	Lexicographic	PB
$x_h{*}s_i[a_i] : s_i[a_{\neg i}]$	17.377	3.120		
$p_w(m){*}s_i[a_i] : s_i[a_{\neg i}]$	3.356	4.188		
$s_i[a_i] : s_i[a_{\neg i}]{*}w_w$	3.036			
$s_i[a_i] : s_i[a_{\neg i}]{*}q_m$	2.067	2.423		
$p_w(l){*}s_i[a_i] : s_i[a_{\neg i}]$	1.721	2.279		4.882
$p_w(l){*}q_m$		1.489		2.346
$p_m(m){*}p_w(l)$			12.068	
$p_m(l){*}p_w(l)$			6.794	
$p_m(l){*}p_m(m)$			5.567	
$p_m(m){*}q_m$			0.886	
$p_m(l){*}q_m$			0.692	
$p_w(l){*}p_w(m)$				8.357
$p_w(l){*}w_w$				7.431
$p_w(l){*}w_m$				2.025

D.4 Between Groups IQR IQR

Table D.10: IQR of median between groups IQR parameter sensitivity

Parameter	Description	Lexicographic	Bayesian Payoff	Bayesian	CPT
$p_w(m)$	Proportion of moderate drinkers	0.691	5.926	1.053	1.265
$p_w(l)$	Proportion of light drinkers	3.664	17.047	4.877	3.656
$p_m(m)$	Proportion of moderate midwives	41.369	1.124	0.814	0.591
$p_m(l)$	Proportion of non-judgemental midwives	7.109	0.739	0.496	0.378
q_w	Probability of women sharing	1.963	2.038	0.733	0.589
w_w	Weight of shared information for women	7.932	11.193	2.289	1.960
q_m	Probability of midwives sharing	0.413	1.972	0.267	0.069
w_m	Weight of shared information for midwives	0.120	2.902	0.150	0.123
x_h	Health payoff for healthy delivery	0.228	3.190	6.308	14.777
$s_i[a_i] : s_i[a_{\neg i}]$	Pseudo-count favouring honesty	0.673	10.411	22.901	26.340
Total	All parameters and two way interactions	85.740	88.611	68.640	84.210

Table D.11: IQR of median between groups IQR emulator statistics

Rule	σ^2	Nugget σ^2	μ	Total output variance	Code uncertainty	RMSSE
Lexicographic	0.826	0.409	0.259	0.002	0.034	2.364
Bayesian Payoff	3.202	0.520	0.328	0.002	0.032	2.452
Bayesian	1.177	0.041	0.133	0.000	0.018	1.152
CPT	0.874	0.118	0.126	0.000	0.017	1.570

Table D.12: Top five interaction terms for between groups IQR IQR

Parameter	Variance			
	CPT	Bayesian	Lexicographic	PB
$x_h * s_i[a_i] : s_i[a_{\neg i}]$	18.626	4.048		2.294
$p_w(l) * s_i[a_i] : s_i[a_{\neg i}]$	3.312	7.947		3.401
$s_i[a_i] : s_i[a_{\neg i}] * w_w$	2.823			
$s_i[a_i] : s_i[a_{\neg i}] * q_m$	2.694	3.134		
$x_h * q_m$	1.022			
$p_w(l) * q_m$		2.307		2.385
$p_m(m) * s_i[a_i] : s_i[a_{\neg i}]$		2.232		
$p_m(l) * p_m(m)$			8.659	
$p_m(m) * p_w(l)$			3.726	
$p_m(l) * p_w(l)$			3.237	
$p_m(l) * w_w$			1.564	
$p_m(m) * w_w$			1.558	
$q_m * w_w$				3.830
$p_w(l) * p_w(m)$				3.808

Appendix E

Older Adult Model Simulation Schedule

This section gives the step by step process for a single run of the older adult care simulation.

1. Generate 1000 older adults, with need status drawn from the empirical distribution and place them in a queue
2. Generate 100 benevolent care providers
3. Initialise care providers with beliefs drawn from the empirical social trust distribution
4. Initialise older adults with beliefs drawn from the empirical help provision, and stigma distributions
5. For each round of the game
    (a) Take 100 older adults from the queue
    (b) Pair each one with a random care provider
    (c) For each pair
        i. The older adult sends a signal
        ii. The care provider helps or not based on the signal
        iii. The older adult is observes the care provider's type, and whether they are helped

iv. The older adult updates her beliefs

v. The care provider stores the game in their memory

vi. If the older adult is helped

A. The care provider is informed of the older adult's true need status

B. The care provider retrospectively updates their beliefs using the true type, and memories of any games with this older adult

C. The care provider forgets any existing memory of giving help

D. The care provider is now eligible to share their memories of games with this older adult

(d) Older adults who have not been helped or left the game, join the back of the queue

(e) New older adults are generated with need status drawn from the empirical distribution, to replace those helped, or departed

(f) Initialise the new older adults with beliefs drawn from the empirical help provision, and stigma distributions

(g) The new older adults are added to the back of the queue

(h) For each helped or departed older adult

i. With probability p, their memory of games is shared with the active older adults

ii. They are removed from simulation

(i) The active older adults update their beliefs

(j) For each care provider with information to share

i. With probability p, their memory of games with the helped older adult is shared

ii. The memory is no longer eligible to be shared

(k) The care providers update their beliefs

Appendix F

Older Adult Care Model Sensitivity Analysis

This section provides complete variance based sensitivity analysis results for the older adult care model. Each subsection gives results for one simulation output under all four decision rules, with tables providing the percentage of overall variance attributable to the individual parameters, emulator quality statistics, and the five most important interaction contributions to variance in the output.

F.1 Mutual Information

Table F.1: Older adult signalling mutual information parameter sensitivity

Parameter	Description	Lexicographic	Bayesian Payoff	Bayesian	CPT
G	Payoff from good outcome	19.762	0.311	0.622	5.138
B	Cost of a bad outcome	15.120	0.061	1.205	0.653
C	Cost of stigma	6.715	12.940	25.788	20.290
P_d	Probability of decider sharing	0.037	0.146	1.602	0.124
W_d	Weight of shared information for deciders	0.089	0.041	0.237	0.055
Q_d	Weight of priors for deciders	0.012	0.063	0.033	0.140
P_a	Probability of asker sharing	7.674	0.396	4.905	0.124
W_a	Weight of shared information for asker	0.085	0.786	4.305	3.008
Q_a	Weight of priors for askers	0.032	6.849	1.280	11.967
Total	All parameters and two way interactions	79.738	62.475	73.275	68.215

Table F.2: Older adult signalling mutual information emulator statistics

Rule	σ^2	μ	Code uncertainty	Total output variance	RMSSE
Lexicographic	0.731	0.606	0.000	0.086	1.186
Bayesian Payoff	1.541	0.017	0.000	0.010	1.076
Bayesian	0.857	0.243	0.000	0.176	1.137
CPT	1.022	0.148	0.000	0.112	1.098

Table F.3: Top five interaction terms for older adult signalling mutual information

Parameter	Variance			
	CPT	Bayesian	Lexicographic	PB
G*C	5.932		6.980	
G*Q_a	3.342	1.625		
G*B	2.689	3.503	8.665	
C*Q_a	2.564	2.075		27.641
C*W_a	1.465	1.462		3.807
P_a*Q_a		1.226		1.224
B*C			5.045	
C*P_a			1.042	1.330
G*P_a			0.0456	
P_a*W_a				1.020

F.2 Referral Proportions

F.2.1 Healthy Older Adults

Table F.4: Parameter sensitivity for health older adult referral proportion

Parameter	Description	Lexicographic	Bayesian Payoff	Bayesian	CPT
G	Payoff from good outcome	30.553	38.189	35.231	66.605
B	Cost of a bad outcome	28.392	39.179	34.919	15.682
C	Cost of stigma	0.007	0.001	0.044	0.003
P_d	Probability of decider sharing	0.009	0.002	0.017	0.004
W_d	Weight of shared information for deciders	0.008	0.001	0.042	0.013
Q_d	Weight of priors for deciders	0.075	0.021	3.484	0.754
P_a	Probability of asker sharing	0.037	0.002	0.006	0.003
W_a	Weight of shared information for asker	0.019	0.002	0.014	0.010
Q_a	Weight of priors for askers	0.009	0.004	0.085	0.052
Total	All parameters and two way interactions	97.101	99.312	95.883	97.490

Table F.5: Healthy older adult referral proportion emulator statistics

Rule	σ^2	μ	Code uncertainty	Total output variance	RMSSE
Lexicographic	1.014	0.980	0.000	0.473	1.456
Bayesian Payoff	0.096	0.557	0.000	0.395	1.072
Bayesian	0.233	0.149	0.000	0.049	1.158
CPT	0.151	0.359	0.000	0.293	1.019

Table F.6: Top five interaction terms for healthy older adult referral proportion

Parameter	Variance			
	CPT	Bayesian	Lexicographic	PB
$G*B$	17.103	16.406	37.058	21.750
$G*Q_d$	0.727	2.592	0.271	0.047
$G*Q_a$	0.141			0.027
$B*Q_d$	0.133	2.281	0.220	0.045
$B*Q_a$	0.048	0.124		0.222
$B*W_d$		0.118		
$B*W_a$			0.109	
$B*P_a$			0.106	

F.2.2 In-need Older Adults

Table F.7: Parameter sensitivity for in-need older adult referral proportion

Parameter	Description	Lexicographic	Bayesian Payoff	Bayesian	CPT
G	Payoff from good outcome	30.592	38.192	28.074	57.227
B	Cost of a bad outcome	28.353	39.154	31.810	17.499
C	Cost of stigma	0.007	0.001	3.367	0.767
P_d	Probability of decider sharing	0.009	0.001	0.016	0.009
W_d	Weight of shared information for deciders	0.008	0.001	0.008	0.047
Q_d	Weight of priors for deciders	0.081	0.021	0.206	0.322
P_a	Probability of asker sharing	0.035	0.002	1.641	0.220
W_a	Weight of shared information for asker	0.018	0.002	0.802	0.012
Q_a	Weight of priors for askers	0.009	0.004	3.732	0.322
Total	All parameters and two way interactions	97.086	99.302	91.797	94.339

Table F.8: In need older adult referral proportion emulator statistics

Rule	σ^2	μ	Code uncertainty	Total output variance	RMSSE
Lexicographic	1.015	0.978	0.000	0.472	1.458
Bayesian Payoff	0.0100	0.557	0.000	0.395	1.090
Bayesian	0.595	0.320	0.000	0.215	1.193
CPT	0.446	0.429	0.000	0.356	1.195

Table F.9: Top five interaction terms for in-need older adult referral proportion

Parameter	Variance			
	CPT	Bayesian	Lexicographic	PB
G*B	15.332	11.551	37.039	21.760
G*C	0.55	1.142		
G*Q_d	0.356		0.272	0.048
G*Q_a	0.335	1.689		0.026
B*Q_a	0.239	1.728	0.219	0.022
B*C		2.345		
B*W_a			0.108	
B*P_a			0.106	
B*Q_d				0.456

Appendix G

Paired Gambles Experimental Protocol

This appendix contains the experimental protocol for the three coin experiment, as approved by the University of Southampton ethics board.

UNIVERSITY OF Southampton

Protocol

Study Title:
Experimental Study of Decision Making from Experience

Researcher(s)
Jonathan Gray
Funder: University of Southampton
Sponsor (if known):

Background

This study is intended to investigate economic decision making from experience, with paired gambles. We aim to test whether participants reason, and learn, about individual gambles within a pair, or combine them together, hypothesising that they will treat them as separate. The secondary aim, is to test which, if any, of four models of decision making used in an agent-based model best predict human decision making, thereby validating the individual level of the simulation.

Method

The study will take the form of twenty short games, administered using the ZTree software. In each game, participants are presented with three biased coins (A, B, and C), which have payoffs associated for heads and tails. They are then given the choice of flipping A and B, or B and C. Coins A and C have different biases, but the same payoffs for heads and tails. Coin B has a payoff for heads, but only if flipped with coin A. Participants must sample the outcomes in each game ten times, before making a final choice.

To minimise order effects, games will be presented in a random order, and participants will be told the expected duration of the experiment, rather than the exact number of games to avoid changes in risk attitudes associated with any 'end-of-task' effect. The games will be preceded by a short quiz to ensure that participants understand the game, and followed by a brief questionnaire to assess the demographics of the participants (gender, and age).

The questions have been developed to maximise the probability that in two cases each decision rule takes a minority position, such that under a large number of possible sampling sequences, they would predict that participants should choose one pair where the other rules predict the opposite.

The individual sampling sequences will be used to parameterise an instance of each decision rule, which will then make a prediction. The probability of observing the final choices of the participants, given each model will then be used to perform Bayesian model selection. A Bayesian approach is appropriate here, because of the requirement to quantify support for a particular model, rather than test a null hypothesis.

UNIVERSITY OF Southampton

Materials

An experimental design has been developed using the ZTree software (attached separately as three_coins.ztt), which administers the questions in a random order for each participant. The full list of decision problems is given in **Table 1**, with the true expected value for choosing A and B, or B and C in each problem shown in **Table 2**. Sample screenshots of the screens participants will encounter are provided in **A3: Sample Decision Problem Screens**
3. In addition to the participant information sheet included with this submission, participants will also be issued with instructions (**Error! Reference source not found.**), which will also be read aloud to them.
Participants will also be asked to complete a short comprehension test, to ensure that they have understood the instructions, which is shown in **A2:** Comprehension Test (the ZTree file is also included with this submission as comprehension.ztt).

Participants

50-100 members of the University of Southampton who have signed up to participate in experiments in the Social Sciences Experimental lab (http://econexp.soton.ac.uk), although the exact number of participants will depend on recruitment success.

Procedure

Each experimental session will take approximately 90 minutes.

1. Participants arrive
2. Participants identity is confirmed (5 minutes)
3. Participants are issued a random identity number (1 minute)
4. Participants are issued information sheet, instructions, and consent forms (1 minute)
5. Instructions are read to participants using text-to-speech software (5 minutes)
6. Participants complete consent forms, or withdraw (2 minutes)
7. After all consent forms are collected, experiment begins
8. Participants complete comprehension test (5 minutes)
9. Computer randomises the order of games for each participant
10. Computer selects, at random, one question for which the participant will receive payment in cash at the end of the experiment
11. Computer screen displays the payoffs attached to the three coins
12. Participants choose to observe the outcome of flipping A and B, or B and C, by clicking a button
13. After all participants have chosen, the computer displays the outcome (heads, or tails) for the two coins, the payoff associated with each coin, and the total payoff for that round.
14. Participants repeat 11, 12, 13 for ten rounds. (8 minutes)
15. Participants make final choice for this game
16. Participants repeat 11, 12, 13, 14, 15 for remaining 7 questions (56 minutes)
17. Participants complete short survey (2 minutes)
18. Participants are called individually by number to collect payment, and sign a receipt to confirm that they have been paid (10 minutes)
19. Experiment ends

UNIVERSITY OF Southampton

Statistical analysis

The primary method of analysis will be Bayesian model selection using the WinBUGS software package, which will allow us to quantify how far the balance of probability favours each of the four models individually, and the extent to which the evidence supports the use of a mental model of the problem.

UNIVERSITY OF Southampton

Ethical issues

In terms of experimental design, the notable ethical issue is that participants' payment will be partially dependent on the responses they give to the decision problems. This may influence their perceived freedom to withdraw from the experiment, and will be managed by making payment partially in the form of a 'show up fee', which is not dependent on completing the experiment.

Because in some games, participants may face a choice between only negative outcomes, in the event that a participant's payment from the randomly selected decision problem would be negative, they will instead only receive the fee for showing up.

Data protection and anonymity

Individual sampling sequences, and decision responses will be coded as anonymous, and not associated to the demographic data. Questionnaire responses will be analysed and reported only in aggregate, and the individual responses will not be made publically available.

Our intention is to make the anonymised individual responses, and sampling sequences publically available from the University of Southampton ePrints repository, to facilitate further research on the topic, and subsequent analysis by other researchers. Explicit consent will be obtained from participants to do so, and they may decline to make their data publically available while otherwise participating. Because the responses are anonymised, and contain no information beyond binary choice responses, the risk of identification is extremely low, and does not raise significant data protection issues.

UNIVERSITY OF Southampton

Appendix

A1: Decision Problems Listing

Table 1: Decision problems

	Payoffs			Probabilities		
	A and C		B			
	Heads	Tails	Heads	P(H|A)	P(H|B)	P(H|C)
1	-67	-64	58	0.990	0.040	0.010
2	4	-13	65	0.010	0.010	0.997
3	57	18	-20	0.004	0.996	0.020
4	-80	29	-87	0.001	0.999	0.997
5	-96	-95	67	0.850	0.100	0.100
6	-21	-17	1	0.102	0.500	0.103
7	95	22	-52	0.880	0.530	0.100
8	-76	-64	10	1.000	1.000	0.002

Table 2: True expected value of choosing A and B, or B and C for the eight decision problems.

	A and B	B and C
1	-64.65	-64.03
2	-12.18	3.95
3	36.69	18.78
4	-58.13	-79.67
5	-89.15	-95.10
6	-16.91	-17.41
7	58.68	29.30
8	-66.00	-64.02

A2: Comprehension Test

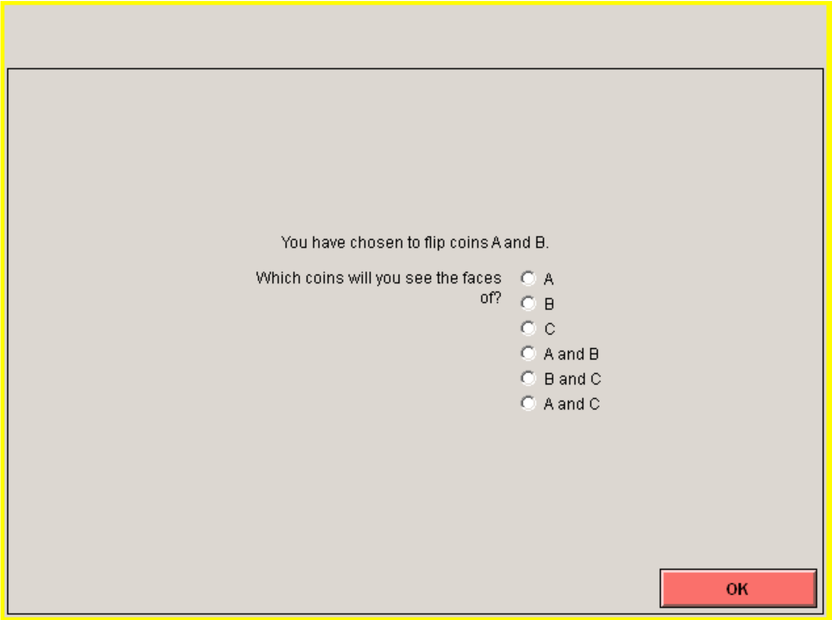


Figure 1: First comprehension test screen, checking that participants understand which coin faces will be revealed.

UNIVERSITY OF Southampton

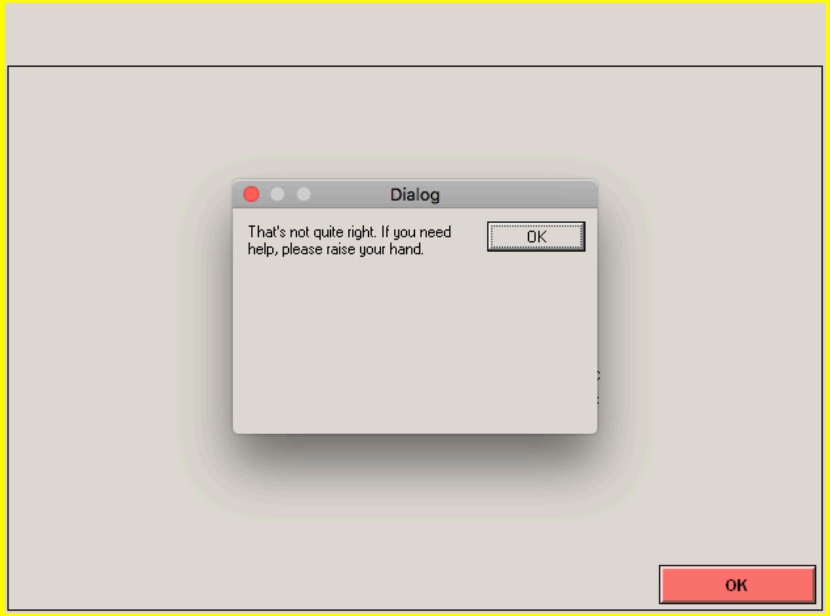


Figure 2: Alert displayed if the participant selects and incorrect answer, prompting them to try again, or to request support if they need it.

UNIVERSITY OF Southampton

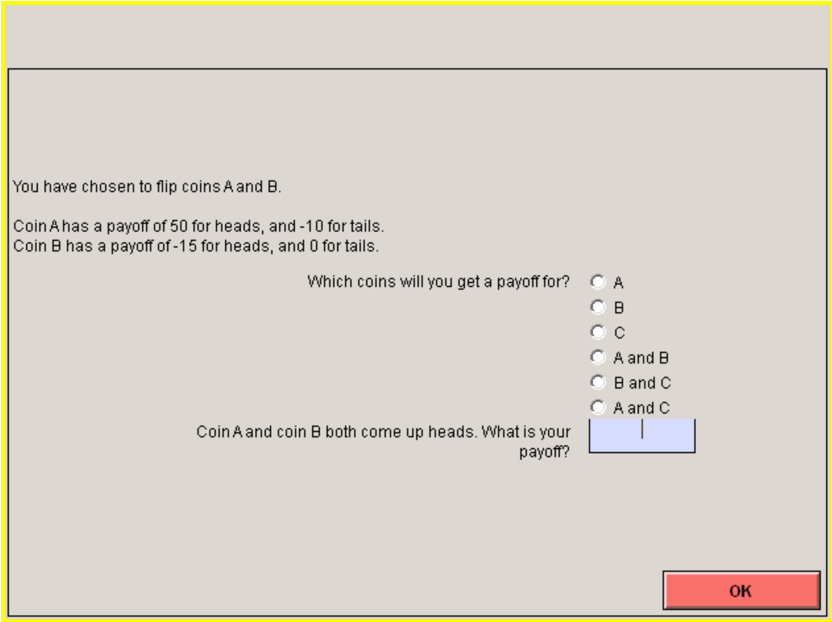


Figure 3: Second comprehension question, checking that participants understand that they receive the combined payoff if they flip coins A and B.

UNIVERSITY OF Southampton

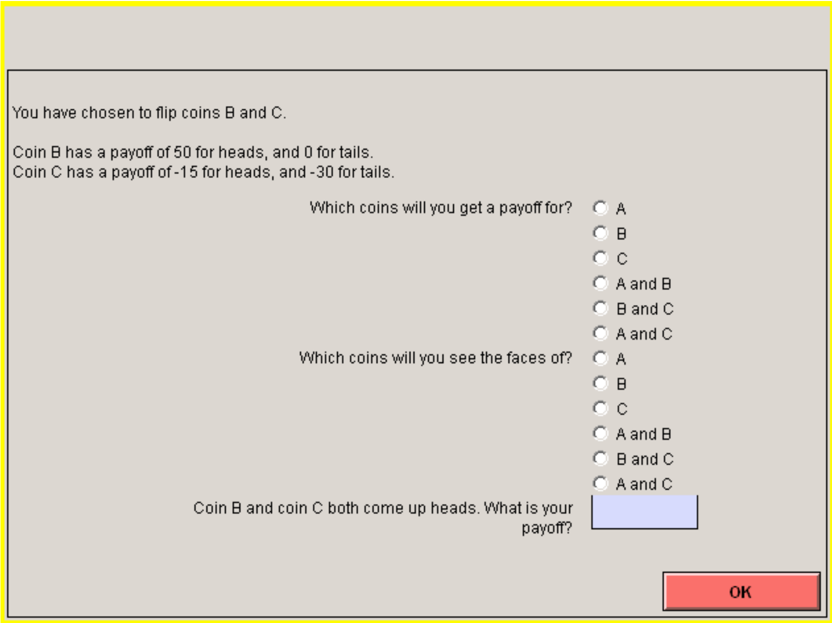


Figure 4: Final comprehension question, checking overall understanding, as well as reminding participants that they receive only the payoff from coin C when flipping B and C.

UNIVERSITY OF Southampton

A3: Sample Decision Problem Screens

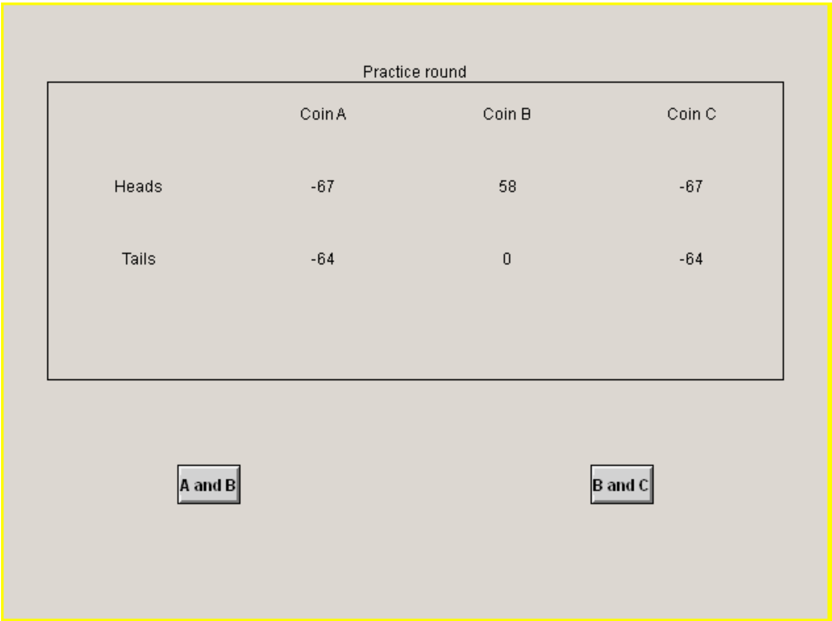


Figure 5: Decision problem sampling screen, showing the possible payoffs for the three coins individually.

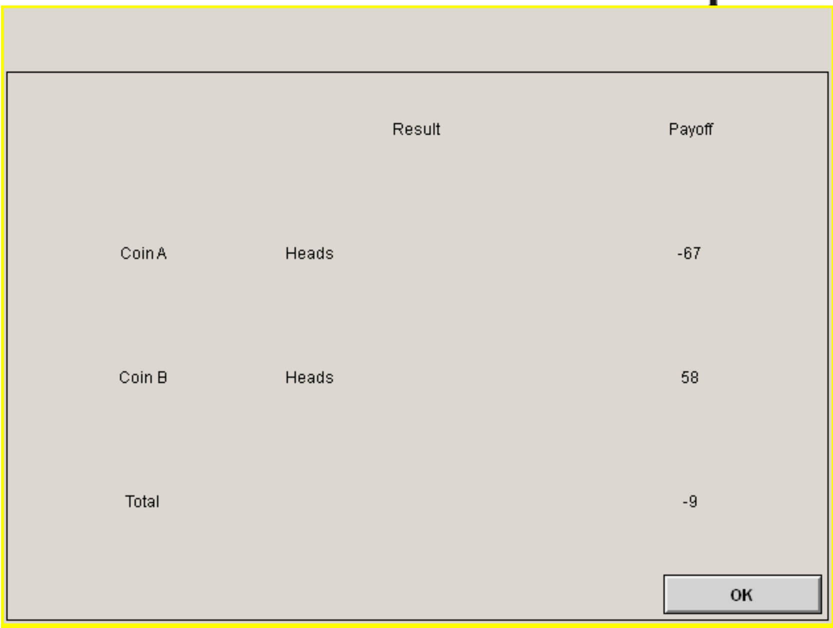


Figure 6: Choice result screen, showing the outcome after choosing to flip coins A and B. The screen shows both individual, and combined payoffs for the two coins, as well as the faces for both.

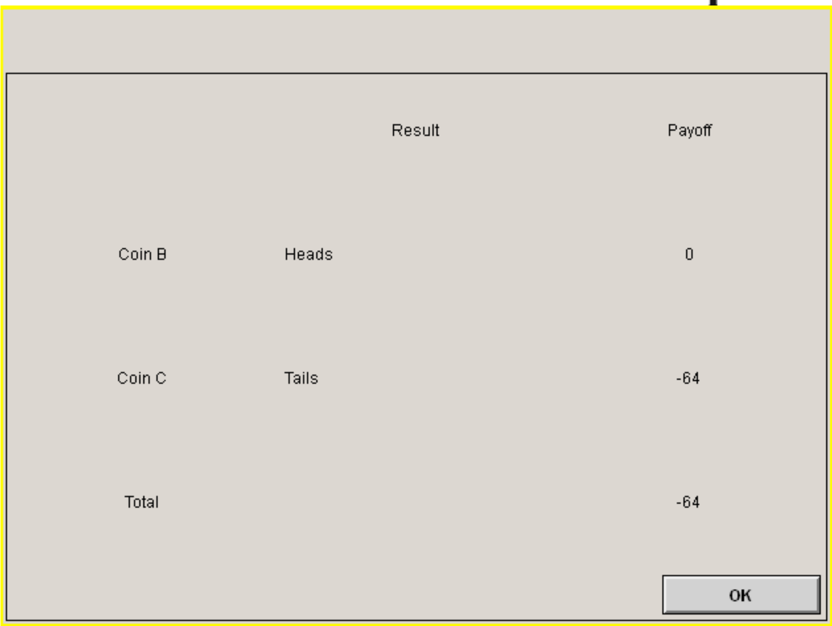


Figure 7: Decision problem results screen, for the B and C choice. The screen shows the faces for both coins, but only displays the payoff for coin C.

UNIVERSITY OF Southampton

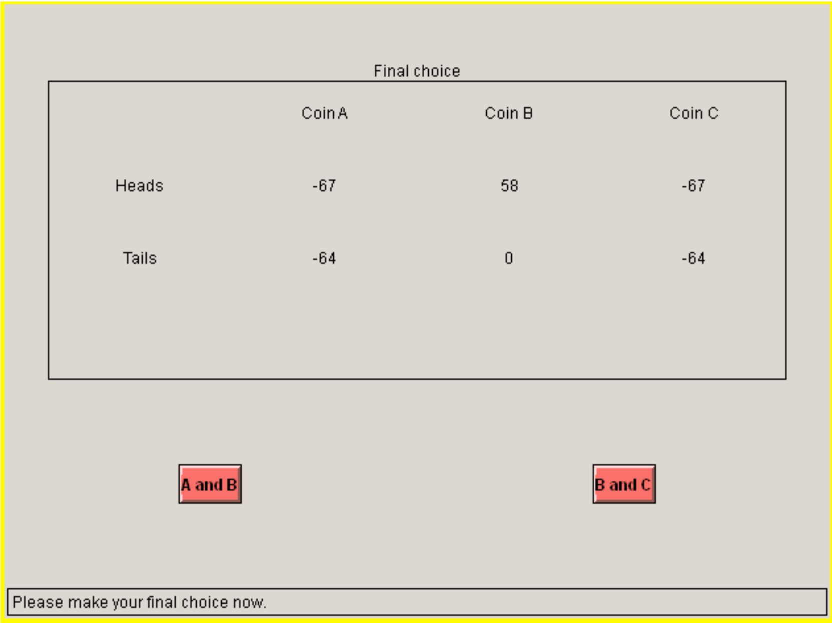


Figure 8: Final choice screen for a decision problem, shown after the 10 sampling rounds are completed.

Figure 9: Questionnaire screen.

Figure 10: Final results screen, showing the payment the subject will receive based on their question answers, the show up fee, and the total of the two.

UNIVERSITY OF Southampton

A4: Instructions Sheet

Instructions

Today, you will be playing several games. In every game, there are three coins – A, B, and C. The three coins are unfair coins, so each coin is more or less likely to come up heads. Coins A, B, and C all have different biases from one another, and the bias of the coins will change for each new game.

Each coin has a payoff for heads, and a payoff for tails, in Experimental Monetary Units, or EMUs. After the experiment, EMUs can be exchanged for pounds at a rate of 1 EMU to ten pence, so 10 EMU are worth 1 pound. Just like the bias of the coins, the amount of EMU for the faces of each coin will change in every game.

In a game, you can choose whether you want to to flip TWO of the coins, at the same time. You can either flip coins A and B, or B and C. If you flip coins A and B, you will get the payoff from coin A, and coin B. BUT, if you flip coin B, and coin C, you will only get the payoff from coin C. After you flip a pair of coins, you will be shown whether they came up heads or tails, what the payoff from the individual coins was, and what the total payoff from the coins was.

In each game, you will have TEN practice rounds, where you can try flipping coins A and B, or B and C to see what happens. After the practice rounds, you will then make a final choice about which pair of coins to flip, and be shown what the outcome was.

After you play all the games, you will be asked to complete a short questionnaire. Finally, ONE of the outcomes of your final choices will be chosen at random, and you will be paid the outcome, in addition to your fee for showing up.

Appendix H

Paired Gambles Experiment Ethics Application and Participant Information

This appendix provides the complete ethics application for the three coins experiment, as approved by the University of Southampton ethics board, in addition to the participant information sheet furnished to participants.

H.1 Ethics Application

UNIVERSITY OF Southampton

This version updated December 2013

SSEGM ETHICS SUB-COMMITTEE APPLICATION FORM

Please note:

- ***You must not begin data collection for your study until ethical approval has been obtained.***
- ***It is your responsibility to follow the University of Southampton's Ethics Policy and any relevant academic or professional guidelines in the conduct of your study. This includes providing appropriate information sheets and consent forms, and ensuring confidentiality in the storage and use of data.***
- ***It is also your responsibility to provide full and accurate information in completing this form.***

1. **Name(s):** Jonathan Gray

2. **Current Position:** PhD Researcher

3. **Contact Details:**

   Division/School Social Statistics & Demography, University of Southampton

   Email j.gray@soton.ac.uk

   Phone 07528843417

4. **Is your study being conducted as part of an education qualification?**

   Yes ☒ **No** ☐

5. **If Yes, please give the name of your supervisor**

   Jakub Bijak, and Seth Bullock

6. **Title of your project:**

   Experimental Study of Decision Making from Experience

7. **Briefly describe the rationale, study aims and the relevant research questions of your study**

The study is intended to investigate how information gained from experience is used to make economic decisions about pairs of choices under uncertainty. The aims of the study are to determine which, if any, of four candidate decision rules best models human decision making from experience, and to test the hypothesis that people use a mental model of the outcome generating process, rather than just observed outcomes, to make economic decisions. The broader context of the study is to validate the learning, and decision making mechanism of agents in an agent-based model.

Research Questions:

1. Which, if any, of four decision rules (lexicographic heuristic, Bayesian risk minimisation, Bayesian, Bayesian risk minimisation with mental representation, and Cumulative Prospect Theory with Bayesian updating), best predicts human decision making from experience, under uncertainty?
2. Do people use a mental representation of the outcome generating process to make decisions from experience, when outcomes are uncertain?

8. Describe the design of your study

This study will use the ZTree software to present participants with several binary decision problems, all of which take the same form. The decision problem consists of three biased coins (A, B, and C), which have a payoff in Experimental Monetary Units (EMU) associated with heads and tails. Coins A, and C share the same payoffs, but have different biases, coin B has a single payoff for getting heads, and a different bias. Participants must choose whether to flip coins A and B, or B and C. If they flip A and B, they receive the combined payoff from both. If they flip B and C, they receive the payoff from C, and observe, but do not receive, the payoff from B. Participants will undertake 10 `sampling` rounds, where they choose a paired coin flip to observe on each round, before making a final choice and being shown the result of their chosen coin flips. After completing the set of eight decision problems, a subset of the rounds will be randomly chosen, and participants will be paid the sum of the outcomes for the coin flips in GBP, at a predefined exchange rate with EMU, in addition to an appearance fee.

Data from each participants' sampling rounds will be used as input to the four decision rules, and their predictions will then be compared to the participants' actual choices. The resulting proportions of the binary choices will be analysed using Bayesian model selection, using R, Python, and WinBUGS.

Participants will also be asked to complete a short questionnaire (attached), after completing the decision problems, to gather demographic details. These details will not be linked to the participants' responses, and will be aggregated.

9. Who are the research participants?

50-100 members of the University of Southampton, aged 18-65. The exact number of participants will depend on recruitment success.

10. If you are going to analyse secondary data, from where are you obtaining it?

Not applicable

UNIVERSITY OF Southampton

11. If you are collecting primary data, how will you identify and approach the participants to recruit them to your study?

Please upload a copy of the information sheet if you are using one – or if you are not using one please explain why.

Participants will be recruited from within the University of Southampton, via leaflets. On attending the experiment, participants will be required to show a valid University of Southampton ID card. Participants will be provided with a participant information sheet, which is included in this submission.

12. Will participants be taking part in your study without their knowledge and consent at the time (e.g. covert observation of people)? If yes, please explain why this is necessary.

No

13. If you answered 'no' to question 13, how will you obtain the consent of participants?

Please upload a copy of the consent form if you are using one – or if you are not using one please explain why.

Participants will be provided with an information sheet (included in this submission), which provides sufficient information to give informed consent to participate. They will also be provided with a consent form covering exactly what they are consenting to.

14. Is there any reason to believe participants may not be able to give full informed consent? If yes, what steps do you propose to take to safeguard their interests?

No

15. If participants are under the responsibility or care of others (such as parents/carers, teachers or medical staff) what plans do you have to obtain permission to approach the participants to take part in the study?

Not applicable

UNIVERSITY OF Southampton

16. Describe what participation in your study will involve for study participants. Please attach copies of any questionnaires and/or interview schedules and/or observation topic list to be used

Participants will use a computer to answer eight decision problems, which each consist of 10 sampling rounds, followed by a final choice. The full set of decision problems is included in the protocol document, as part of this submission.

17. How will you make it clear to participants that they may withdraw consent to participate at any point during the research without penalty?

In the participant information sheet, verbally before beginning the experiment, and on the consent form. Because participants' payment is partially dependent on their answers to the decision problems, it will also be made clear on the participant information sheet that they will receive the appearance fee irrespective of whether they complete the full set of decision problems.

18. Detail any possible distress, discomfort, inconvenience or other adverse effects the participants may experience, including after the study, and you will deal with this.

We do not anticipate any adverse effects from this study.

19. How will you maintain participant anonymity and confidentiality in collecting, analysing and writing up your data?

Participants will be assigned a number against which their responses will be recorded, their name will be used only to issue a receipt for payment.

Questionnaire responses will not be associated to the participants' decision problem responses, will be collected anonymously, and will be analysed and reported only in aggregate.

20. How will you store your data securely during and after the study?

The University of Southampton has a Research Data Management Policy, including for data retention. The Policy can be consulted at http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html

All response data will be kept in accordance with the 1998 Data Protection Act, and University of Southampton Research Data Management Policy. The anonymised decision problem responses may be made available through the University of Southampton ePrints repository, with participants' explicit consent.

21. Describe any plans you have for feeding back the findings of the study to participants.

UNIVERSITY OF Southampton

A summary of findings will be provided to any interested participants.

22. What are the main ethical issues raised by your research and how do you intend to manage these?

The primary ethical issue is that participants will be paid for taking part in the experiment, based on the responses they give to the decision problems. This could potentially impact their perceived freedom to withdraw from the experiment without penalty. To address this, payment will be split into an appearance fee, which is not contingent on completing the full set of decision problems, in addition to any earnings from the decision problems. This will also be made clear to participants on the participant information sheet.

The intention to make anonymised responses publically available for other researchers raises the issue of confidentiality and anonymity of participants. However, all data will be anonymised, and the likelihood of identification from the binary responses to the choice problems is extremely low. We will explicitly obtain consent from participants to make this data available, and make it clear that they are under no obligation to do so. The questionnaire contains potentially identifying details, and as such, individual responses will not be made publically available, and will not be linked to the participants' responses to the decision problems. Responses to the questionnaire will be analysed and reported only in aggregate.

23. Please outline any other information you feel may be relevant to this submission.

Not applicable.

H.2 Participant Information Sheet

UNIVERSITY OF Southampton

Participant Information Sheet

Study Title: Experimental Study of Decision Making from Experience

Researcher: Jonathan Gray, Jakub Bijak, and Seth Bullock
Ethics number: 20700

Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

What is the research about?

You are taking part in an economics experiment, intended to examine decision making from experience. The research is being conducted by Jonathan Gray as part of his PhD research, in the Department of Social Statistics and the Institute of Complex Systems Simulation, at the University of Southampton. The research is funded by the EPSRC, under the Care Life Cycle project, as part of the Complexity in the Real World theme.

Why have I been chosen?

You have been chosen because you have registered to participate in experiments at the Social Sciences Experimental Lab at the University of Southampton.

What will happen to me if I take part?

The experiment will take approximately 90 minutes, and will consist of a short comprehension task to ensure you understand the experiment, followed by several simple games. After completing the experiment, you will be asked to take a short questionnaire.

In each game, you will be shown three coins – A, B, and C, which have a payoff in Experimental Monetary Units (EMU) associated with heads and tails, which will be different for each game. All three coins are biased, so they are not equally likely to come up heads, or tails. The bias of the three coins is not the same, and the bias of each coin will change on each new game.

You will be asked to choose to flip two of the coins – either A and B, or B and C. If you flip coins A and B, you will receive the payoff for both. If you flip coins B and C, you will receive the payoff only for coin C.
You will be given 10 practice flips in each game to observe the outcomes, before making a final choice of which pair you wish to flip.

After completing all the games, one will be chosen at random and you will receive the EMU outcome, in addition to your initial endowment of EMU, in cash, at an exchange rate of 0.1 GBP : 1 EMU (each EMU is worth 10 pence).

Are there any benefits in my taking part?

You will be paid for one game, chosen randomly from those you complete, in addition to an initial endowment for participating, at a rate of 0.1 GBP : 1 EMU (1 EMU = 10 pence).
Your exact payment will depend partly on chance, and partly on your own actions.

Are there any risks involved?

There are no anticipated risks to you, beyond what might ordinarily be expected in an office environment.

Will my participation be confidential?

Your data will be coded anonymously. Data will be stored in accordance with the Data Protection Act of 1998, and University of Southampton policies, on a password protected

UNIVERSITY OF Southampton

computer. With your consent, your anonymised data may be made available through the University of Southampton ePrints repository at the time of publication. If you do not wish your anonymised data to be made available, you may still participate in the study. Your individual responses to the concluding questionnaire will not made public, and will be analysed and reported only in aggregate.

What happens if I change my mind?
You may withdraw at any time, without your legal rights being affected. If you withdraw, you will be paid your initial endowment of EMU.

What happens if something goes wrong?
If you have any concerns, or complaints about this study, please contact the Head of Research Governance (02380 595058, rgoinfo@soton.ac.uk).

Where can I get more information?
If you have any questions, please speak to the person conducting the experiment. Alternatively, contact Jonathan Gray (07528843417, j.gray@soton.ac.uk).

If you have any concerns or complaints about this study, please contact the Head of Research Governance at the University of Southampton (02380595058, rgoinfo@soton.ac.uk).

Glossary

antenatal	The time period covering the pregnancy, prior to birth. 33
binge drinking	Drinking with the express purpose of becoming drunk. 34
burnout	Lasting physical and emotional exhaustion, and disillusionment. 38
episiotomies	An episiotomy is a common procedure where an incision is made in the perineum to facilitate delivery, and attempt to reduce the risk of vaginal tearing. 38
term	Gestation of 37-42 weeks, 40 weeks is considered full term. 33

References

Adams-Price, C. E. and Morse, L. W. (2009). Dependency stereotypes and aging: The implications for getting and giving help in later life. *Journal of Applied Social Psychology*, 39(12):2967–2984.

Addis, M. E. and Mahalik, J. R. (2003). Men, masculinity, and the contexts of help seeking. *American Psychologist*, 58(1):5–14.

Adkins, C. J. (1968). *Equilibrium Thermodynamics*. Cambridge University Press, Cambridge, UK, 3rd edition.

Agresti, A. and Hitchcock, D. B. (2005). Bayesian inference for categorical data analysis. *Statistical Methods and Applications: Journal of the Italian Statistical Society*, 14(3):297–330.

Ainslie, G. (1991). Derivation of 'rational' economic behavior from hyperbolic discount curves. *The American Economic Review*, 81(2):334–340.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211.

Albert, S. M. and Duffy, J. (2012). Differences in risk aversion between young and older adults. *Neuroscience and Neuroeconomics*, 1(1):3–9.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, 21(4):503–546.

Alvik, A., Haldorsen, T., Groholt, B., and Lindemann, R. (2006). Alcohol consumption before and during pregnancy comparing concurrent and retrospective reports. *Alcoholism: Clinical & Experimental Research*, 30(3):510–5.

Alvik, A., Haldorsen, T., and Lindemann, R. (2005). Consistency of reported alcohol use by pregnant women: Anonymous versus confidential questionnaires with item nonresponse differences. *Alcoholism: Clinical & Experimental Research*, 29(8):1444–1449.

Andersen, A. N., Andersen, P. K., Olsen, J., Grønbæk, M., and Strandberg-Larsen, K. (2012). Moderate alcohol intake during pregnancy and risk of fetal death. *International Journal of Epidemiology*, 41(2):405–13.

Aparicio Diaz, B. and Fent, T. (2006). An agent-based simulation model of age-at-marriage norms. In Billari, F. C., Fent, T., and Scheffran, J., editors, *Agent-Based Computational Modelling: Applications in Demography, Social, Economic and Environmental Sciences*, pages 85–117. Physica-Verlag, Heidelberg.

Aparicio Diaz, B., Fent, T., Prskawetz, A., and Bernardi, L. (2011). Transition to parenthood: The role of social interaction and endogenous networks. *Demography*, 48(2):559–579.

Arthur, W. B. (2005). Out-of-equilibrium economics and agent-based modeling. *Science*, 2:1551–1564.

Austen-Smith, D. and Fryer Jr., R. G. (2005). An economic analysis of 'acting white'. *Quarterly Journal of Economics*, 120(2):551–583.

Axelrod, R. (1997). Advancing the art of simulation in the social sciences. *Complexity*, 12(3):1–13.

Axelrod, R. (2006). Agent-based modeling as a bridge between disciplines. In Tesfatsion, L. and Judd, K. L., editors, *Agent-based Computational Economics*, Handbook of Computational Economics, chapter 33, pages 1565–1584. Elsevier B.V.

Baltussen, G., Post, T., and van Vliet, P. (2006). Violations of cumulative prospect theory in mixed gambles with moderate probabilities. *Management Science*, 52(8):1288–1290.

Barber, S. J. and Mather, M. (2014). Stereotype threat in older adults: When and why does it occur and who is most affected? In *The Oxford Handbook of Emotion, Social Cognition, and Problem Solving During Adulthood*, pages 302–320. Oxford University Press.

Barr, H. M., Bookstein, F. L., O'Malley, K. D., Connor, P. D., Huggins, J. E., and Streissguth, A. P. (2006). Binge drinking during pregnancy as a predictor of psychiatric disorders on the Structured Clinical Interview for DSM-IV in young adult offspring. *The American Journal of Psychiatry*, 163(6):1061–5.

Bennett, C. M., Baird, A. A., Miller, M. B., and Wolford, G. L. (2010). Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: An argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results*, 1(1):1–5.

Bernoulli, D. (1954/1738). Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36.

Bijak, J. (2006). Forecasting international migration: Selected theories, models, and methods. Technical Report 4/2006, Central European Forum for Migration Research.

Bijak, J., Hilton, J., Silverman, E., and Cao, V. D. (2013). Reforging the wedding ring: Exploring a semi-artificial model of population for the United Kingdom with gaussian process. *Demographic Research*, 29(27):729–766.

Billari, F. C. (2015). Integrating macro- and micro-level approaches in the explanation of population change. *Population Studies*, 69(sup1):S11–S20.

Billari, F. C., Fent, T., Prskawetz, A., and Scheffran, J. (2006). Agent-based computational modelling: An introduction. In Billari, F. C., Fent, T., and Scheffran, J., editors, *Agent-Based Computational Modelling: Applications in Demography, Social, Economic and Environmental Sciences*, pages 1–16. Physica-Verlag, Heidelberg.

Billari, F. C., Ongaro, F., and Prskawetz, A. (2003a). Introduction: Agent-based computational demography. In Billari, F. C. and Prskawetz, A., editors, *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*, pages 1–18. Springer-Verlag, Heidelberg.

Billari, F. C., Prskawetz, A., Aparicio Diaz, B., and Fent, T. (2007). The ”wedding-ring”: An agent-based marriage model based on social interaction. *Demographic Research*, 17:59–81.

Billari, F. C., Prskawetz, A., and Fürnkranz, J. (2003b). On the cultural evolution of age-at-marriage norms. In Billari, F. C. and Prskawetz, A., editors, *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*, pages 139–157. Physica-Verlag, Heidelberg.

Birnbaum, M. H. (2005). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty*, 31(3):263–287.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological review*, 115(2):463–501.

Biró, M. A., Waldenström, U., Brown, S., and Pannifex, J. H. (2003). Satisfaction with team midwifery care for low- and high-risk women: A randomized controlled trial. *Birth*, 30(1):1–10.

Blair, E., Sudman, S., Bradburn, N. M., and Stocking, C. (1977). How to ask questions about drinking and sex: Response effects in measuring consumer behaviour. *Journal of Marketing Research*, 14(3):316–321.

Bongaarts, J. and Bulatao, R. A. (2000a). The accuracy of past projections. In Panel on Population Projections, Commitee on Population, and National Research Council, editors, *Beyond Six Billion: Forecasting the World's Population*, pages 37–52. The National Academies Press, Washington, DC.

Bongaarts, J. and Bulatao, R. A. (2000b). International migration. In Panel on Population Projections, Commitee on Population, and National Research Council, editors, *Beyond Six Billion: Forecasting the World's Population*, pages 156–187. The National Academies Press, Washington, DC.

Bongaarts, J. and Bulatao, R. A. (2000c). The uncertainty of population forecasts. In Panel on Population Projections, Commitee on Population, and National Research Council, editors, *Beyond Six Billion: Forecasting the World's Population*, pages 188–217. The National Academies Press, Washington, DC.

Booij, A. S., Praag, B. M. S., and Kuilen, G. (2009). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision*, 68(1-2):115–148.

Bowers, J. S. and Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3):389–414.

Box, G. (1979). Robustness in the strategy of scientific model building. In Launer, R. L. and Wilkinson, G. N., editors, *Robustness in Statistics*, pages 201–236. Academic Press, New York.

Bradley, K. A., Boyd-Wickizer, J., Powell, S. H., and Burman, M. L. (1998). Alcohol screening questionnaires in women: A critical review. *JAMA: The Journal of the American Medical Association*, 280(2):166–71.

Braitenberg, V. (1986). *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA.

Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113(2):409–32.

Bröder, A. and Newell, B. (2008). Challenging some common beliefs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making*, 3(3):205–214.

Burch, T. K. (2003a). Data, models, theory and reality: The structure of demographic knowledge. In Billari, F. C. and Prskawetz, A., editors, *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*, pages 19–40. Springer-Verlag, Heidelberg.

Burch, T. K. (2003b). Demography in a new key: A theory of population theory. *Demographic Research*, 9:263–284.

Burke, M. S., Carter, J. R., Gominiak, R. D., and Ohl, D. F. (1996). An experimental note on the Allais paradox and monetary incentives. *Empirical Economics*, 21(4):617–632.

Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125(3):367–383.

Camerer, C. F. (2000). Prospect theory in the wild: Evidence from the field. In Kahneman, D. and Tversky, A., editors, *Choices, Values, and Frames*, pages 288–300. Cambridge University Press, Cambridge.

Carnell, R. (2012). *LHS: Latin Hypercube Samples*. R package version 0.10.

Carr, P. B. and Steele, C. M. (2010). Stereotype threat affects financial decision making. *Psychological Science*, 21(10):1411–6.

Chapman, G. B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3):771–91.

Chapman, G. B. and Elstein, A. S. (1995). Valuing the future: Temporal discounting of health and money. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 15(4):373–86.

Chater, N. and Oaksford, M. (2008). The probabilistic mind: Prospects for Bayesian cognitive science. In Chater, N. and Oaksford, M., editors, *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*, pages 2–33. Oxford University Press, Oxford.

Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., and Schultz, W. (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *The Journal of Neuroscience*, 29(40):12574–83.

Conlisk, J. (1989). Three variants on the allais example. *American Economic Review*, 79(3):392–407.

Conte, R. (2000). The necessity of intelligent agents in social simulation. *Advances in Complex Systems*, 3(4):19–38.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., and Helbing, D. (2012). Manifesto of computation social science. *European Physical Journal: Special Topics*, 214(1):325–346.

Courgeau, D., Bijak, J., Franck, R., and Silverman, E. (2016). Model-based demography: Towards a research agenda. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies*, The Springer Series on Demographic Methods and Population Analysis, chapter 2, pages 28–51. Springer, Dordrecht.

Day, N. L., Richardson, G., Robles, N., Sambamoorthi, U., Taylor, P., Scher, M., Stoffer, D., Jasperse, D., and Cornelius, M. (1990). Effect of prenatal alcohol exposure on growth and morphology of offspring at 8 months of age. *Pediatrics*, 85(5):748–52.

Day, R. and Hitchings, R. (2011). 'Only old ladies would do that': Age stigma and older people's strategies for dealing with winter cold. *Health & Place*, 17(4):885–94.

De Mulder, W., Grow, A., Molenberghs, G., and Verbeke, G. (2015). Application of statistical emulation to an agent-based model: Assortative mating and the reversal of gender inequality in education in Belgium. *Proceedings of the 30th International Workshop on Statistical Modelling*, 1:139–144.

DeGroot, M. H. (1962). Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics*, 33:404–419.

Del Boca, F. K. and Darkes, J. (2003). The validity of self-reports of alcohol consumption: State of the science and challenges for research. *Addiction*, 98 Suppl 2:1–12.

Del Boca, F. K. and Noll, J. A. (2000). Truth or consequences: The validity of self-report data in health services research on addictions. *Addiction*, 95(11):347–360.

Dennis, C.-L. and Chung-Lee, L. (2006). Postpartum depression help-seeking barriers and maternal treatment preferences: A qualitative systematic review. *Birth*, 33(4):323–31.

Department of Health (2008). *Pregnancy and alcohol*. Department of Health, London.

Derks, B., Inzlicht, M., and Kang, S. (2008). The neuroscience of stigma and stereotype threat. *Group Processes & Intergroup Relations*, 11(2):163–181.

di Paolo, E. A., Noble, J., and Bullock, S. (2000). Simulation models as opaque thought experiments. In *Artificial Life VII: The Seventh International Conference on the Simulation and Synthesis of Living Systems*, pages 497–506.

Dorazio, R. M. and Johnson, F. A. (2003). Bayesian inference and decision theory-a framework for decision making in natural resource management. *Ecological Applications*, 13(2):556–563.

Elsenbroich, C. (2012). Expanation in agent-based modelling: Functions, causality or mechanisms? *Journal of Artificial Societies and Social Simulation*, 15(3):1.

Epstein, G. S. and Gang, I. N. (2006). The influence of others on migration plans. *Review of Development Economics*, 10(4):652–665.

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, 4(5):41–60.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12.

Epstein, J. M. (2014). *Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science*. Princeton University Press.

Epstein, J. M. and Axtell, R. L. (1994). *Growing Artificial Societies: Social Science from the Bottom Up*. Complex Adaptive Systems. MIT Press/Brookings Institution Press, Cambridge, MA.

Eraker, S. A. and Sox, H. C. (1981). Assessment of patients' preferences for therapeutic outcomes. *Medical decision making: An International Journal of the Society for Medical Decision Making*, 1(1):29–39.

European Commission (2012). *Eurobarometer 67.3 (May-Jun 2007)*. ZA4561 Data file Version 2.1.0. GESIS Data Archive, Cologne.

Ewert, U. C., Roehl, M., and Uhrmacher, A. M. (2003). Consequences of mortality crises in pre-modern European towns: A multiagent-based simulation approach. In Billari, F. C. and Prskawetz, A., editors, *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*, pages 175–196. Physica-Verlag, Heidelberg.

Farquhar, M., Camilleri-Ferrante, C., and Todd, C. (2000). Continuity of care in maternity services: Women's views of one team midwifery scheme. *Midwifery*, 16(1):35–47.

Feltovich, N., Harbaugh, R., and To, T. (2002). Too cool for school? Signalling and countersignalling. *RAND Journal of Economics*, 33(4):630–649.

Fishburn, P. C. (1974). Exceptional paper - lexicographic orders, utilities and decision rules: A survey. *Management Science*, 20(11):1442–1471.

Fox, C. R. and Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science*, 44(7):879–895.

Franck, R. (2002). *The Explanatory Power of Models*. Springer Netherlands, Dordrecht.

Fraser, R. B. (2006). Alcohol consumption and the outcomes of pregnancy. Statement 5, Royal College of Obstetricians and Gynaecologists, London.

Frey, S. (2014). Mixed human/entity games and the anomalous effects of misattributing strategic agency. *Adaptive Behavior*, 22(4):266–276.

Fryer Jr., R. G. and Torelli, P. (2010). An empirical analysis of 'acting white'. *Journal of Public Economics*, 94(5):380–396.

Gallistel, C. R. (2012). Extinction from a rationalist perspective. *Behavioural Processes*, 90(1):66–80.

Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. In Koehler, D. J. and Harvey, N., editors, *Blackwell Handbook of Judgment and Decision Making*, pages 62–88. Blackwell Publishing, Oxford.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3(1):20–29.

Gigerenzer, G. and Goldstein, D. (1999). Betting on one good reason: Take the best and its relatives. In *Simple Heuristics that Make Us Smart*. Oxford University Press, New York.

Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4):650–69.

Gilbert, N. (1999). Simulation a new way of doing social science. *American Behavioral Scientist*, 42(10):1485–1487.

Glöckner, A. and Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, 123(1):21–32.

Godfray, H. C. J. (1991). Signalling of need by offspring to their parents. *Nature*, 352(6333):328–330.

Goffman, E. (1990). *Stigma: Notes on the Management of Spoiled Identity*. Penguin Books, London.

Goldstein, D. G. and Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109(1):75–90.

Gomberg, E. S. (1988). Alcoholic women in treatment: The question of stigma and age. *Alcohol and Alcoholism*, 23(6):507–14.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, 144:517–546.

Gray, J., Bijak, J., and Bullock, S. (2016). Deciding to disclose: A decision theoretic agent model of pregnancy and alcohol misuse. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, chapter 11. Springer, Dordrecht.

Gray, R. and Henderson, J. (2006). Report to the department of healthreview of the fetal effects of prenatal alcohol exposure. Technical report, Oxford: National Perinatal Epidemiology Unit.

Griffiths, T. L., Chater, N., Kemp, C., Tenenbaum, J. B., and Perfors, A. (2010). Probabilistic model of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, 14(8):357–364.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., and DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115–126.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., and Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, 221(23):2760–2768.

Grow, A. (2016). Regression metamodels for sensitivity analysis in agent-based computational demography. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, number 41 in The Springer Series on Demographic Methods and Population Analysis, chapter 7, pages 185–211. Springer.

Grüne-Yanoff, T. (2009). The explanatory potential of artificial societies. *Synthese*, 169(3):539–555.

Guerri, C., Riley, E., and Strömland, K. (1999). Commentary on the recommendations of the Royal College of Obstetricians and Gynaecologists concerning alcohol consumption in pregnancy. *Alcohol and Alcoholism*, 34(4):497–501.

Hamilton, W. D., Axelrod, R., and Tanese, R. (1990). Sexual reproduction as an adaptation to resist parasites (a review). *Proceedings of the National Academy of Sciences of the United States of America*, 87(9):3566–73.

Harrison, G. W. (1994). Expected utility theory and the experimentalists. In Hey, J. D., editor, *Experimental Economics*, pages 42–73. Springer-Verlag, Heidelberg.

Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players, I-III. *Management Science*, 14(3):159–182.

Hatem, M., Sandall, J., Devane, D., Soltani, H., and Gates, S. (2008). Midwife-led versus other models of care for childbearing women. *Cochrane Database of Systematic Reviews*, 2008(3):1–105.

Hau, R., Pleskac, T. J., Kiefer, J., and Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, 21(5):493–518.

Heard, D., Dent, G., Shifeling, T., and Banks, D. (2015). Agent-based models and microsimulation. *Annual Review of Statistics and Its Application*, 2(1):259–272.

Henderson, J., Kesmodel, U., and Gray, R. (2007). Systematic review of the fetal effects of prenatal binge-drinking. *Journal of Epidemiology and Community Health*, 61(12):1069–73.

Henrich, J. (2009). The evolution of costly displays, cooperation and religion. *Evolution and Human Behavior*, 30(4):244–260.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, 466(7302):29–29.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010b). The weirdest people in the world? *The Behavioral and brain sciences*, 33(2-3):61–83; discussion 83–135.

Hertwig, R., Barron, G., Weber, E. U., and Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8):534–9.

Hess, M. and Dikken, J. (2010). The association between ageism and subjective age of older people in Europe. *International Journal of Social Sciences and and Humanity Studies*, 2(1):99–109.

Hilbig, B. E. (2010). Reconsidering ”evidence” for fast-and-frugal heuristics. *Psychonomic bulletin & review*, 17(6):923–930.

Hilton, J. and Bijak, J. (2016). Design and analysis of demographic simulations. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, The Springer Series on Demographic Methods and Population Analysis, chapter 8, pages 211–237. Springer.

Huizink, A. C. (2009). Moderate use of alcohol, tobacco and cannabis during pregnancy: New approaches and update on research findings. *Reproductive Toxicology*, 28(2):143–51.

Huizink, A. C. and Mulder, E. J. H. (2006). Maternal smoking, drinking or cannabis use during pregnancy and neurobehavioral and cognitive functioning in human offspring. *Neuroscience and Biobehavioral Reviews*, 30(1):24–41.

Hume, D. (1739). *A Treatise of Human Nature*. Clarendon Press, Oxford.

Humphriss, R., Hall, A., May, M., Zuccolo, L., and Macleod, J. (2013). Prenatal alcohol exposure and childhood balance ability: Findings from a UK birth cohort study. *BMJ Open*, 3(6).

Hunter, B. (2006). The importance of reciprocity in relationships between community-based midwives and mothers. *Midwifery*, 22(4):308–22.

Improving Access to Psychological Therapies, DrugScope, and National Treatment Agency for Substance Misuse (2012). IAPT positive practice guide for working with people who use drugs and alcohol.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–291.

Kahneman, D. and Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4):341–350.

Kashyap, R. and Villavicencio, F. (2016a). An agent-based model of sex ratio at birth distortions. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, The Springer Series on Demographic Methods and Population Analysis, chapter 12, pages 343–369. Springer, Dordrecht.

Kashyap, R. and Villavicencio, F. (2016b). The dynamics of son preference, technology diffusion, and fertility decline underlying distorted sex ratios at birth: A simulation approach. *Demography*, 53(5):1261–1281.

Kaskutas, L. A. and Graves, K. (2000). An alternative to standard drinks as a measure of alcohol consumption. *Journal of Substance Abuse*, 12(1-2):67–78.

Kennedy, H. P., Shannon, M. T., Chuahorm, U., and Kravetz, M. K. (2004). The landscape of caring for women: A narrative study of midwifery practice. *Journal of Midwifery & Women's Health*, 49(1):14–23.

Kennedy, M. C. (2004). *GEM Software*. Centre for Terrestrial Carbon Dynamics, Sheffield.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

Kesmodel, U., Wisborg, K., Olsen, S. F., Henriksen, T. B., and Secher, N. J. (2002). Moderate alcohol intake during pregnancy and the risk of stillbirth and death in the first year of life. *American Journal of Epidemiology*, 155(4):305–12.

Klabunde, A. (2014). Computational economic modeling of migration. *Ruhr Economic Papers 471*.

Klabunde, A. and Willekens, F. (2016). Decision-making in agent-based models of migration: State of the art and challenges. *European Journal of Population*, 32(1):73–97.

Kovalchik, S., Camerer, C. F., Grether, D. M., Plott, C. R., and Allman, J. M. (2005). Aging and decision making: A comparison between neurologically healthy elderly and young individuals. *Journal of Economic Behavior & Organization*, 58(1):79–94.

Kreps, D. M. and Cho, I. (1987). Signalling games and stable equilibria. *The Quarterly Journal of Economics*, 102(2):179–221.

Kristensen, S. (1997). Sensor planning with Bayesian decision theory. *Robotics and Autonomous Systems*, 19(3-4):273–286.

Langenbucher, J. and Merrill, J. (2001). The validity of self-reported cost events by substance abusers: Limits, liabilities, and future directions. *Evaluation Review*, 25(2):184–210.

Latino-Martel, P., Chan, D. S. M., Druesne-Pecollo, N., Barrandon, E., Hercberg, S., and Norat, T. (2010). Maternal alcohol consumption during pregnancy and risk of childhood leukemia: Systematic review and meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention*, 19(5):1238–60.

Leombruni, R. and Richiardi, M. (2005). Why are economists sceptical about agent-based simulations? *Physica A: Statistical Mechanics and its Applications*, 355(1):103–109.

Loewenstein, G. and Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, 107(2):573–597.

Lorek, H. and Sonnenschein, M. (1999). Modelling and simulation software to support individual-based ecological modelling. *Ecological Modelling*, 115(2-3):199–216.

Luce, B. Y. R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24(2):178–191.

Macy, M. W. and Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, 28(1):143–166.

Maier, S. E. and West, J. R. (2001). Drinking patterns and alcohol-related birth defects. *Alcohol Research & Health*, 25(3):168–174.

Malvin, J. H. and Moskowitz, J. M. (1983). Anonymous versus identifiable self-reports of adolescent drug attitudes, intentions, and use. *Public Opinion Quarterly*, 47(4):557.

Marks, M. N., Siddle, K., and Warwick, C. (2003). Can we prevent postnatal depression? A randomized controlled trial to assess the effect of continuity of midwifery care on rates of postnatal depression in high-risk women. *The Journal of Maternal-fetal & Neonatal Medicine*, 13(2):119–27.

Marmot, M., Oldfield, Z., Clemens, S., Blake, M., Phelps, A., Nazroo, J., Steptoe, A., Rogers, N., and Banks, J. (2015). English longitudinal study of ageing: Waves 0-6, 1998-2013. data source.

McNamara, J. and Houston, A. (1980). The application of statistical decision theory to animal behaviour. *Journal of Theoretical Biology*, 85(4):673–90.

Mesoudi, A. (2009). The cultural dynamics of copycat suicide. *PLoS ONE*, 4(9).

Meyer-Leu, Y., Lemola, S., Daeppen, J.-B., Deriaz, O., and Gerber, S. (2011). Association of moderate alcohol use and binge drinking during pregnancy with neonatal health. *Alcoholism: Clinical & Experimental Research*, 35(9):1669–77.

Mikels, J. A. and Reed, A. E. (2009). Monetary losses do not loom large in later life: Age differences in the framing effect. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 64(4):457–60.

Miller, R. R. (2012). An assessment of Gallistel's (2012) rationalistic account of extinction phenomena. *Behavioural Processes*, 90(1):81–3; discussion 87–8.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, 54(2):286–295.

National Institute for Health and Care Excellence (2007). *Antenatal and Postnatal Mental Health. CG45*. NICE, London.

National Institute for Health and Care Excellence (2010a). *Alcohol-use Disorders: Preventing Harmful Drinking. PH24*. NICE, London.

National Institute for Health and Care Excellence (2010b). *Antenatal Care. CG62*. NICE, London.

National Institute for Health and Care Excellence (2010c). *Pregnancy and Complex Social Factors. CG110*. NICE, Manchester.

Neilson, W. and Stowe, J. (2002). A further examination of cumulative prospect theory parameterizations. *Journal of Risk and Uncertainty*, 24(1):31–47.

Newell, B. R., Weston, N. J., and Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone ”takes-the-best”. *Organizational Behavior and Human Decision Processes*, 91(1):82–96.

Niazi, M. and Hussain, A. (2011). Agent-based computing from multi-agent systems to agent-based models: A visual survey. *Scientometrics*, 89(2):479–499.

Nilsson, H., Rieskamp, J., and Wagenmakers, E.-J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55(1):84–93.

Norwegian Social Science Data Services (2008). ESS round 4: European social survey round 4 data.

Oakley, J. and O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4):769–784.

Oakley, J. E. and O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: A Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769.

Oliver, A. (2003). A quantitative and qualitative test of the Allais paradox using health outcomes. *Journal of Economic Psychology*, 24(1):35–48.

Oppenheimer, D. M. (2003). Not so fast! (and not so frugal!): Rethinking the recognition heuristic. *Cognition*, 90(1):1–9.

Padoa-Schioppa, C. and Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090):223–6.

Padoa-Schioppa, C. and Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience*, 11(1):95–102.

Parsons, S. and Wooldridge, M. (2002). Game theory and decision theory in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, pages 1–14.

Peterson, M. (2009). *An Introduction to Decision Theory*. Cambridge University Press, Cambridge, UK.

Phillips, D., Thomas, K., Cox, H., Ricciardelli, L. A., Ogle, J., Love, V., and Steele, A. (2007). Factors that influence women's disclosures of substance use during pregnancy: A qualitative study of ten midwives and ten pregnant women. *The Journal of Drug Issues*, 37(2):357–376.

Piccinelli, M. and Tessari, E. (1997). Efficacy of the alcohol use disorders identification test as a screening tool for hazardous alcohol intake and related disorders in primary care: A validity study. *British Medical Journal*, 314:420–424.

Platt, M. L. and Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741):233–8.

Prior, L., Wood, F., Lewis, G., and Pill, R. (2003). Stigma revisited, disclosure of emotional problems in primary care consultations in Wales. *Social Science & Medicine*, 56(10):2191–200.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Radcliffe, P. (2011). Substance-misusing women: Stigma in the maternity setting. *British Journal of Midwifery*, 19(8):497–506.

Redshaw, M. and Henderson, J. (2014). *Safely Delivered: A National Survey of Womens' Experiences of Maternity Care 2014*. The National Perinatal Epidemiology Unit, Oxford.

Resnick, M. (1994). *Turtles, Termites and Traffic Jams: Explorations in Massively Parallel Microworlds*. The MIT Press, Cambridge, MA.

Ríos Insua, D., Banks, D., and Rios, J. (2015a). Modeling opponents in adversarial risk analysis. *Risk Analysis*, 46(4):742–755.

Ríos Insua, D., Rios, J., and Banks, D. (2009). Adversarial risk analysis. *Journal of the American Statistical Association*, 104(486):841–854.

Ríos Insua, D., Rios, J., and Banks, D. (2015b). *Adversarial Risk Analysis*. CRC Press, Boca Raton.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *The Annals of Mathematical Statistics*, 35(1):1–20.

Royal College of Obstetricians and Gynaecologists (1996). Alcohol consumption in pregnancy. In *RCOG Guideline No. 9*. RCOG, London.

Russell, M. (1994). New assessment tools for risk drinking during pregnancy. *Alcohol Health and Research World*, 18(1):55–67.

Russell, M. and Martier, S. S. (1996). Detecting risk drinking during pregnancy: A comparison of four screening questionnaires. *American Journal of Public Health*, 86(10):1435–1439.

Rustichini, A. (2009). Neuroeconomics: What have we found, and what should we search for. *Current Opinion in Neurobiology*, 19(6):672–7.

Samejima, F. (1970). Graded response model of the latent trait theory and tailored testing. *Psychometrika*, 35(1):5–17.

Samuelson, P. A. (1937). Note on measurement of utility. *The Review of Economic Studies*, 4(2):155–161.

Sandall, J. (1997). Midwives' burnout and continuity of care. *British Journal of Midwifery*, 5(2):106–111.

Savitz, D. A. (2012). Re: Moderate alcohol intake during pregnancy and risk of fetal death. *International Journal of Epidemiology*, 41(6):1847; author reply 1848.

Silverberg, J. L., Bierbaum, M., Sethna, J. P., and Cohen, I. (2013). Collective motion of moshers at heavy metal concerts. *Physical Review Letters*, 110(22):228701.

Silverman, E., Bijak, J., Hilton, J., Cao, V. D., and Noble, J. (2013). When demography met social simulation: A tale of two modelling approaches. *Journal of Artificial Societies and Social Simulation*, 16(4):9.

Silverman, E., Bijak, J., and Noble, J. (2011). Feeding the beast: Can computational demographic models free us from the tyranny of data? In Lenaerts, T., Giacobini, M., Bersini, H., Bourgine, P., Dorigo, M., and Doursat, R., editors, *Advances in Artificial Life, ECAL 2011*, volume 276, pages 747–754. MIT Press.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138.

Simon, H. A. (1959). Theories of decision-making in economics and behavioural science. *The American Economic Review*, 49(3):253–283.

Simon, H. A. (1996). *The Sciences of the Artificial*. MIT Press, Cambridge, MA, 3rd edition.

Simon, H. A. (2000). Bounded rationality in social science: Today and tomorrow. *Mind & Society*, 1(1):25–39.

Smith, J. A., Braunack-Mayer, A., Wittert, G., and Warin, M. (2007). 'I've been independent for so damn long!': Independence, masculinity and aging in a help seeking context. *Journal of Aging Studies*, 21(4):325–335.

Smith, J. M. and Price, G. R. (1973). The logic of animal conflict. *Nature*, 246(5427):15–18.

Sokol, R. J., Martier, S. S., and Ager, J. W. (1989). The T-ACE questions: Practical prenatal detection of risk-drinking. *American Journal of Obstetrics and Gynecology*, 160(4):863 – 870.

Soon, C. S., Brass, M., Heinze, H.-j., and Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, 11(5):543–5.

Sosis, R. (2003). Why aren't we all hutterites? *Human Nature*, 14(2):91–127.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Stevens, T. and McCourt, C. (2002). One-to-one midwifery practice part 2: The transition period. *British Journal of Midwifery*, 10(1):45–50.

Stockwell, T., Donath, S., Cooper-Stanbury, M., Chikritzhs, T., Catalano, P., and Mateo, C. (2004). Under-reporting of alcohol consumption in household surveys: A comparison of quantity-frequency, graduated-frequency and recent recall. *Addiction*, 99(8):1024–33.

Storn, R. and Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous space. *Journal of Global Optimization*, 11:341–359.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, 32(2):101–130.

Strandberg-larsen, K., Nielsen, N. R., Andersen, P. K., and Andersen, A. N. (2008). Binge drinking in pregnancy and risk of fetal death. *Obstetrics & Gynecology*, 111(3):602–609.

Streissguth, A. P., Barr, H. M., Olson, H. C., Sampson, P. D., Bookstein, F. L., and Burgess, D. M. (1994). Drinking during pregnancy decreases word attack and arithmetic scores on standardized tests: Adolescent data from a population-based prospective study. *Alcoholism: Clinical & Experimental Research*, 18(2):248–54.

Sun, Y., Strandberg-Larsen, K., Vestergaard, M., Christensen, J., Andersen, A. N., Grønbæk, M., and Olsen, J. (2009). Binge drinking during pregnancy and risk of seizures in childhood: A study based on the danish national birth cohort. *American Journal of Epidemiology*, 169(3):313–22.

ten Broeke, G., van Voorn, G., and Ligtenberg, A. (2016). Which sensitivity analysis method should i use for my agent-based model. *Journal of Artificial Societies and Social Simulation*, 19(1):1–35.

Tenenbaum, J. B., Griffiths, T. L., and Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10(7):309–18.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, 8(3):201–207.

Thaler, R. (2000). From homo economicus to homo sapiens. *The Journal of Economic Perspectives*, 14(1):133–141.

The NHS Information Centre Social Care Statistics (2010). *Personal Social Services: Expenditure and Unit Costs, England - 2008-09*. The Health and Social Care Information Centre, Leeds.

Thiele, J. C., Kurth, W., and Grimm, V. (2014). Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *Journal of Artificial Societies and Simulation*, 17(3):11.

Thompson, N. S. and Derr, P. (2009). Contra Epstein, good explanations predict. *Journal of Artificial Societies and Social Simulation*, 12(1):9.

Todd, C. J., Farquhar, M. C., and Camilleri-Ferrante, C. (1998). Team midwifery: The views and job satisfaction of midwives. *Midwifery*, 14(4):214–24.

Todd, P. M., Billari, F. C., and Simão, J. (2005). Aggregate age-at-marriage patterns from individual mate-search heuristics. *Demography*, 42(3):559–574.

Todd, P. M. and Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology*, 24(2):143–165.

Toll, B. A., O'Malley, S. S., and Katulak, N. A. (2007). Comparing gain- and loss-framed messages for smoking cessation with sustained-release bupropion: A randomized controlled trial. *Psychology of Addictive Behaviour*, 21(4):534–544.

Turner, C. (1990). How much alcohol is in a 'standard drink'? An analysis of 125 studies. *British Journal of Addiction*, 85(9):1171–5.

Tversky, A. (1975). A critique descriptive of expected utility theory: Descriptive and normative considerations. *Erkenntnis*, 9:163–173.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–31.

Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions rational choice and the framing of decisions. *The Journal of Business*, 59(4):251–278.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Van Bavel, J. and Grow, A. (2016). Introduction: Agent-based modelling as a tool to advance evolutionary population theory. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, The Springer Series on Demographic Methods and Population Analysis, chapter 1, pages 3–27. Springer, Dordrecht.

Van Walraven, C., Oake, N., Jennings, A., and Forster, A. J. (2010). The association between continuity of care and outcomes: A systematic and critical review. *Journal of Evaluation in Clinical Practice*, 16(5):947–56.

Vlachantoni, A., Shaw, R., Evandrou, M., and Falkingham, J. (2015). The determinants of receiving social care in later life in England. *Ageing and Society*, 35(2):321–345.

Vlachantoni, A., Shaw, R., Willis, R., Evandrou, M., Falkingham, J., and Luff, R. (2011). Measuring unmet need for social care amongst older people. *Population Trends*, 145(1):60–76.

Von Neumann, J. and Morgenstern, O. (1953). *The Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 3rd edition.

Wakefield, J. R. H., Hopkins, N., and Greenwood, R. M. (2013). Meta-stereotypes, social image and help seeking: Dependency-related meta-stereotypes reduce help-seeking behaviour. *Journal of Community & Applied Social Psychology*, 23(5):363–372.

Waldherr, A. and Wijermans, N. (2013). Communicating social simulation models to sceptical minds. *Journal of Artificial Societies and Social Simulation*, 16(4):13.

Walters, K., Iliffe, S., and Orrell, M. (2001). An exploration of help-seeking behaviour in older people with unmet needs. *Family Practice*, 18(3):277–82.

Watson, A. C., Corrigan, P., Larson, J. E., and Sells, M. (2007). Self-stigma in people with mental illness. *Schizophrenia Bulletin*, 33(6):1312–1318.

Watson, K. K. and Platt, M. L. (2012). Social signals in primate orbitofrontal cortex. *Current Biology*, 22(23):2268–73.

Watts, D., Dodds, P., and Newman, M. (2002). Identity and search in social networks. *Science*, 296:1302–1305.

Wenger, G. C. (1991). A network typology: From theory to practice. *Journal of Aging Studies*, 5(2):147–162.

Wildman, W. J. and Sosis, R. (2011). Stability of groups with costly beliefs and practices. *Journal of Artificial Societies and Social Simulation*, 14(3).

Willekens, F. (2016). The decision to emigrate: A simulation model based on the theory of planned behaviour. In Grow, A. and Van Bavel, J., editors, *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, The Springer Series on Demographic Methods and Population Analysis, chapter 10, pages 257–299. Springer, Dordrecht.

Willis, M. L., Palermo, R., Burke, D., McGrillen, K., and Miller, L. (2010). Orbitofrontal cortex lesions result in abnormal social judgements to emotional faces. *Neuropsychologia*, 48(7):2182–7.

Wooldridge, M. (1999). Intelligent agents. In Weiss, G., editor, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, pages 27–78. MIT Press, MA.

Wu, B. M., Birkin, M. H., and Rees, P. H. (2008). A spatial microsimulation model with student agents. *Computers, Environment and Urban Systems*, 32(6):440–453.

Zahavi, A. (1975). Mate selection-a selection for a handicap. *Journal of Theoretical Biology*, 53(1):205–14.