University of Southampton
Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

FEASIBILITY OF BAYESIAN DECISION THEORETIC AGENTS AS A MODEL FOR DISCLOSURE OF DRINKING BEHAVIOUR IN PREGNANCY

by JONATHAN GRAY

Completed: September 2013

A dissertation submitted in partial fulfilment of the degree of
MSc Complex Systems Simulation
by examination and dissertation

University of Southampton
Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

FEASIBILITY OF BAYESIAN DECISION THEORETIC AGENTS AS A MODEL FOR DISCLOSURE OF DRINKING BEHAVIOUR IN PREGNANCY

by JONATHAN GRAY

Completed: September 2013

A dissertation submitted in partial fulfilment of the degree of
MSc Complex Systems Simulation
by examination and dissertation

ABSTRACT

This dissertation presents a method for modelling disclosure behaviour by treating the interaction as paired signalling games played by decision theoretic agents. Two theories of decision making - Bayesian risk minimisation, and Cumulative Prospect Theory (CPT) - are investigated, and a simulation developed using Python.

The feasibility of the method is examined through a case study, which considers the disclosure of the drinking behaviour of pregnant women to their midwives.

The essence of the scenario is that there is considered to be a long term benefit to disclosure - common in the healthcare arena, but an opportunity cost associated with disclosure. In the case study, this is conceptualised as arising from the perceived undesirability of drinking while pregnant. More generally this could derive from any imbalance in the long term benefit and the opportunity cost, for example the subjective benefit of a cigarette in the near future, versus the discounted benefit of better health later.

Theories of decision are driven by the weighing of probabilities and subjective gains or losses. In this case, the probabilities are generated by individual agents based on their initial preconceptions and their experiences across the simulation, using Bayesian inference.

The Bayesian risk minimising model is able to reproduce qualitative trends around increased honesty over appointments, and a negative impact of harsh judgement of drinkers on disclosure. The CPT model is less successful, which may be a result of improper, or excessively homogenous parameters, in combination with unrealistic payoffs.

A global sensitivity analysis is also conducted using Gaussian Emulation Machines, and finally recommendations for further work are dervived, along with a few key recommendations for practice - assume people are being honest, and be non-judgemental.

The more accurate the map, the more it resembles the territory. The most accurate map possible would be the territory, and thus would be perfectly accurate and perfectly useless.

— Mr. Ibis Gaiman [33]

ACKNOWLEDGMENTS

My thanks to the ICSS, and the Care Lifecycle Project for ecosystem and funding, and to Jakub Bijak for supervising.

And thanks to Jen, for the conversation that sparked the idea, and everything else.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

FAS Foetal Alcohol Syndrome

CPT Cumulative Prospect Theory

PT Prospect Theory

NICE National Institute for Health and Care Excellence

IAPT Improving Access to Psychological Therapies

AUDIT Alcohol Use Disorders Identification Test

T-ACE Tolerance, Annoyance, Cut down, Eye-opener

RCOG Royal College of Obstetricians and Gynecologists

AML acute myeloid leukemia

ADHD Attention Defecit Hyperactivity Disorder

PND Postnatal Depression

RCT randomised control trial

ESS Evolutionarily Stable Strategy

MCMC Markov chain monte carlo methods

DU Discounted Utility

AIC Akaike information criterion

ANOVA analysis of variance

GEM Gaussian Emulation Machine

GEM-SA Gaussian Emulation Machine for Sensitivity Analysis

JIT Just-in-time

GLOSSARY

antenatal	The time period covering the pregnancy, prior to birth. 7
binge drinking	Drinking with the express purpose of becoming drunk. 8
burnout	Lasting physical and emotional exhaustion, and disillusionment. 12
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. 12
gravid	Pregnant. 3
term	Gestation of 37-42 weeks, 40 weeks is considered full term. 7

Part I

INTRODUCTION AND BACKGROUND

1 INTRODUCTION

This dissertation considers the feasibility of signalling games [60] played by populations of decision theoretic Bayesian agents as a model of disclosure behaviour.

To test the principle, a case study of the disclosure of drinking behaviour by pregnant women to their midwives is presented. The case study was first framed as a signalling game, then translated into two interlinked decision problems. A simulation was then implemented in Python, where populations of midwife, and gravid agents played out the signalling game. Both classes of agents used Bayesian inference, and decision rules to reason about what moves to play over a number of iterations of the game. This simulation model was then used to examine whether this could produce an adequate representation of women's behaviour when choosing how honest to be about their drinking behaviour at appointments with a midwife.

The overall dynamics of the model are contrasted to qualitative trends reported by Phillips et al. [84], and Alvik et al. [3] with promising results. The sensitivity of model parameters was also explored, along with the impact of the prior beliefs of midwives on the honesty of their patients, and the feasibility of a descriptive decision rule in place of Bayesian risk minimisation. Recommendations for practice, and further research were then derived from the results of simulation.

Taking the conception of simulation models as 'opaque thought experiments' as mooted by Di Paolo et al. [26], arguably the first question to ask is whether they offer any advantage over a thought experiment in this context. A simulation model is interrogable, it can be passed around, automated, the outputs optimised. Crucially, a simulation model can have a far wider scope, and deeper complexity than a simple thought experiment. This makes them ideal tools for policy makers in the role of hypothesis generators, and as an aid to shared problem solving.

Ideally, the model can effectively act as a shared ontology and actually make some aspects of the thought experiment more transparent by providing a concrete and inspectable model alongside individual mental models. If the randomised control trial (RCT) is a tool to evaluate the impact of policy decisions in the real world, then simulation models can be an important first step in deciding what should reach that point. This has the potential to substantially reduce the costs in time, and money of failed policy by pointing up potential adverse outcomes, or interventions unlikely to produce any change in the system.

Simulation can also offer the ability to speculatively explore systems which are in some way opaque to investigation in the real world, often because of the expense, or ethical concerns involved in doing so. In the spirit of both putative advantages, the case study focuses on a system which represents a high stakes policy environment - maternity claims are the largest legal costs faced by the NHS [77], and is difficult to access because it turns on deception.

Current advice is that women should not drink during pregnancy, or in the run up to it [74, 25]. However the 2010 Infant Feeding Survey [46] found that 40% of women continued to drink during pregnancy, and that only 70% received information on the risks, or advice on how to cut down. The extent to which drinking during pregnancy which is harmful below levels that might be considered disordered is contentious (section 2.1.1), but the perception of it as socially undesirable is not. This balance between socially undesirable, yet potentially medically relevant information makes the disclosure situation interesting. There is, or may be perceived to be, an immediate social cost, yet the information might provide some long term benefit if revealed.

The methodological aspect to this work is the more significant feature, because it is can be generalised to a variety of different application domains. Eliciting the right information is always important, but critical in health and social care. Much of this comes down to asking the right questions, but in many cases the onus is on the client to give the right information. For example reporting side effects of drugs accurately, or even to proactively provide information on abuse.

Approaching this class of interaction by framing it as a signalling game provides a well understood framework for teasing out the key features, or at least debating which features those are. Translating the game to a decision problem serves to abstract away some of the difficulties inherent in a game theoretic approach. This is particularly useful where the desire is to describe behaviour rather than obtain the rational equilibrium solution. Rather than attempt to replicate a complex process of mental modelling and find the equilibrium, it is assumed instead that this process can be treated probabilistically and that system is more revealing in motion than at rest. Other minds are black boxes, and act according to some underlying probability distribution than can be inferred. This shifts the emphasis onto providing a plausible model of how decisions are made. However, it does not preclude the possibility of incorporating strategic sophistication into the model, since it is inherently modular. Having cast it as a decision problem, the rules which govern decisions can effectively swapped out, and the rules are agnostic in terms of how the information is derived. This is demonstrated to some extent by the inclusion of CPT agents, which use the same inferential process, and play the same game as their Bayesian counterparts but lead to markedly different overall dynamics.

The decision problem itself can also be altered without a wholesale reconstruction of the agents, and beliefs derived in one environment can inform others. This last is an interesting prospect, for example consider a population of agents which have played a mental health signalling game transitioning to a smoking behaviour game, where a history of mental illness would indicate against some lines of treatment.

A decision problem formulation with Bayesian inference to derive probabilities is also attractive because of the balance it strikes between computational tractability, and cognitive plausibility. The suggestion that this offers better cognitive plausibility than a classic thresholds and conditionals type agent model is motivated by the ongoing work by proponents of 'neuroeconomics', which as noted by Rustichini [94] aims to provide a test of the numerous theories of decision. The evidence supports many key aspects of the decision theory model as corresponding to real decision making, particularly the notion of a common currency for comparing the desirability of outcomes.

Potentially then, this approach offers a structured but flexible way to model some classes of human system. The simulation method is tractable enough to scale, but captures some of the sophistication of decision making in a way that is cognitively plausible, and supported by evidence from both empirical economics and neurology.

2 BACKGROUND

This chapter presents an overview of literature focusing on the impact of drinking behaviour in pregnancy, and factors affecting disclosure behaviour in the midwifery context. This is followed by a review of literature supporting the theoretical underpinning of the modelling approach, with particular reference to statistical decision theory.

2.1 ALCOHOL, AND DISCLOSURE IN THE MATERNITY SETTING

2.1.1 *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. [5] 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 [97] 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. [59] 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. [5] this was significant at term.

Considering longer term negative outcomes, a metastudy by Latino-Martel et al. [63] 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 leukemia (AML), but note the rarity of the condition as a limitation.

Huizink and Mulder [50] 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 Defecit 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 [49], 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 [40] found there was insufficient evidence to suggest any harm arising from moderate (under 1.5 UK units per day) alcohol consumption. This ties to the current guidance from the National Institute for Health and Care Excellence (NICE) [74], 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 Gynecologists (RCOG) [91] suggesting no evidence of harm below 15 units per week, and subsequent criticism by Guerri et al. [42], 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. [21]). A subsequent RCOG statement [90] 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 [66], 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. [104] 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. [106] looked at seizure disorders in children whose mothers binged during pregnancy. They reported significantly greater risk of both neonatal seizures (~3 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. [105] 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. [47] cautiously 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. [70] 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. [7] 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.

2.1.2 *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 [23], but do not correspond perfectly to alternative methods. Del Boca and Darkes [22] 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 [37], disclosure equates to revelation of the mark. This is not immediately contentious, for example Gomberg [38] identified stigma surrounding alcohol abusing women in particular, an issue also highlighted by Improving Access to Psychological Therapies (IAPT) guidance [52], as well as a number of other studies relating response effects to perceived negative consequences [62, 23, 11]. In the maternity context, Radcliffe [88] 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 [75] 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 [73], there is no specific policy for routine antenatal care beyond providing information on possible impacts of alcohol consumption [74]. NICE guidance on pregnancy and complex social factors [75] 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 2.1.1), this suggests that much of the onus is on the patient to volunteer information.

Where screening is used, Kaskutas and Graves [56] 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 [103], and a lack of consistency in the standard drink measure [115]. Alternative screening tools, for example AUDIT, and T-ACE are available and have been shown to perform well in identifying problematic levels of drinking [85, 14, 92, 93], although the emphasis in these cases is on consumption at disordered levels.

Prior et al. [87] considered a different health arena (mental health problems and GPs), 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. [4], 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. [3], which found that contemporaneous reports of consumption were significantly lower than those postpartum. 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 wellbeing during pregnancy was associated with lower retrospective reports. Taken together with [4], these results could be seen as conflicting, but may suggest self-stigmatisation [121], or reflect a lack of distinction between anonymity, and confidentiality [67].

In summation then, there is a consensus that alcohol consumption is generally underreported 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 underreporting and consumption, i.e. that heavier drinking is associated with a greater tendency to understate intake.

2.1.3 *Practice Implications*

Given that alcohol consumption is thought to be underreported some consideration must be given to the implications for midwifery practice, in terms of eliciting more accurate self-reporting. Phillips et al. [84] 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 [102] 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. [57] 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 unidirectionally didactic relationship.

Hunter [51] 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 practiced 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. [113] 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. [28] 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. [10] looked at a 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. [68] 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 [24]. 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. [45] 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.. 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 [119], but is clearly not a cost free endeavour, with particular concern arising from the emotional cost [113], and increased rates of 'burnout' in practitioners [96].

2.2 GAMES, SIGNALS, AND DECISIONS

2.2.1 *Signalling Games*

Game theory generally deals with strategic decision making in the unusual circumstance of complete information, that is, every player has at least complete knowledge of all possible outcomes, who their opponents are, and so forth. Arguably more generally applicable is the incomplete information scenario, where players lack information about the rules of play in some fashion. Harsanyi [43] proposed a method for effectively transforming such games into games of complete information by treating the possible variations on the rules as subgames. To determine which subgame is to be played, an additional player - nature - is introduced to make the first move, where nature conducts a lottery according to some probability distribution. If it is assumed that the underlying probability distribution is known to all players, the game is then one of complete information.

Perhaps the best known example of Bayesian games, are the signalling games codified by Kreps and Cho [60], after initially being framed by Spence [101] in the context of employment markets. The general form of such a game is that one player holds information known only to them, on the basis of which they send a signal to the other player(s), which the other player(s) then act upon. Much of the interest in signalling games turns on what conditions are necessary for honest signalling to be a Nash equilibrium, or in the context of evolutionary game theory, an Evolutionarily Stable Strategy (ESS) .

One approach to this requires that signalling is a costly exercise, as proposed by Grafen [39] in examining biological signals (for example, the eye-catching but unwieldy peacock tail). Grafen demonstrated that an earlier suggestion by Zahavi [125], who proposed that such signals were in effect a handicap demonstrating fitness, would lead to an ESS because of the costly nature of the signalling. This solution is also noted by Spence [101], who showed that a separating equilibrium exists[1] contingent on signals being more costly for some types.

Costly signalling has been applied to explain a variety of apparently contradictory behaviours, for example Godfray [36] in the context of offspring soliciting food from parents, where the key question is why a behaviour with potentially very high costs (namely, being eaten) would be preferred to a less risky method. In a social context, costly signalling has been proposed as an explanation for religion in human societies. Sosis [100] developed a model of religious ritual as an exercise in costly signalling, showing that higher costs to engagement in rituals for skeptics maintains the stability of religious groups and the presumed benefits that membership confers. Henrich [48] extended this idea, and developed an evolutionary model combining cultural transmission with costly signalling in a population, finding that for even modest costs the system moved towards universal belief. Wildman and Sosis [123] subsequently extended the model, to address the fact that both stable equilibria are binary states, finding that the incorporation of group differentiation allowed subgroups to persist.

Signalling games have also been extended to provide models of other observed human behaviour, for example Austen-Smith and Fryer Jr. [6] attempted to explain the observed poor academic attainment of some social groups by positing a multiple audience signalling game. They found that the introduction of a secondary signalling game with a peer audience, alongside the prototypical Spence model introduced a pooling equilibrium. Subsequent empirical work by Fryer Jr. and Torelli [32] has provided some support for this idea. Along similar lines, Feltovich et al. [31] examine an observed failure by high quality types to signal as would be anticipated, introducing the concept of countersignaling in scenarios where there is noisy leakage of type information. They found that where there is added noisy type information available, separating equilibriums exist where high quality senders signal either as low quality, or not at all.

2.2.2 *Bayesian Decision Theory and Expected Utility*

Decision theory is the theory of rational decision making [83], this contrasts with game theory which is concerned with strategic decision making. In the broadest sense, the field can be divided into two types of theories: normative, and descriptive. Normative theories are those which attempt to give the rational answer to a decision problem, descriptive or behavioural theories focus instead on characterising the process of human decision making. In this instance, the particular concern is with theories of decision making under uncertainty.

1 In fact, an infinite number of them.

Underpinning almost all theories of decision making, and much of economic theory in general is the concept of expected utility, originally proposed by Bernoulli [9]. This casts decisions as choices between lotteries or gambles, with differing payoffs and probabilities.

Under this model, the expected utility of any gamble is a function of the probability of the outcomes, their utility to the gambler, and the gambler's risk aversion. Essentially this is an extension of the expected value criterion, which assumes that the expected value is based only on the probability and objective value of outcomes. By contrast, the utility framing is a subjective measure, allowing differing preferences between gamblers. Von Neumann and Morgenstern [120] later formalised the theory, defining rational decision as acting to maximise expected utility, where an individual's preferences are shown to fulfil four axioms, namely completeness, transitivity, independence, and continuity. Completeness requires that for any two lotteries A and B, the decision maker prefers one to the other, or is indifferent. Transitivity requires that if A is preferred to B, and B is preferred to C, then A is also preferred to C. Continuity states that given a scenario as in the transitivity axiom, there is some combination of lotteries A and C where the decision maker is indifferent between that combined lottery and B. Finally, independence maintains that if one were to prefer gamble A to B, that preference holds if both are combined with lottery C.

While vastly influential, the expected utility theory has been substantially criticised, generally for failing to predict real behaviour. Allais [2] attacked the independence axiom in particular, suggesting that in some scenarios people's choices would be inconsistent where expected utility implies otherwise. A number of studies (e.g. [80, 15]) have since supported the intuition to some extent.

More recently, support for some aspects of the expected utility theory, particularly the concept of utility as a common currency for comparison, has come from neurology, for example following work by Platt and Glimcher [86], Padoa-Schioppa and Assad [81, 82] report neuronal firing corresponding to economic value in decision making tasks undertaken by monkeys, while Christopoulos et al. [19] found similarly indicative results for risk aversion. The suggestion implicit in the model proposed here, that this also applies to social judgements, is less investigated, although both Watson and Platt [122], and Willis et al. [124] found that lesions in the brain area[2] identified by Padoa-Schioppa and Assad lead to abnormal social judgements in humans and primates.

Bayesian decision theory, as expounded by Robbins [89] applies Bayesian inference to the process of decision making under some degree of uncertainty, on the basis that the decision is a repeated one. The central idea is relatively straightforward, and assumes that the

2 The orbitofrontal cortex.

loss or gain of some action to resolve a decision is contingent on an unknown parameter. To solve the problem, the decision maker chooses whichever action will minimise the risk, where the risk of an action is $\sum_i \lambda(a_j|w_i)P(w_i|x)$, i.e. the loss incurred for taking action a_j given that the true state of the world is w_i, multiplied by the belief that this is the true state of world given evidence x, summed across all possible worlds. Essentially this is identical with expected value, with Bayesian style probabilities. This allows an additional process of inference to progressively update the distribution from which $P(w_i|x)$ derives, as new evidence is obtained after each decision.

This approach has been used in a wide variety of scenarios, for example McNamara and Houston [69] have applied statistical decision theory as a framework for understanding animal learning[3], while Harsanyi [44] has derived an ethical framework from the principles. Less controversially, in contexts where optimality is desirable as an outcome, Dorazio and Johnson [27] have used Bayesian decision methods in combination with Markov chain monte carlo methods (MCMC) to solve complex waterfowl habitat management problems, and Kristensen [61] has developed robots which utilise Bayesian decision analysis to plan sensor operations.

As with standard expected utility, the Bayesian approach can be criticised, in this case on the grounds of plausibility. The question of plausibility arises from the suggestion that Bayesian inference is in some way a model of human inductive reasoning, as argued by some branches of cognitive science. For example, Tenenbaum et al. argue for the Bayesian approach as a top-down model of inductive reasoning in humans [110, 41], a general approach criticised by Bowers and Davis [13] as unfalsifiable, overcomplicated, and relying on an unrealistic conceptualisation of the brain as optimal. Miller [71] also applied similar criticism to claims by Gallistel [34] that Bayesian inference better characterises learning as opposed to associative conditioning type models, suggesting that this relies on an assumption of optimality which is unfounded.

2.2.3 *Descriptive Decision Theory*

Arguably the most significant criticism of theories of decision making, is their failure to correspond to empirically observed decision making [4]. This was probably first raised by Simon [98], who proposed that the apparent divergence derived from a tendency to satisfice, rather than optimise. This suggestion rests on the not unreasonable assumption

3 Although they note that this is in the sense of how animals 'should' learn, rather than how they do learn

4 This critique is not unique to decision theory, and has also been levelled at game theory (e.g. Fehr and Fischbacher [30] on the irrational altruism of humans playing the prisoners' dilemma).

that people do not have unlimited cognitive capacity (i.e. bounded rationality [99]), and hence use heuristic means to make decisions, namely by choosing the first 'good enough' option. Simon suggests that this process nevertheless leads to the optimal solution is most cases.

Subsequent work on descriptive theories largely follows the same framework in assuming that in reality, human decision making is a heuristic process. Tversky and Kahneman [116] developed three heuristics to explain observed systematic errors in reasoning - representativeness, availability, and anchoring. Representativeness suggests that when asked to judge how related one object or event is to another, they do this based on the extent to which they resemble one another - crucially they will ignore additional, better information when available. Availability claims that when tasked with estimating probabilities, people will rely on the ease with which they can call examples to mind (note that this might be considered an example of satisficing). Finally, anchoring proposes that when estimating, people start with some initial value and progressively update from there, i.e. they will tend to overweight prior evidence at the expense of new information.

Subsequently, Kahneman and Tversky [55], Tversky and Kahneman [117] also identified framing effects, which imply that the decisions people make are impacted by the fashion in which the problem is presented. The essential outcome from these findings is that people are risk seeking when faced with outcomes framed as losses, but risk averse towards gains, and regard any loss as greater than an equivalent gain. The impact of framing in itself has been shown to be significant, for example Toll et al. [114] found improved abstinence rates in smoking cessation where quitting was framed as a gain, and NICE recommend considering the framing of treatment outcomes when presenting options to patients [72].

Prospect Theory (PT) [54] attempts to provide a decision rule accounting for the heuristic nature of decision making and incorporate framing effects, which successfully explains many perceived failures of rationality. A revised version, CPT [118] addressing a violation of first order stochastic dominance possible under the original formulation, extends the theory to allow decisions with more than two options, but sacrifices the editing phase. Camerer [17] reviews a number of successes in explaining apparent anomalies with CPT, and argues that should replace expected utility in general usage. Thaler [112] regards the theory as promising, but points out that it is in many ways incomplete, citing the lack of explanation as to how people construct frames as an example of this.

A significant weakness of CPT as a general theory of decision making is that it fails to account for behaviour under intertemporal choice, or rather does not attempt to address it. Generally, intertemporal choice is assumed to be underpinned by the Discounted Utility (DU) model of Samuelson [95], which proposes that the value of a thing right now is greater than the value of it at some point in the future (jam today has more utility than jam tomorrow), following an exponential relationship. A more nuanced view of this has been proposed by Ainslie [1], suggesting that the relationship is hyperbolic rather than exponential. Both models however fail to explain several inconsistencies, for example Thaler [111] found that discounting rates were different between gains and losses. Loewenstein and Prelec [65] report a number of additional inconsistencies that are not adequately resolved by DU models, and propose an alternative along the lines of CPT to resolve them while retaining the capabilities of Kahneman and Tversky's model in immediate term choices.

Part II

SIMULATION AND EXPERIMENTS

3 SIMULATIONS AND MODEL

3.1 MIDWIVES AND WOMEN

Before describing the model, it is perhaps useful to sketch the interaction it is intended to capture. In this case, this is the sequence of appointments a woman has with their midwife, or midwives over the course of a pregnancy.

Following the NICE guidelines [74] then, first contact is usually at the booking appointment, where the midwife will ask a number of standard questions, about health and lifestyle. In this instance the only question of interest is about routine alcohol consumption, which may be asked in terms of average units per week, or by using one of a number of screening tools. While midwives may raise this on subsequent appointments if they feel it is warranted, the general case is that this is raised at the booking appointment only. Based on the information provided by the woman at that time, or volunteered later, the practitioner may choose to refer the woman to a specialist service (e.g. a substance misuse midwife), for further evaluation and perhaps treatment.

Taking this interaction at the most abstract level, the woman sends a signal that may or may not convey some information known only to them, and the midwife takes some action in response to the signal.

In this scenario, if it is assumed that referring women who drink above some threshold always yields a better health outcome, the rational course of action is for women to be honest about their drinking habits, since doing so can only be to their benefit. However, people do not reveal everything to their health care providers, despite an apparent unconditional gain from doing so, or substantial, if often delayed, penalties for holding back.

Returning to the real encounter, it seems reasonable to suppose that the midwife responds in some social sense to the reported alcohol consumption and may express disapproval based on their own views. In the model this is interpreted as being a form of social cost, applied to those who drink heavily while pregnant. Rational behaviour on the part of the woman is then dependent on the balance between the social cost of honesty, and the health cost of dishonesty.

For the midwife's part, failing to identify the woman as needing further treatment potentially risks the health of the baby which is emotionally taxing, and carries a risk of legal action. On the other hand, referrals represent an additional cost to the organisation, and generating too many false positives will reflect badly on the midwife.

3.2 THE SIGNALLING GAME

The interaction described above is analogous to a pair of repeated signalling games, where both players are privately informed of their type rather than just the signaller (similar to the courtship game).

To simplify the situation, no particular effort is made to deduce realistic payoffs, women and midwives are assumed to not recognise each other[1], and the number of types is limited to three. For the women, these are light drinkers who drink at a level that is both socially acceptable and poses no risk to their baby, moderate drinkers who are putting the child at some risk and might well be frowned upon, and heavy drinkers who need treatment in order to have a healthy child. Types are similar for midwives, and reflect how harshly they judge different drinking types - harsh midwives regard even moderate drinking as completely unacceptable, moderate midwives only object to heavy drinking, and nonjudgemental midwives pass no judgement on anything. Social judgement type does not in this case reflect the decision to refer, since it is assumed that all midwives are well versed in the risks to mother and child and differ only in how they express their decision.

Figure 1 shows a simplified version of a single round of the extensive form signalling game (the complete tree is available in appendix A). The first two moves are made by nature, assigning a drinking type to the woman, and a social judgement type to the midwife. The woman then signals one of the three drinking types, and the midwife chooses to refer or not, based on the signal and their beliefs about which types are likely to have sent it.

Both players receive a common interest payoff (table 2b), based on the health outcome, which is contingent on the true drinking type of the woman, and referral choice. Midwives' receive an additional payoff (table 2c) reflecting the additional cost to the institution of a referral. Women take a social cost based on their signal, and the social judgement type of the midwife (table 2a).

3.3 DECISION PROBLEM FRAMING

Transformation from a game theoretic framing to a decision problem is relatively simple. Moves which would otherwise belong to the other player are in effect assigned to nature, with the outcomes treated as lotteries. This shifts the emphasis from the problem of second-guessing the other player, to inferring the probability distribution governing the lotteries, and using that information to make the decision on what signal to send, or act to take.

1 This is of course unrealistic, and perhaps corresponds better to a scenario where each woman has a hundred babies, or each appointment is a subsequent generation.

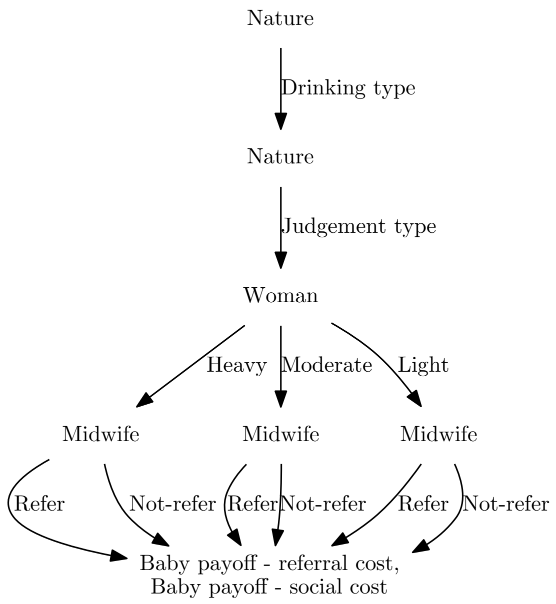


Figure 1: Simplified signalling game tree

		Woman		
		Heavy	Moderate	Light
Midwife	Harsh	0, -2	0, -1	0, 0
	Medium	0, -1	0, 0	0, 0
	Low	0, 0	0, 0	0, 0

(a) Social cost

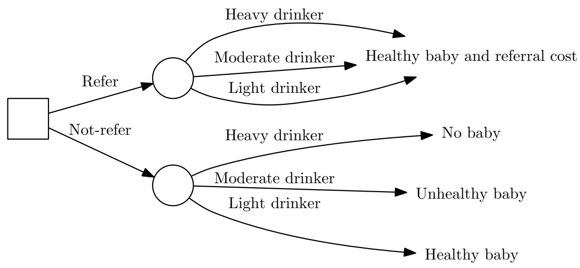
		Woman		
		Heavy	Moderate	Light
Midwife	Refer	2, 2	2, 2	2, 2
	Don't refer	-2, -2	-1, -1	2, 2

(b) Health outcome payoff matrix

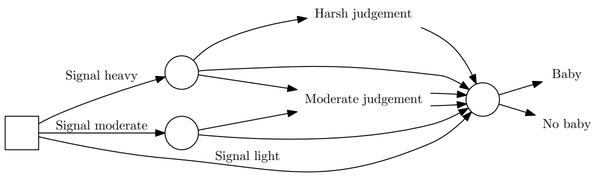
		Woman		
		Heavy	Moderate	Light
Midwife	Refer	-1, 0	-1, 0	-1, 0
	Don't refer	0, 0	0, 0	0, 0

(c) Referral cost matrix

Table 2: Payoff matrices



(a) Midwife's perspective



(b) Woman's perspective (heavy drinker)

Figure 2: Interaction diagrams for both player types

3.3.0.1 *Women's Decision Problem*

From the perspective of the women, there are two lotteries to be considered, governing the social judgement cost applied in response to their signal, and whether or not they are referred. The outcomes of the two lotteries are assumed to be independent, in that the level of social judgement cost applied does not impact the decision to refer. Figure 2b shows the structure of the decision problem for a heavy drinker, where circles represent 'chance' nodes, and squares are decision nodes.

3.3.0.2 *Midwives' Decision Problem*

Midwives have an ostensibly less complex problem, with only one lottery to consider. The problem in this case is to estimate the likelihood of the true state of the world (i.e. the true type of the woman), given that they have observed a particular signal.

3.4 ESTIMATING RISK

Thus far, a significant feature of both formulations has been effectively handwaved, namely that of probabilities. In general, probabilities in decision problems are assumed to be both known, and accurate. In practice, this is not usually the case outside of the laboratory. The probabilities of events are more usually estimates, or following the

Bayesian perspective, a measure of belief in the likelihood of an outcome. In this case, there are a number of discrete lotteries to choose amongst on any given round of play. Aside from the distribution of midwife types, the probability distributions change over time; for example, the probability of a referral from a particular midwife given any signal is, in part, dependent on their historical interactions.

In this model agents use Bayesian updating to attempt to improve the accuracy of their beliefs. Given the discrete nature of agent types, and actions, coupled with a desire for tractability of the simulation, the Dirichlet distribution is employed. The probability density function takes the form -

$$D(\Theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \Theta_i^{\alpha_{i-1}}$$

Where $\alpha = \{\alpha_1 \ldots \alpha_k\}$, k is the number of discrete types, $\Theta = \{x_1, \ldots, x_{k-1}\}$ all more than zero and summing to less than 1, and α_i is the prior probability for type i, which can be thought of as a pseudo-count.

The distribution is particularly convenient, in that to infer the probability of a new observation being of some particular type becomes simply -

$$P(x = j|D, \alpha) = \frac{\alpha_j + n_j}{\sum_j (\alpha_j + n_j)} \tag{1}$$

Where n_j is simply the count of occurrences of type j, so that the belief that a new observation is of type j the number of times that type 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 with no particular concern as to their order.

3.4.1 *Midwives' Beliefs*

Midwives hold beliefs purely about the relationship of signals to types, i.e. whether a low type signal truly means high, or medium. This entails three separate distributions (one per type), with the same number of parameters. After a round is played, the true type of the woman is revealed to the midwife agent, and the parameter corresponding to that type, in the distribution corresponding to that signal is incremented.

Midwives' priors are of the form $s_i\{\alpha_l, \alpha_m, \alpha_h\}$, where s_i is the set of beliefs about the meaning of signal i, $i \in [l, m, h]$, and α_n is the strength of prior belief that signal i was sent by type n.

3.4.2 *Women's Beliefs*

A slightly more complex structure of beliefs is necessary for the women, who are required to estimate the distribution of midwife types, and in addition, how likely each signalling choice is to lead to referral.

In the first case, a simple three parameter distribution is used, in the second the structure is similar to that used by the midwife agents, but with a binomial distribution (refer, not-refer). Note that while the belief updating for these two lotteries is distinct, subsequent reasoning on them treats them as a combined lottery, i.e. given that the probability of a type m midwife is p, and the probability of referral given signal m is q, the total probability of being referred, and harshly judged, is pq.

Women's priors take the form $\{\alpha_l, \alpha_m, \alpha_h\}$ for the distribution of midwife types, where α_h is the prior belief that the next midwife will be a harsh judge, and $s_i\{\alpha_r, \alpha_n\}$ where α_r is the strength of prior belief that signal i will lead to referral, and α_n that it will not.

3.5 DECISION RULES

Decision rules codify strategies for making decisions, given known payoffs and risks. In the broadest sense, they fall into one of two classes: normative, and descriptive. Normative decision rules are concerned with how to rationally make decisions, whereas descriptive rules attempt to represent how decisions are actually made.

3.5.1 *Bayesian Risk*

A common normative strategy is Bayesian risk minimisation. Under this method, an agent is viewed as attempting to minimise the risk of a decision, where the risk of some particular decision is the sum of each outcome, multiplied by the probability of it occurring. Or formally -

$$R(a_i|x) = \sum_{j=1}^{c} \lambda(a_i|w_j)P(w_j|x)$$

Where $\lambda(a_i|w_j)$ is the loss incurred for taking action a_i, when the true state of the world is w_j, and $P(w_j|x)$ is as in equation 1. To decide what to do, the agent then chooses a_i to minimise $R(a_i|x)$ given x, the set of observations about the world. In this case, the loss function is simple the negation of the payoff function, and the set of actions and world states is sufficiently small that this can be done exhaustively.

3.5.2 *Cumulative Prospect Theory*

Rather than the psychologically more interesting PT, the CPT decision rule is used in this instance, because of the requirement for women to evaluate more than two 'prospects'. A prospect in this instance is a paired outcome and probability, and the set of prospects for an action hence define the outcome space. CPT introduces the concept of a probability weighting function, which underweights small probabilities, and overweights large ones in an effort to capture the tendency of humans to treat high probability events as sure things, and small probabilities as 'never going to happen'. A number of different weighting functions have been proposed, but in this instance the original formulation by Tversky and Kahneman [118] is used. This distinguishes between weighting for gains, and losses -

$$\begin{aligned} w^{+}(p) &= \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{\frac{1}{\gamma}}} \\ w^{-}(p) &= \frac{p^{\delta}}{(p^{\delta}+(1-p)^{\delta})^{\frac{1}{\delta}}} \end{aligned}$$

Where p is the unweighted probability, and γ and δ are the weights for gain and loss probabilities respectively. Humans have also been observed to value gains and losses differently, with a loss being 'worse' than the equivalent gain is 'good'. This entails a transformed value function -

$$v(x) = \begin{cases} f(x) & if\, x > 0 \\ 0 & if\, x = 0 \\ g(x) & if\, x < 0 \end{cases}$$

Where,

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

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

And α, and β are respectively the power of a gain, and a loss. The CPT value of outcome x is $v(x)w^{+}(x)$ if $x \geq 0$, and $v(x)w^{-}(x)$ otherwise. For an action the CPT value is the sum of the value of the prospects of that action, as in the Bayesian risk model. The decision rule then requires the agent to choose the action which maximises the prospect theory value.

3.6 IMPLEMENTATION

The model was implemented in python, a freely available and widely used interpreted programming language. Python does introduce some limitations on the speed of the implementation, which have been mitigated to some extent by introducing multiprocessing to allow running of experiments across logical CPUs. In addition, the pypy runtime, with JIT compiler was used as a drop in replacement for the standard python implementation. This effectively doubled the speed of execution, without requiring modification to the code, or reimplementation of costly parts in a compiled language. The code is relatively naive, and could be optimised to realise further speed increases and enhance scalability. Alternatively, porting to a language engineered with a focus on speed (e.g. Julia) would perform a similar function.

To facilitate replication, extension, and general borrowing of the model and approach, the code for both classes of agent, together with that for running the experiments has been made available under the GPL via github, at `https://github.com/greenape/disclosure-game`.

4 EXPERIMENTS

4.1 TRUST AND CASELOADING

The significance of relationship building in eliciting disclosure is highlighted throughout the midwifery literature, with a strong recognition that building up trust between practitioner, and woman facilitates honest communication. Phillips et al. [84], report that a number of appointments may be necessary to build up a sufficiently strong relationship for women to be willing to disclose potentially damaging information, suggesting a qualitative trend towards increased honesty across appointments.

Some studies have suggested that a caseloading approach, where women have a dedicated midwife throughout the course of pregnancy offers an advantage in this respect (e.g. [28]). No studies providing quantitative data on the extent to which this impacts disclosure were available, instead describing a qualitative trend in behaviour.

On this basis, the model should capture the impact of the developing relationship, with honest signalling increasing with the number of appointments, and this trend should be amplified where women are assigned to a single midwife for all appointments.

In terms of the dynamics of the model, this would be expected in a situation where the population of women have heterogenous prior beliefs about the midwives, which do not reflect reality. Where the makeup of the midwife population is broadly sympathetic (i.e. contains only a minority of types which impose social penalties), this would be anticipated to produce the reported increase in disclosure.

Following the trend described by Alvik et al. [3] in comparing concurrent and retrospective accounts of alcohol consumption in the antenatal period, it would also be expected that heavy drinkers would show an increased tendency to underreport alcohol usage compared with moderate drinkers under either condition. The structure of the model, which makes a referral advantageous to midwives for all but light drinkers should capture this, since heavy drinkers are able to signal a lower consumption while still avoiding negative payoffs.

4.1.1 *Hypothesis*

A. Honest signalling by women will increase with the number of appointments.

B. Caseloaded women will show a faster rate of increase as compared to non-caseloaded.

Parameter	Value
Number of women	1000
Number of midwives	100
Caseload weight	10:1
Runs	1000
Number of harsh midwives	5
Number of moderate midwives	15
Number of non-judgemental midwives	80
Midwives' priors	10:1

Table 3: Parameters

C. Heavy drinkers will signal less honestly than moderate drinkers.

D. Harsh judgement by midwives will lead to less honest signalling.

4.1.2 *Design*

In this experiment, two models of care provision were contrasted. In one, the individual woman's experience was of seeing a randomly chosen midwife on each appointment. The second scenario randomly assigned women to a particular midwife for all appointments, such that each midwife had the same patient load.

Midwives' use homogenous priors biased towards honesty, such that the dirichlet distribution $\{\alpha_{low}, \alpha_{moderate}, \alpha_{high}\}$ for signal i has $\alpha_i = 10$ and 1 otherwise. Women's beliefs are heterogenous, and stochastic such that $\alpha_k = 10$ (i.e. each α_i is a pseudo-count out of ten), with equal numbers of drinking types. The degree of randomness was controlled to facilitate comparability between conditions by using a common random seed for each condition that varied between runs.

Table 3 gives the full set of parameters used in the design.

4.2 TRUST AND BELIEF

The dynamics of the model suggest a mechanism for an intuition about trust, namely that it is necessarily a two way process. One conceptualisation of the beliefs of agents, is that they are effectively modelling trust in an outcome. This is particularly marked in the case of the midwife agents, where high belief in a particular signal honestly implying a type represents a bias towards trusting the signal. Strong prior beliefs reduce the weight of subsequent evidence, and where those beliefs are strongly in favour of signals being honest, will yield more trusting behaviour. This impacts on the long term behaviour of women in the model, because the advantage of signalling a lower drinking type is contingent on receiving a referral despite the signal.

This suggests that the dynamics of the model should be sensitive to the strength of midwives' prior beliefs, with strong bias towards honesty leading to greater swings towards disclosure, and sufficiently weak bias allowing the meaning of low type signals to be co-opted.

4.2.1 *Hypothesis*

A. Stronger priors biased towards honest signalling in midwives will result in more honest signals from women.

4.2.2 *Design*

To investigate the impact of strength of priors on change towards an honest signalling strategy, simulations covering increasingly strong priors in favour of honesty were run and the average change in honest signalling for heavy and moderate drinkers recorded. Beginning from an uninformed prior (i.e. $\alpha_i = 1$ for all types), up to entirely entrenched priors in excess of the average number of appointments ($\alpha_i = 50$ for signal i, and 1 for all others). A hundred random starting conditions were generated, with priors increasing from 1-5, and then in increments of 5 up to a maximum of 50.

4.2.3 *Balancing Beliefs*

To further explore the impact of midwives' priors on honesty, an additional experiment examined the effects of different combinations of bias in the presumed meaning of signals on change in frequency of honest signals. Combinations were generated by producing permutations of the each set of priors used in the preceding experiment, such that a prior structured {{5, 1, 1}, {1, 5, 1}, {1, 1, 5}} yields additional permutations {{1, 5, 1}, {5, 1, 1}, {1, 1, 5}, {{1, 5, 1}, {1, 1, 5}, {5, 1, 1}} etc. for a total of 64 sets of priors. Other parameters and methodology were as in the preceding experiment.

4.3 NORMATIVE VS. DESCRIPTIVE

In addition to the homo economicus type model of the Bayesian risk minimising rule, the CPT was also evaluated using the experimental design of the caseloading experiment. The motivation in this instance is to test the feasibility of applying a more complex decision rule, with better behavioural plausibility, and investigate whether this yields an improved fit to real world behaviour. Appropriate parameters for the

Parameter	Value
Midwives' priors	10:1
Women's priors	random
Rounds	100
Runs	1000

Table 4: CPT decision rule parameters

Parameter	Minimum	Maximum
Priors	1	50
Midwife type proportion	1	100
Woman type proportion	1	100

Table 5: Parameter space

CPT decision rule are subject to some debate [76, 78], and a meta-study by Byrnes et al. [16] found that risk aversion is significantly affected by gender, with women being more risk averse than men. On this basis, the parameters used were those determined by Booij et al. [12] for women in particular (table 4), with other parameters as in table 3.

4.3.1 *Design*

4.4 SENSITIVITY

A global sensitivity analysis was conducted on change in frequency of honest signalling by heavy and moderate drinkers, using the GEM-SA [58]. Parameters were sampled based on LP-Tau sampling of 250 points within the parameter space defined in table 5 with the populations rescaled to 1000 women, and 100 midwives. The analysis considered only the 'core' Bayesian decision model, excluding caseloading, and using the same priors regime for women as in the preceding designs.

Part III

RESULTS AND DISCUSSION

5 RESULTS AND ANALYSIS

5.1 RESULTS

5.1.1 *Trust and Caseloading*

The results of the trust experiment broadly speaking follow the qualitative trend described by Phillips et al. [84], with honest signalling tending to increase over time. Interestingly, this dynamic does not hold exactly for heavy drinking types, who show a tendency to signal as moderate drinkers in preference to their true type as shown in figure 3. The impact of caseloading is counterintuitive, and both figure 3, and 4 suggest that effect is to decrease honest signalling rather than the anticipated improvement.

Visually, the relationship between number of appointments and frequency of signal appears to be non-linear, with the plots suggestive of a logarithmic relationship. This reflects the initially unstable state of the system, where a majority of moderate and heavy drinkers initially signal as low types before moving to a new balance. The instability arises because the heterogenous priors of the woman do not initially reflect the true type balance of the midwife population, or a realistic likelihood of referral for a given signal. Belief on type distribution is essentially identical between the caseloading conditions, with a significant difference visible in beliefs on signal referral (figure 5).

Linear models were constructed on signal frequencies with number of appointments, a logarithmic transform of appointments, caseloading as a boolean variable, and interactions between caseloading, and appointments as predictors. For heavy drinkers signalling as light drinkers[1], the model has significant results ($F_{5,197,994} = 115,600$, $p < 2.2 \times 10^{-16}$) and an R^2 value of 75%. All four predictors are significant at $p < 2.2 \times 10^{-16}$. Stepped Akaike information criterion (AIC) reduction retained all terms, supporting the suggestion that the relationship is non-linear. The intercept for the regression line was 0.499, with the full coefficients given in table 6. An analysis of variance (ANOVA) suggests that the majority of variation is attributable to caseloading, and secondarily to the transformed appointments predictor.

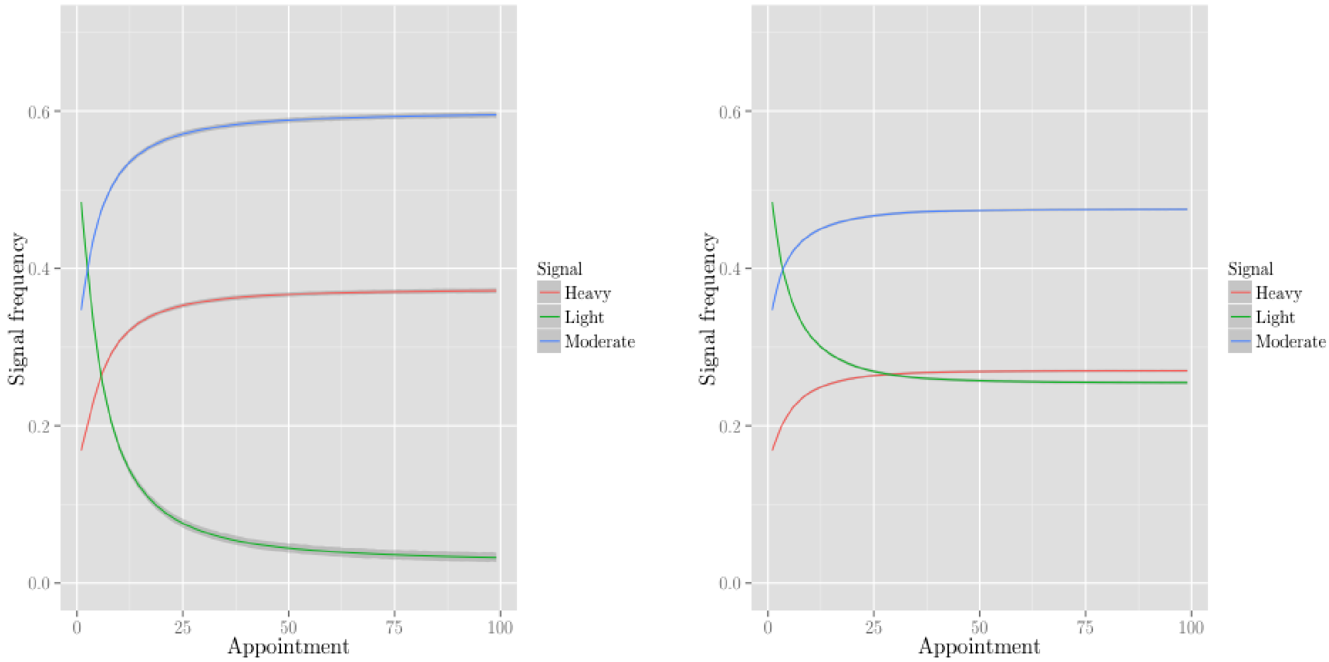
The plots suggest that all conditions show less honest signalling by heavy drinkers as compared to moderate drinkers. Both groups have a similar tendency to signal as low drinkers, but heavy drinkers are more likely to signal as moderate, whereas the honest signalling

1 Full coefficients and R^2 values are reported in appendix B for all six models, which are similar.

Predictor	Coefficient
Appointment	0.0022
ln(Appointment)	-0.1465
Caseloading	-0.0198
Appointment with caseloading	-0.001
ln(appointment) with caseloading	0.0727

Table 6: Regression coefficients for frequency of heavy drinkers signalling as light

strategy dominates for moderate drinkers. Moderate and heavy drinkers have equal opportunity to be deceptive, but heavy drinkers have more change of doing so without negative consequences.

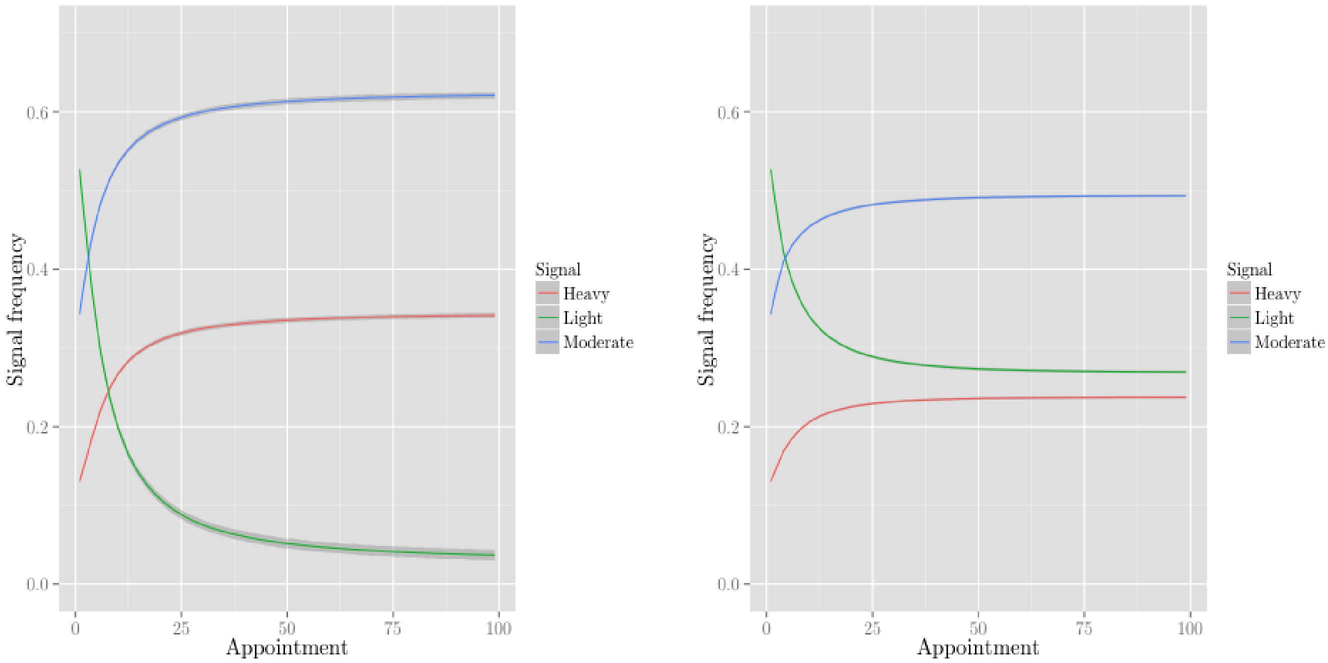


(a) Without caseloading

(b) With caseloading

Figure 3: Mean frequency of signals by heavy drinkers with 95% confidence limit

The caseloading scenario also allows a breakdown of signalling behaviour by assigned midwife type. Figure 6 show the frequency of honest signals by moderate and heavy drinkers, broken down by assigned midwife type. The general tendency is towards more honest signalling with less harsh social judgement. Moderate drinkers actually signal more honestly with a commensurately moderate midwife; more honest signalling would be expected here in any case, since moderate drinkers receive no negative payoff from an honest signal



(a) Without caseloading (b) With caseloading

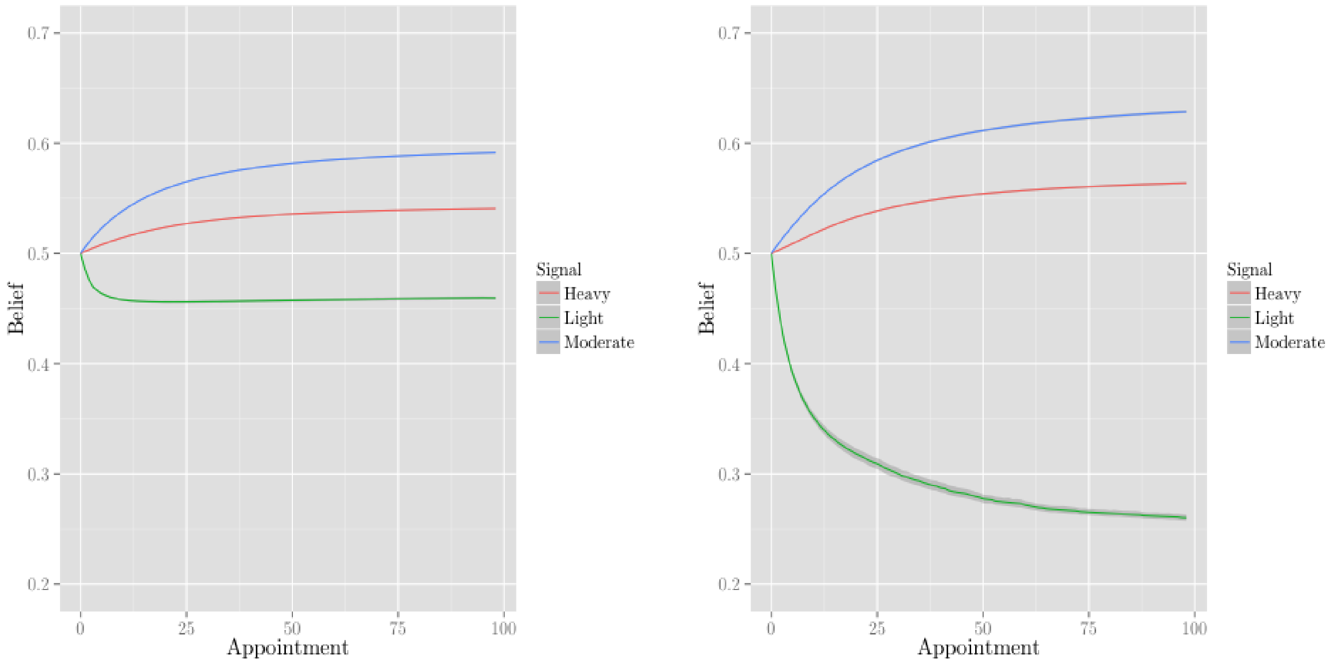
Figure 4: Mean frequency of signals by moderate drinkers with 95% confidence limit

in this context. A possible explanation for the lower honest signalling with the non-judgemental midwife, is that this allows moderate drinkers to exaggerate their drinking without penalty to ensure a referral. In both drinking types, harsh social judgement is associated with a greater tendency to signal as low drinkers, which suggests that if the motivation is to maximise the chance of identifying that women are not truly low drinkers, midwives should be as non-judgemental as possible.

The results suggest that light drinkers will always signal honestly, since there is nothing to be gained from an exaggeration without the motivation to gain a referral. There is however something to be lost, or the perception of that. Were the belief in non-judgemental midwife types to reach 1, then there would be no perceived loss in sending false signals for these types and they would choose stochastically.

5.1.2 *Trust and Belief*

Figure 7 shows average change in frequency of honest signals by heavy and moderate drinkers after 100 appointments with respect to the bias of midwives' priors towards honesty. As anticipated, the results suggest that a stronger bias towards interpreting signals as honest is associated with more honest signalling. The relationship ap-



(a) Without caseloading

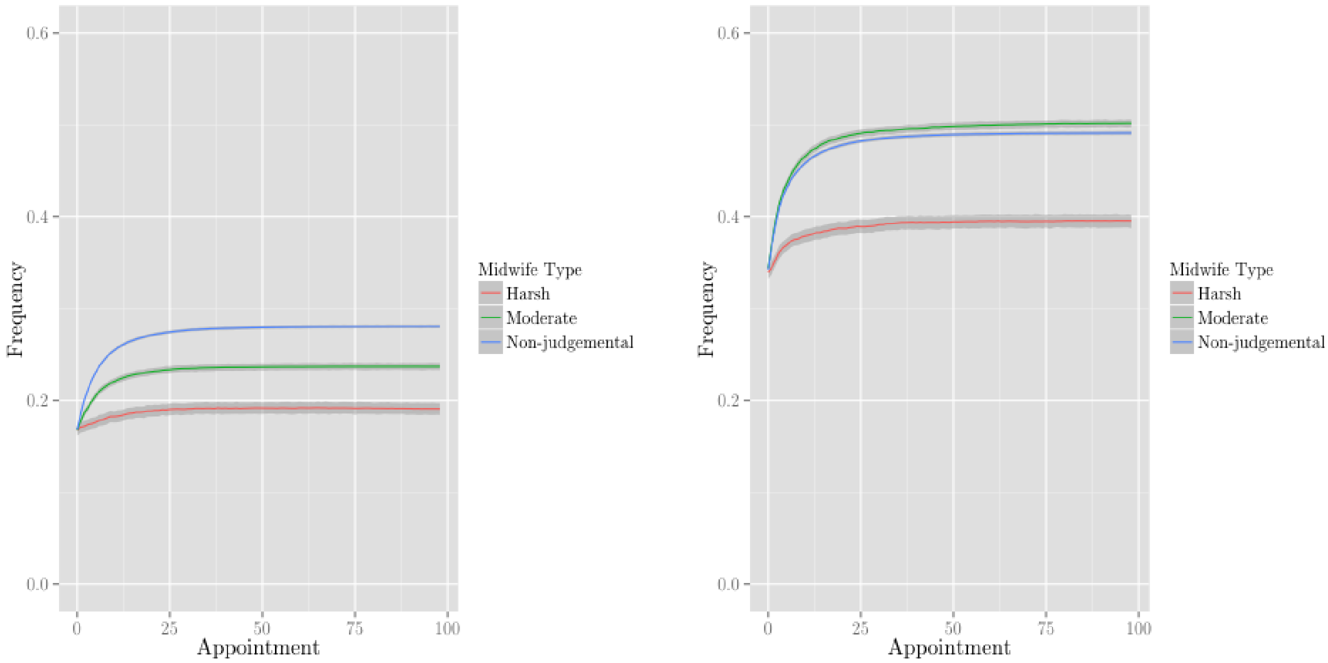
(b) Without caseloading

Figure 5: Average strength of women's belief that a signal will lead to referral, with 95% confidence limit

pears to be non-linear, stronger for moderate drinkers, and weaker than might be expected. Contrasting with figure 8 which shows the change in low type signals, is instructive, suggesting that much of the change in signalling behaviour for heavy drinkers is actually to switch to reporting as moderate rather than heavy. Broadly speaking, the significant change in honest signalling appears to stem from midwives' having a degree of belief in signals being honest which is sufficient to survive dishonest signals, and allow women to update their beliefs on the likelihood of referral.

5.1.3 *Balancing Beliefs*

To examine the contribution of combinations of biased priors to changes in honest signalling, linear models were constructed on change in honest signalling for both heavy, and moderate drinkers with significant results ($F_{9,5058} = 5058$, $p < 2.2 \times 10^{-16}$, and $F_{9,6590} = 4843$, $p < 2.2 \times 10^{-16}$ respectively) and R^2 values of 87% in both cases. An initial minimal model using the full set of priors as predictors was constructed, with interaction terms incorporated by forwards-backwards stepped AIC. The AIC process removed priors for the meaning of heavy signals, and heavy drinking as a meaning of moderate signals, and introduced several two-way interactions between priors for light



(a) Heavy drinkers (b) Moderate drinkers

Figure 6: Mean frequency of honest signals by assigned midwife type, with 95% confidence interval

and moderate signals. Predictors were consistent between the two models, which differ in their coefficients (table 7). An ANOVA suggests that the majority of variation in both cases can be explained in terms of belief that low signals are honest, and moderate signals actually imply low consumption. Taken together with the differing coefficients, this suggests that balancing prior beliefs to maximise honest signalling for both groups is a challenging multi objective optimisation problem. Given that under this model, referring moderate drinkers is still preferable, a reasonable approach would seem to be to maximise the honesty of moderate drinkers and allow pooling on the basis that the referral will still capture the majority of heavy drinkers.

5.1.4 *Normative vs. Descriptive*

Results from the CPT type agents differ substantially from their Bayesian companion, showing reduced change over appointments. Unlike the Bayesian risk minimising model, the results show no change in signalling behaviour without caseloading. However as shown in figure 9 there is a modest increase in honest signalling behaviour over initial appointments where women are caseloaded. Unlike the Bayesian model, honest signalling does not dominate for either group, with the majority tending to persist in signalling as light drinkers. Much

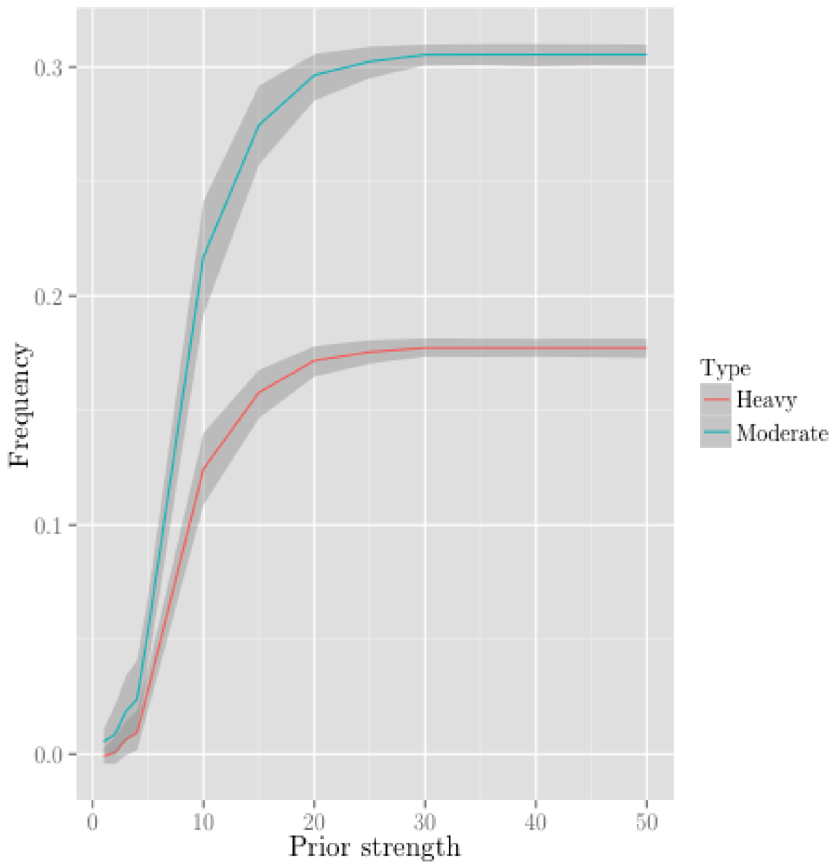


Figure 7: Mean change in frequency of honest signals after 100 appointments for varying presumptions of honesty

of this behaviour actually appears to derive from that of the midwives, with the loss aversion tending to very rapidly drive behaviour to universal referral.

As in the caseloading experiment, results for caseloaded women are also broken down by assigned midwife type in figure 10. The same trend seen in the Bayesian case is also present here, albeit at a smaller scale of change as might be anticipated from the overall results. More honesty by moderate drinkers with moderate midwives also manifests, suggesting that this is not a feature of the Bayesian risk minimising rule in particular.

5.1.5 *Sensitivity Analysis*

To examine the sensitivity to a subset of parameters, Gaussian Emulation Machines (GEMs) were built change in the frequency of moderate and heavy drinkers signalling honestly, and as light drinkers. Table 8 gives the percentages of variation which can be explained in terms of each parameter varied. The model is most sensitive to the proportion of light drinking women, which begins to exert a significant impact as

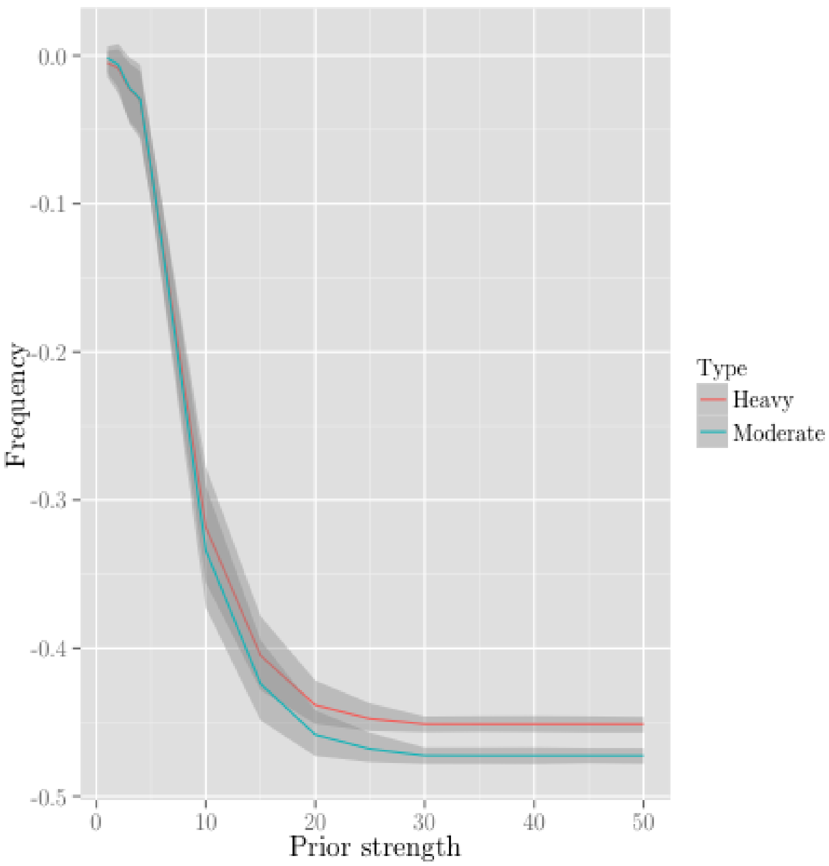
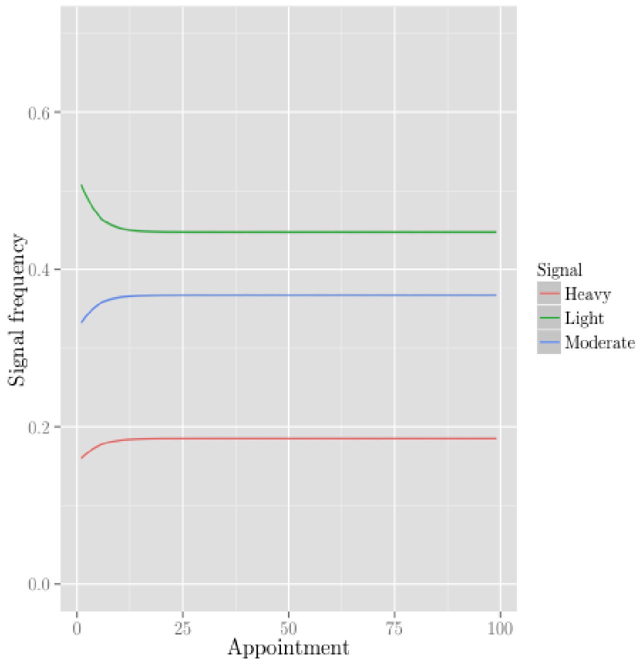


Figure 8: Mean change in frequency of low type signals after 100 appointments for varying presumptions of honesty

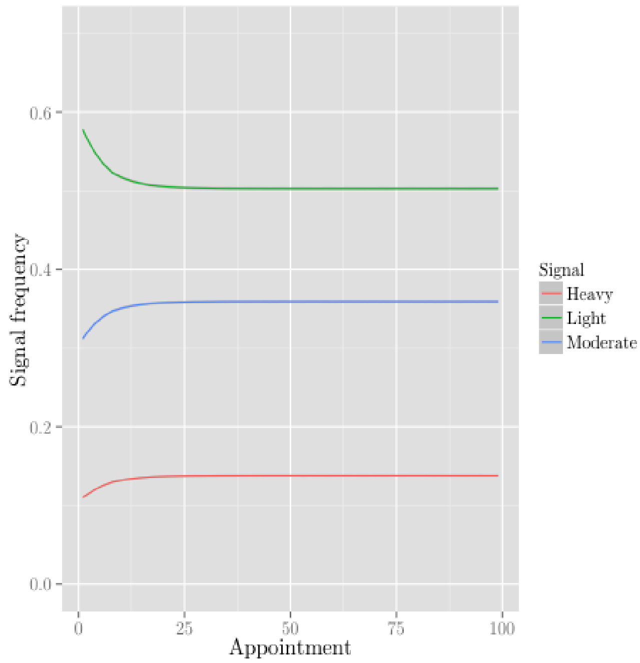
they become more than 50% of the population and are able to effectively prevent low type signals being co-opted by the other drinking types. Variance associated with midwife judgement types is low, but tends to have a higher total effect than would be expected which suggests that they have interactions with other parameters.

Predictor	Moderate Coefficient	Heavy coefficient	Moderate p value	Heavy p value
Intercept	0.0276	0.0049	$< 2 \times 10^{-16}$	0.0012
$\alpha_{l,l}$	0.0080	0.0051	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
$\alpha_{l,m}$	0.0003	-0.0049	0.0006	$< 2 \times 10^{-16}$
$\alpha_{l,h}$	-0.0044	-0.0107	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
$\alpha_{m,l}$	-0.0036	0.0079	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
$\alpha_{m,m}$	0.0071	0.0033	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
$\alpha_{l,l} : \alpha_{m,m}$	-0.0002	-0.0001	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
$\alpha_{l,h} : \alpha_{m,l}$	0.0001	0.0001	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
$\alpha_{l,h} : \alpha_{m,m}$	-0.0001	0.0001	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
$\alpha_{l,m} : \alpha_{m,l}$	-0.0001	0	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$

Table 7: Balanced belief coefficients

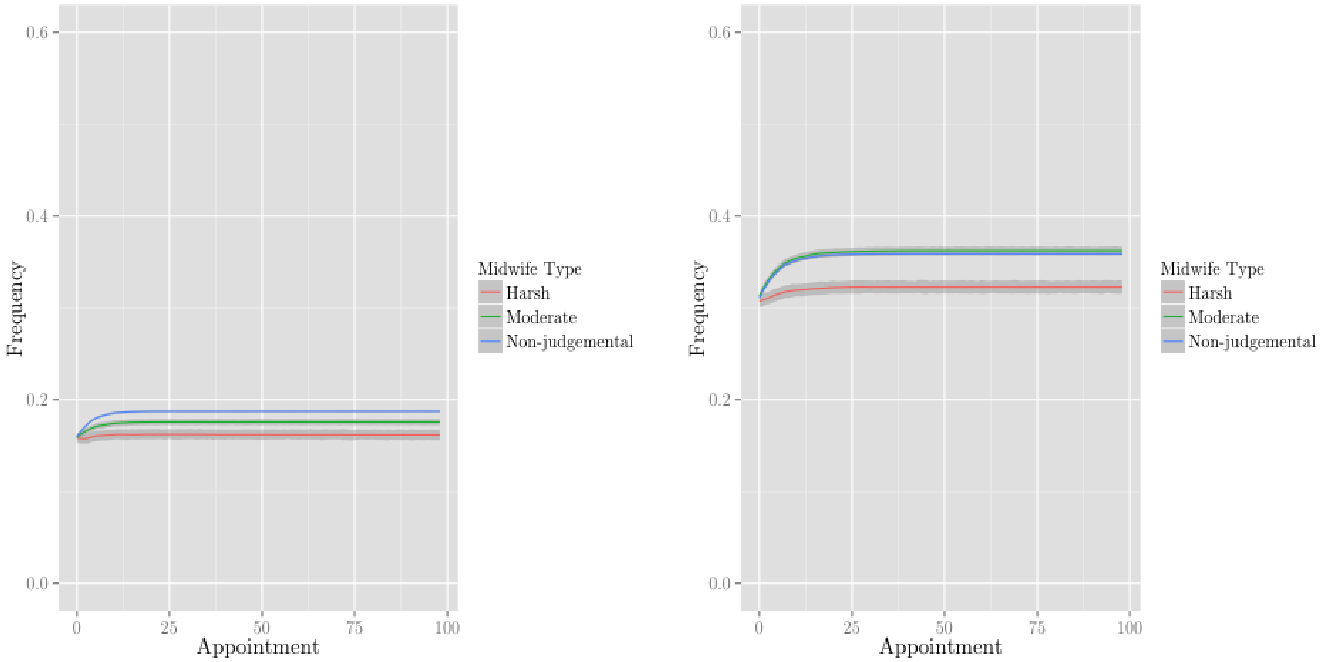


(a) Heavy drinkers



(b) Moderate drinkers

Figure 9: Mean frequency of signals by caseloaded CPT women with 95% confidence interval



(a) Heavy drinkers

(b) Moderate drinkers

Figure 10: Mean frequency of honest signals by assigned midwife type for CPT agents, with 95% confidence interval

Parameter	Moderate low	Moderate honest	Heavy low	Heavy honest
$\alpha_{l,l}$	0.66	0.47	0.92	0.86
$\alpha_{l,m}$	4.67	3.00	2.84	1.61
$\alpha_{l,h}$	3.56	3.45	3.25	1.71
$\alpha_{m,l}$	0.11	0.13	0.13	0.27
$\alpha_{m,m}$	0.29	0.93	0.18	0.23
$\alpha_{m,h}$	0.11	0.17	0.18	0.28
$\alpha_{h,l}$	0.07	0.08	0.09	0.32
$\alpha_{h,m}$	0.10	0.05	0.11	0.73
$\alpha_{h,h}$	0.07	0.31	0.08	0.25
m_l	1.24	0.60	0.64	1.57
m_m	0.60	1.06	0.68	0.55
m_h	0.15	0.44	0.24	0.92
w_l	64.57	53.65	62.82	57.88
w_m	3.74	7.40	2.81	7.39
w_h	7.02	14.96	8.52	11.97
Total	86.94	86.70	83.52	86.54

Table 8: Sensitivity analysis

6 DISCUSSION AND RECOMMENDATIONS

6.1 DISCUSSION

6.1.1 *Case Study*

The experimental results from the case study are encouraging, with the model able to reproduce two empirically observed qualitative trends in behaviour, i.e. more honesty over time, and less honest behaviour by heavy drinkers. More substantial validation of results is hampered by a paucity of real world data arising from the problem of verifying alcohol intake. This is, in part, a legal question which arises from British law assuming no rights on the part of the foetus [20], which implies that women cannot be compelled to undertake treatment or screening. More pragmatically, the evidence for harm is inconclusive, and available tests are limited in what can be detected to either very recent consumption, or long term abuse [64]. This forces a reliance on self-report screening tools, although a review by Bearer et al. [8] reported some progress in biomarkers specifically targeting pregnant women.

The results of the caseloading experiment are interesting, but may in fact reflect the variance attributable to the relative proportions of drinking types rather than the impact of caseloading in a real world situation. A possible explanation for the decrease in honest signalling is that the reduced group size, and continuity of care alters the balance of drinking types dealt with by individual midwives. Whereas without caseloading, midwives will in principle encounter an even distribution of all types across a run of the simulation, the caseloading scenario suggests that some midwives will inevitably have biased caseloads and deal with groups where light or non drinkers are a minority. This allows other drinking types to co-opt the light drinking signal, and to do so sufficiently rapidly that the midwife learns to refer for all signals before the women come to believe that the potential disadvantage of signalling honestly can be disregarded. In effect, women and midwives are lead to collude in the deception. While the unrealistic proportions of drinking types may explain much of this, the impact is not entirely implausible and reflects the emotional cost, and increased risk of burnout associated with this model of care [96].

There is also an interesting potential feedback loop arising from harsh social judgement with caseloading which may suggest an avenue for research. The model suggests that midwives who judge harshly may have higher referral rates, because they will come to distrust light drinking signals, which are in themselves encouraged by the harsher judgement. This hints at a possible topic for empirical research, looking at the relationship between referral rates by midwives and their attitudes to drinking behaviours.

6.1.2 *Method Feasibility*

The larger question here, is whether the methodology is feasible, rather than whether this specific case study is adequately modelled. Scenarios centred around disclosure abound in the health and social care context, a few examples are mental health, smoking, substance misuse, abuse, and my tendency to not go to the doctor because 'it will go away by itself'.

From a logistical perspective, the approach is certainly feasible. The Bayesian approach to updating beliefs is computationally inexpensive, particularly with the discrete classes considered here, and the relatively complex CPT decision rule performs well implemented in python, which is a relatively slow language.

The CPT model performs significantly less well in capturing reality than had been expected, much of this presumably arises from the unrealistic payoffs (addressed in the next section), but may also suggest that the estimation of probabilities is not ideal. In some senses the Bayesian updating method is actually quite similar to the availability heuristic proposed by Tversky and Kahneman [116], but lacks some of the implicit nuance in the latter. These Bayesian agents have perfect recall of their personal history, and beyond the balance between priors and evidence, take no account of the ordering or subjective significance of events. Based on results from Glöckner and Pachur [35] showing that individual variation in CPT parameters has a signicant impact on predictions, the homogeneity of the parameters may also be a contributor here.

With this said, the performance of the model at the qualitative level is promising, and a more plausible risk estimation method might represent excessive detail or be adequately dealt with by simply introducing noise to the process.

At a higher level, the basic approach is to frame the scenario as a Bayesian game, translate the game into a pair of decision problems, and then construct a simulation where populations of decision theoretic agents learn the risks by interaction. The case study suggests that this basic framework works well as an exploratory method for examining the dynamics of the system, and can form a useful basis for more detailed investigation.

6.2 LIMITATIONS AND FUTURE WORK

There are a number of limitations in the present model which potentially limit its usefulness as a tool in this context, the majority of which centre around the abstract nature of the representation. In particular, the structure of the repeated game fails in some way to capture the reality by assuming that drinking type is static for all time, and that the type is accurately revealed at the end of a round. The reality is more complex, and the evidence suggests that intervention does reduce consumption e.g. O'Connor and Whaley [79], and findings by Chang et al. [18] showing that even a thorough assessment of drinking behaviour reduced consumption. The type revelation is also more nuanced, since a more accurate appraisal of drinking behaviour requires a referral. This would, if honesty is initially presumed, would allow heavy and moderate drinkers signalling as low types to remain undetected. Introducing noisy type information, in the fashion of Feltovich et al. [31], might resolve this by relaxing the assumption that either party in the model makes decisions based purely on an past experience. Another option would be to introduce a cheap talk round preceding the subgame could be introduced as suggested by Farrell and Rabin [29]. Alternatively, the type revelation could be dropped altogether and only occur where a referral takes place, making the decision attractive for informational reasons rather than simply avoiding a loss.

This type of scenario is clearly one of intertemporal choice. Payoffs intuitively divide into an up front cost representing the perceived loss from disclosing, and a longer term payoff (e.g. a healthy baby, or not getting lung cancer), which is not captured here. This suggests that a decision rule which explicitly accounts for the observed inconsistencies in human decision making would be applicable, for example the Loewenstein and Prelec [65] model, or the less complex hyperbolic discounting model proposed by Ainslie [1].

Decisions in this model occur in a social vacuum, where women only ever interact with midwives, and their decisions are informed only by those interactions. By contrast, real decisions are informed by social interactions on both sides of the equation, and one of the more interesting aspects of disclosure to a health professional is that it represents disclosure to all agents of the organisation. These agents will subsequently act based on this information as well as their own preconceptions. Incorporating these features would require introducing a social network model, and making the identity of individuals persistent across time (by contrast, under the present model a woman and midwife could have a hundred appointments together and never know it). A potential avenue to explore here is including social interactions as a second audience, as in Austen-Smith and Fryer Jr. [6].

Finally, the utility function aspect has been effectively handwaved in this study and treated as homogenous for all agents. Kable and Glimcher [53] suggest that the values assigned to choices in decision making are learnt, and that this can be explained in terms of reinforcement learning. This suggests that a potentially fruitful avenue is to explore the incorporation of reinforcement learning models, for example the temporal difference learning model proposed by Sutton and Barto [108, 107, 109], to allow agents to learn their own utility functions from experience[1]. Simply introducing a stochastic element to the utility function might also achieve much the same end, but the cognitive plausibility and better explanatory power of the the former is intuitively appealing if less parsimonious.

Aside from structural concerns, the lack of a strong quantitative evidence base makes validation problematic, and introduces some difficulties in determining realistic proportions of types. This suggests there is a strong case for exploring alternative application domains which have a stronger qualitative evidence base to facilitate validation of the model. Arguably the best target for this would be smoking interventions, which has been extensively studied, and now incorporates physiological verification via carbon monoxide testing as standard practice in the maternity context.

6.3 RECOMMENDATIONS

Based on the results of the trust and belief experiment (5.1.1), the key practice recommendation would be to presume that women are actually being honest when they report consumption. Simulations, and evidence (e.g. Alvik et al. [3]), indicate that this is not always the case, but these results suggest that lack of presumption of honesty may lead to a breakdown of trust, and many wasted referrals. Secondarily, and echoing NICE guidance [75], that midwives' should be non-judgemental in dealing with drinking behaviour.

In terms of research recommendations, there are a number of factors discussed in the limitations that warrant particular investigation, particularly more accurate proportions of drinking types in the typical population, which the global sensitivity analysis identifies as the most sensitive set of parameters. A more challenging task is eliciting more realistic parameters for the payoffs involved, which could potentially be achieved by using hypothetical questioning to reveal preferences.

Finally, the beliefs of women are deliberately glossed over here to evade an excess of parameters. While it is reasonable to assume that these are heterogenous within the population, the distribution of beliefs is likely to have a significant impact on the dynamics of the model. Taking one example, if some proportion of light drinkers were

1 In fact, Kable and Glimcher go so far as to suggest that reinforcement learning can effectively give the expected utility of uncertain, temporally delayed choices.

certain that there were only non-judgemental midwives, they might effectively muddy the meaning of the moderate, and heavy drinking signals by signaling at random. Future work should therefore explore this aspect, which is particularly of interest in the light of designing public health campaigns, and 'nudge' type interventions.

Part IV

APPENDIX

A

SIGNALLING GAME TREE

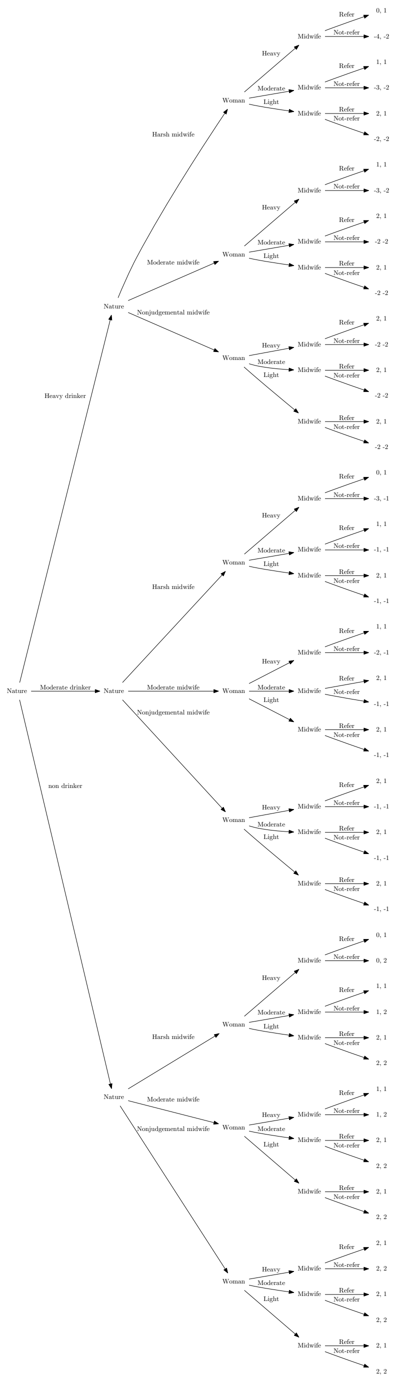


Figure 11: Complete game tree for the signalling game

B

CASELOADING REGRESSION MODELS

Predictor	Coefficient	p-value
Intercept	0.499	$< 2 \times 10^{-16}$
ln(Appointment)	-0.147	$< 2 \times 10^{-16}$
Caseloading	-0.020	$< 2 \times 10^{-16}$
Appointment	0.002	$< 2 \times 10^{-16}$
ln(Appointment):Caseloading	0.073	$< 2 \times 10^{-16}$
Caseloading:Appointment	-0.001	$< 2 \times 10^{-16}$

Table 9: Heavy drinkers signalling as light, $R^2 = 74\%$, $F_{5,197994} = 115,600$, $p < 2.2 \times 10^{-16}$

Predictor	Coefficient	p-value
Intercept	0.345	$< 2 \times 10^{-16}$
ln(Appointment)	0.078	$< 2 \times 10^{-16}$
Caseloading	0.008	$< 2.61 \times 10^{-13}$
Appointment	-0.001	$< 2 \times 10^{-16}$
ln(Appointment):Caseloading	-0.038	$< 2 \times 10^{-16}$
Caseloading:Appointment	0.001	$< 2 \times 10^{-16}$

Table 10: Heavy drinkers signalling as moderate, $R^2 = 67\%$, $F_{5,197994} = 79,750$, $p < 2.2 \times 10^{-16}$

Predictor	Coefficient	p-value
Intercept	0.156	$< 2 \times 10^{-16}$
ln(Appointment)	0.068	$< 2 \times 10^{-16}$
Caseloading	0.011	$< 2 \times 10^{-16}$
Appointment	-0.001	$< 2 \times 10^{-16}$
ln(Appointment):Caseloading	-0.034	$< 2 \times 10^{-16}$
Caseloading:Appointment	4.76×10^{-4}	$< 2 \times 10^{-16}$

Table 11: Heavy drinkers signalling as heavy, $R^2 = 66\%$, $F_{5,197994} = 77,650$, $p < 2.2 \times 10^{-16}$

Predictor	Coefficient	p-value
Intercept	0.555	$< 2 \times 10^{-16}$
ln(Appointment)	-0.161	$< 2 \times 10^{-16}$
Caseloading	-0.029	$< 2 \times 10^{-16}$
Appointment	0.002	$< 2 \times 10^{-16}$
ln(Appointment):Caseloading	0.078	$< 2 \times 10^{-16}$
Caseloading:Appointment	-0.001	$< 2 \times 10^{-16}$

Table 12: Moderate drinkers signalling as light, $R^2 = 74\%$, $F_{5,197994} = 113,900$, $p < 2.2 \times 10^{-16}$

Predictor	Coefficient	p-value
Intercept	0.336	$< 2 \times 10^{-16}$
ln(Appointment)	0.089	$< 2 \times 10^{-16}$
Caseloading	0.012	$< 2 \times 10^{-16}$
Appointment	-0.001	$< 2 \times 10^{-16}$
ln(Appointment):Caseloading	-0.042	$< 2 \times 10^{-16}$
Caseloading:Appointment	0.001	$< 2 \times 10^{-16}$

Table 13: Moderate drinkers signalling as moderate, $R^2 = 67\%$, $F_{5,197994} = 80,720$, $p < 2.2 \times 10^{-16}$

Predictor	Coefficient	p-value
Intercept	0.109	$< 2 \times 10^{-16}$
ln(Appointment)	0.072	$< 2 \times 10^{-16}$
Caseloading	0.002	$< 2 \times 10^{-16}$
Appointment	-0.001	$< 2 \times 10^{-16}$
ln(Appointment):Caseloading	-0.036	$< 2 \times 10^{-16}$
Caseloading:Appointment	4.67×10^{-4}	$< 2 \times 10^{-16}$

Table 14: Moderate drinkers signalling as heavy, $R^2 = 67\%$, $F_{5,197994} = 78,650$, $p < 2.2 \times 10^{-16}$

BIBLIOGRAPHY

[1] Ainslie, G., 1991. Derivation of "Rational" Economic Behavior from Hyperbolic Discount Curves. The American Economic Review 81 (2), 334–340. (Cited on pages 17 and 47.)

[2] 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. (Cited on page 14.)

[3] Alvik, A., Haldorsen, T., Groholt, B., Lindemann, R., Mar. 2006. Alcohol consumption before and during pregnancy comparing concurrent and retrospective reports. Alcoholism: Clinical & Experimental Research 30 (3), 510–5. (Cited on pages 3, 10, 29, and 48.)

[4] Alvik, A., Haldorsen, T., Lindemann, R., Aug. 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. (Cited on page 10.)

[5] Andersen, A.-M. N., Andersen, P. K., Olsen, J. r., Grø nbæk, M., Strandberg-Larsen, K., Apr. 2012. Moderate alcohol intake during pregnancy and risk of fetal death. International Journal of Epidemiology 41 (2), 405–13. (Cited on page 7.)

[6] Austen-Smith, D., Fryer Jr., R. G., 2005. An economic analysis of "acting white". Quarterly Journal of Economics (May), 551–583. (Cited on pages 13 and 47.)

[7] Barr, H. M., Bookstein, F. L., O'Malley, K. D., Connor, P. D., Huggins, J. E., Streissguth, A. P., Jun. 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. (Cited on page 9.)

[8] Bearer, C. F., Stoler, J. M., Cook, J. D., Carpenter, S. J., 2003. Biomarkers of alcohol use in pregnancy. Alcohol Research & Health 28 (1), 38–43. (Cited on page 45.)

[9] Bernoulli, D., 1954. Exposition of a New Theory on the Measurement of Risk. Econometrica 22 (1), 23–36. (Cited on page 14.)

[10] Biró, M. A., Waldenström, U., Brown, S., Pannifex, J. H., Mar. 2003. Satisfaction with team midwifery care for low- and high-risk women: a randomized controlled trial. Birth 30 (1), 1–10. (Cited on pages 11 and 12.)

[11] Blair, E., Sudman, S., Bradburn, N., 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. (Cited on page 9.)

[12] Booij, A. S., Praag, B. M. S., Kuilen, G., Apr. 2009. A parametric analysis of prospect theoryâs functionals for the general population. Theory and Decision 68 (1-2), 115–148. (Cited on page 32.)

[13] Bowers, J. S., Davis, C. J., May 2012. Bayesian just-so stories in psychology and neuroscience. Psychological Bulletin 138 (3), 389–414. (Cited on page 15.)

[14] Bradley, K. A., Boyd-Wickizer, J., Powell, S. H., Burman, M. L., Jul. 1998. Alcohol screening questionnaires in women: a critical review. JAMA: the Journal of the American Medical Association 280 (2), 166–71. (Cited on page 10.)

[15] Burke, M. S., Carter, J. R., Gominiak, R. D., Ohl, D. F., 1996. An Experimental Note on the Allais Paradox and Monetary Incentives. Empirical Economics 21 (4), 617–632. (Cited on page 14.)

[16] Byrnes, J. P., Miller, D. C., Schafer, W. D., 1999. Gender differences in risk taking: A meta-analysis. Psychological Bulletin 125 (3), 367–383. (Cited on page 32.)

[17] Camerer, C. F., 2004. Prospect theory in the wild: Evidence from the field. (Cited on page 16.)

[18] Chang, G., Wilkins-Haug, L., Berman, S., Goetz, M. A., Oct. 1999. Brief intervention for alcohol use in pregnancy: a randomized trial. Addiction 94 (10), 1499–508. (Cited on page 47.)

[19] Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., Schultz, W., Oct. 2009. Neural correlates of value, risk, and risk aversion contributing to decision making under risk. The Journal of Neuroscience 29 (40), 12574–83. (Cited on page 14.)

[20] Davies, M., Brannan, S., Chrispin, E., Mason, S., Mussell, R., Sheather, J., Sommerville, A., Sep. 2010. Ethics briefings. Journal of Medical Ethics 36 (9), 574–576. (Cited on page 45.)

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

[22] Del Boca, F. K., Darkes, J., Dec. 2003. The validity of self-reports of alcohol consumption: state of the science and challenges for research. Addiction 98 Suppl 2, 1–12. (Cited on page 9.)

[23] Del Boca, F. K., Noll, J. A., Nov. 2000. Truth or consequences: the validity of self-report data in health services research on addictions. Addiction 95 (11), 347–360. (Cited on page 9.)

[24] Dennis, C.-L., Chung-Lee, L., Dec. 2006. Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review. Birth 33 (4), 323–31. (Cited on page 11.)

[25] Department of Health, 2008. Pregnancy and alcohol. Department of Health, London. (Cited on page 4.)

[26] Di Paolo, E. A., Noble, J., Bullock, S., 2000. Simulation Models as Opaque Thought Experiments. (Cited on page 3.)

[27] Dorazio, R. M., Johnson, F. A., 2003. Bayesian inference and decision theory-a framework for decision making in natural resource management. Ecological Applications 13 (2), 556–563. (Cited on page 15.)

[28] Farquhar, M., Camilleri-Ferrante, C., Todd, C., Mar. 2000. Continuity of care in maternity services: women's views of one team midwifery scheme. Midwifery 16 (1), 35–47. (Cited on pages 11 and 29.)

[29] Farrell, J., Rabin, M., 1996. Cheap Talk. The Journal of Economic Perspectives 10 (3), 103–118. (Cited on page 47.)

[30] Fehr, E., Fischbacher, U., 2003. The nature of human altruism. Nature 425 (6960), 785–91. (Cited on page 15.)

[31] Feltovich, N., Harbaugh, R., To, T., 2002. Too cool for school? Signalling and countersignalling. RAND Journal of Economics 33 (4), 630–649. (Cited on pages 13 and 47.)

[32] Fryer Jr., R. G., Torelli, P., 2010. An empirical analysis of 'acting white'. Journal of Public Economics 94 (5), 380–396. (Cited on page 13.)

[33] Gaiman, N., 2001. American Gods. Headline Review, London. (Cited on page vii.)

[34] Gallistel, C. R., May 2012. Extinction from a rationalist perspective. Behavioural Processes 90 (1), 66–80. (Cited on page 15.)

[35] Glöckner, A., Pachur, T., Apr. 2012. Cognitive models of risky choice: parameter stability and predictive accuracy of prospect theory. Cognition 123 (1), 21–32. (Cited on page 46.)

[36] Godfray, H. C. J., 1991. Signalling of Need by Offspring to their Parents. Nature 352, 328–330. (Cited on page 13.)

[37] Goffman, E., 1990. Stigma: Notes on the Management of Spoiled Identity. Penguin Books, London. (Cited on page 9.)

[38] Gomberg, E. S., Jan. 1988. Alcoholic women in treatment: the question of stigma and age. Alcohol and Alcoholism 23 (6), 507–14. (Cited on page 9.)

[39] Grafen, A., 1990. Biological Signals as Handicaps. Journal of Theoretical Biology 144, 517–546. (Cited on page 12.)

[40] Gray, R., Henderson, J., 2006. National Perinatal Epidemiology Unit. (Cited on page 8.)

[41] Griffiths, T. L., Chater, N., Kemp, C., Tenenbaum, J. B., Perfors, A., 2010. Probabilistic Modelf of Cognition: Exploring Representations and Inductive Biases. Trends in Cognitive Sciences 14 (8), 357–364. (Cited on page 15.)

[42] Guerri, C., Riley, E., 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. (Cited on page 8.)

[43] Harsanyi, J. C., 1967. Games with Incomplete Information Played by "Bayesian" Players, I-III. Management Science 14 (3), 159–182. (Cited on page 12.)

[44] Harsanyi, J. C., 1978. Bayesian Decision Theory and Utilitarian Ethics. The American Economic Review 68 (2), 223–228. (Cited on page 15.)

[45] Hatem, M., Sandall, J., Devane, D., Soltani, H., Gates, S., 2009. Midwife-led versus other models of care for childbearing women (Review). Cochrane Database of Systematic Reviews (3). (Cited on page 12.)

[46] Health and Social Care Information Centre, IFF Research, 2012. Infant feeding survey 2010. (Cited on page 4.)

[47] Henderson, J., Kesmodel, U., Gray, R., Dec. 2007. Systematic review of the fetal effects of prenatal binge-drinking. Journal of Epidemiology and Community Health 61 (12), 1069–73. (Cited on pages 8 and 9.)

[48] Henrich, J., Jul. 2009. The evolution of costly displays, cooperation and religion. Evolution and Human Behavior 30 (4), 244–260. (Cited on page 13.)

[49] Huizink, A. C., Sep. 2009. Moderate use of alcohol, tobacco and cannabis during pregnancy: new approaches and update on research findings. Reproductive Toxicology 28 (2), 143–51. (Cited on page 8.)

[50] Huizink, A. C., Mulder, E. J. H., Jan. 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. (Cited on page 7.)

[51] Hunter, B., Dec. 2006. The importance of reciprocity in relationships between community-based midwives and mothers. Midwifery 22 (4), 308–22. (Cited on page 11.)

[52] Improving Access to Psychological Therapies, DrugScope, National Treatment Agency for Substance Misuse, 2012. IAPT positive practice guide for working with people who use drugs and alcohol. (Cited on page 9.)

[53] Kable, J. W., Glimcher, P. W., Sep. 2009. The neurobiology of decision: consensus and controversy. Neuron 63 (6), 733–45. (Cited on page 48.)

[54] Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision Under Risk. Econometrica 47 (March), 263–291. (Cited on page 16.)

[55] Kahneman, D., Tversky, A., 1984. Choices, values, and frames. American Psychologist 39 (4), 341–350. (Cited on page 16.)

[56] Kaskutas, L. A., Graves, K., Jan. 2000. An alternative to standard drinks as a measure of alcohol consumption. Journal of Substance Abuse 12 (1-2), 67–78. (Cited on page 10.)

[57] Kennedy, H. P., Shannon, M. T., Chuahorm, U., 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. (Cited on page 11.)

[58] Kennedy, M., 2005. CTCD: GEM Software. `http://ctcd.group.shef.ac.uk/gem.html`, accessed: 3/8/2013. (Cited on page 32.)

[59] Kesmodel, U., Wisborg, K., Olsen, S. F., Henriksen, T. B., Secher, N. J. r., Feb. 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. (Cited on page 7.)

[60] Kreps, D. M., Cho, I., 1987. Signalling Games and Stable Equilibria. The Quarterly Journal of Economics 102 (2), 179–221. (Cited on pages 3 and 12.)

[61] Kristensen, S., Mar. 1997. Sensor planning with Bayesian decision theory. Robotics and Autonomous Systems 19 (3-4), 273–286. (Cited on page 15.)

[62] Langenbucher, J., Merrill, J., Apr. 2001. The Validity of Self-Reported Cost Events by Substance Abusers: Limits, Liabilities, and Future Directions. Evaluation Review 25 (2), 184–210. (Cited on page 9.)

[63] Latino-Martel, P., Chan, D. S. M., Druesne-Pecollo, N., Barrandon, E., Hercberg, S., Norat, T., May 2010. Maternal alcohol consumption during pregnancy and risk of childhood leukemia: systematic review and meta-analysis. Cancer Epidemiology, Biomarkers & Prevention 19 (5), 1238–60. (Cited on page 7.)

[64] Litten, R. Z., Bradley, A. M., Moss, H. B., Jun. 2010. Alcohol biomarkers in applied settings: recent advances and future research opportunities. Alcoholism, Clinical and Experimental Research 34 (6), 955–67. (Cited on page 45.)

[65] Loewenstein, G., Prelec, D., 1992. Anomalies in intertemporal choice: Evidence and an interpretation. The Quarterly Journal of Economics 107 (2), 573–597. (Cited on pages 17 and 47.)

[66] Maier, S. E., West, J. R., 2001. Drinking Patterns and Alcohol-Related Birth Defects. Alcohol Research & Health 25 (3), 168–174. (Cited on page 8.)

[67] Malvin, J. H., Moskowitz, J. M., 1983. Anonymous Versus Identifiable Self-Reports of Adolescent Drug Attitudes, Intentions, and Use. Public Opinion Quarterly 47 (4), 557. (Cited on page 10.)

[68] Marks, M. N., Siddle, K., Warwick, C., Feb. 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. (Cited on page 11.)

[69] McNamara, J., Houston, A., Aug. 1980. The application of statistical decision theory to animal behaviour. Journal of Theoretical Biology 85 (4), 673–90. (Cited on page 15.)

[70] Meyer-Leu, Y., Lemola, S., Daeppen, J.-B., Deriaz, O., Gerber, S., Sep. 2011. Association of moderate alcohol use and binge drinking during pregnancy with neonatal health. Alcoholism: Clinical & Experimental Research 35 (9), 1669–77. (Cited on page 8.)

[71] Miller, R. R., May 2012. An assessment of Gallistel's (2012) rationalistic account of extinction phenomena. Behavioural Processes 90 (1), 81–3; discussion 87–8. (Cited on page 15.)

[72] National Institute for Health and Care Excellence, 2007. Antenatal and Postnatal Mental Health. CG45. NICE, London. (Cited on page 16.)

[73] National Institute for Health and Care Excellence, 2010. Alcohol-use disorders: preventing harmful drinking. PH24. NICE, London. (Cited on page 9.)

[74] National Institute for Health and Care Excellence, 2010. Antenatal care. CG62. NICE, London. (Cited on pages 4, 8, 9, and 21.)

[75] National Institute for Health and Clinical Excellence, 2010. Pregnancy and complex social factors. CG110. NICE, Manchester. (Cited on pages 9 and 48.)

[76] Neilson, W., Stowe, J., 2002. A further examination of cumulative prospect theory parameterizations. Journal of Risk and Uncertainty 24 (1), 31–47. (Cited on page 32.)

[77] NHS Litigation Authority, 2012. Ten Years of Maternity Claims. (Cited on page 4.)

[78] Nilsson, H. k., Rieskamp, J., Wagenmakers, E.-J., Feb. 2011. Hierarchical Bayesian parameter estimation for cumulative prospect theory. Journal of Mathematical Psychology 55 (1), 84–93. (Cited on page 32.)

[79] O'Connor, M. J., Whaley, S. E., 2007. Brief Intervention for Alcohol Use by Pregnant Women. American Journal of Public Health 92 (2), 252–258. (Cited on page 47.)

[80] Oliver, A., 2003. A quantitative and qualitative test of the Allais paradox using health outcomes. Journal of Economic Psychology 24 (1), 35–48. (Cited on page 14.)

[81] Padoa-Schioppa, C., Assad, J. A., May 2006. Neurons in the orbitofrontal cortex encode economic value. Nature 441 (7090), 223–6. (Cited on page 14.)

[82] Padoa-Schioppa, C., Assad, J. A., Jan. 2008. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nature neuroscience 11 (1), 95–102. (Cited on page 14.)

[83] Peterson, M., 2009. An introduction to decision theory. Cambridge University Press, Cambridge, UK. (Cited on page 13.)

[84] Phillips, D., Thomas, K., Cox, H., Ricciardelli, L. A., Ogle, J., Love, V., 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. (Cited on pages 3, 10, 29, and 35.)

[85] Piccinelli, M., 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 (February), 420–424. (Cited on page 10.)

[86] Platt, M. L., Glimcher, P. W., Jul. 1999. Neural correlates of decision variables in parietal cortex. Nature 400 (6741), 233–8. (Cited on page 14.)

[87] Prior, L., Wood, F., Lewis, G., Pill, R., May 2003. Stigma revisited, disclosure of emotional problems in primary care consultations in Wales. Social Science & Medicine 56 (10), 2191–200. (Cited on page 10.)

[88] Radcliffe, P., 2011. Substance-misusing women: Stigma in the maternity setting. British Journal of Midwifery 19 (8), 497–506. (Cited on page 9.)

[89] Robbins, H., Mar. 1964. The Empirical Bayes Approach to Statistical Decision Problems. The Annals of Mathematical Statistics 35 (1), 1–20. (Cited on page 14.)

[90] Royal College of Obstetricians and Gynaecologists, 2006. Alcohol Consumption and the Outcomes of Pregnancy. RCOG Statement No. 5. RCOG, London. (Cited on page 8.)

[91] Royal College of Obstetricians and Gynecologists, 1996. Alcohol Consumption in Pregnancy. RCOG Guideline No. 9. RCOG, London. (Cited on page 8.)

[92] Russell, M., 1994. New assessment tools for risk drinking during pregnancy. Alcohol Health and Research World 18 (1), 55–67. (Cited on page 10.)

[93] Russell, M., Martier, S. S., 1996. Detecting risk drinking during pregnancy: a comparison of four screening questionnaires. American Journal of Public Health 86 (10), 1435–1439. (Cited on page 10.)

[94] Rustichini, A., Dec. 2009. Neuroeconomics: what have we found, and what should we search for. Current Opinion in Neurobiology 19 (6), 672–7. (Cited on page 5.)

[95] Samuelson, P. A., 1937. Note on Measurement of Utility. The Review of Economic Studies 4 (2), 155–161. (Cited on page 17.)

[96] Sandall, J., 1997. Midwives' burnout and continuity of care. British Journal of Midwifery 5 (2), 106–111. (Cited on pages 12 and 45.)

[97] Savitz, D. A., Dec. 2012. Re: Moderate alcohol intake during pregnancy and risk of fetal death. International Journal of Epidemiology 41 (6), 1847; author reply 1848. (Cited on page 7.)

[98] Simon, H. A., 1956. Rational choice and the structure of the environment. Psychological Review 63 (2), 129–138. (Cited on pages 15 and 16.)

[99] Simon, H. A., 2000. Bounded Rationality in Social Science: Today and Tomorrow. Mind & Society 1 (1), 25–39. (Cited on page 16.)

[100] Sosis, R., Jun. 2003. Why aren't we all hutterites? Human Nature 14 (2), 91–127. (Cited on page 13.)

[101] Spence, M., 1973. Job Market Signaling. The Quarterly Journal of Economics 87 (3), 355–374. (Cited on pages 12 and 13.)

[102] Stevens, T., McCourt, C., 2002. One-to-one Midwifery Practice Part 2: the Transition Period. British Journal of Midwifery 10 (1), 45–50. (Cited on page 11.)

[103] Stockwell, T., Donath, S., Cooper-Stanbury, M., Chikritzhs, T., Catalano, P., Mateo, C., Aug. 2004. Under-reporting of alcohol consumption in household surveys: a comparison of quantity-frequency, graduated-frequency and recent recall. Addiction 99 (8), 1024–33. (Cited on page 10.)

[104] Strandberg-larsen, K., Nielsen, N. R., Andersen, P. K., Andersen, A.-m. N., 2008. Binge Drinking in Pregnancy and Risk of Fetal Death. Obstetrics & Gynecology 111 (3), 602–609. (Cited on page 8.)

[105] Streissguth, A. P., Barr, H. M., Olson, H. C., Sampson, P. D., Bookstein, F. L., Burgess, D. M., Apr. 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. (Cited on page 8.)

[106] Sun, Y., Strandberg-Larsen, K., Vestergaard, M., Christensen, J., Nybo Andersen, A.-M., Grø nbaek, M., Olsen, J. r., Feb. 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. (Cited on page 8.)

[107] Sutton, R. S., 1988. Learning to predict by the methods of temporal differences. Machine learning 3 (1), 9–44. (Cited on page 48.)

[108] Sutton, R. S., Barto, A. G., 1987. A temporal-difference model of classical conditioning. In: Proceedings of the Ninth Annual Conference of the Cognitive Science Society. pp. 355–378. (Cited on page 48.)

[109] Sutton, R. S., Barto, A. G., 1990. Time-Derivative Models of Pavlovian Reinforcement. In: Gabriel, M., Moore, J. (Eds.), Learning and computational neuroscience: Foundations of Adaptive Networks. MIT Press, Cambridge, MA, Ch. 12, pp. 497–537. (Cited on page 48.)

[110] Tenenbaum, J. B., Griffiths, T. L., Kemp, C., Jul. 2006. Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences 10 (7), 309–18. (Cited on page 15.)

[111] Thaler, R., 1981. Some Empirical Evidence on Dynamic Inconsistency. Economics Letters 8, 201–207. (Cited on page 17.)

[112] Thaler, R., 2000. From homo economicus to homo sapiens. The Journal of Economic Perspectives 14 (1), 133–141. (Cited on page 16.)

[113] Todd, C. J., Farquhar, M. C., Camilleri-Ferrante, C., Dec. 1998. Team midwifery: the views and job satisfaction of midwives. Midwifery 14 (4), 214–24. (Cited on pages 11 and 12.)

[114] Toll, B. A., O'Malley, S. S., 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. (Cited on page 16.)

[115] Turner, C., Sep. 1990. How much alcohol is in a 'standard drink'? An analysis of 125 studies. British Journal of Addiction 85 (9), 1171–5. (Cited on page 10.)

[116] Tversky, A., Kahneman, D., Sep. 1974. Judgment under Uncertainty: Heuristics and Biases. Science 185 (4157), 1124–31. (Cited on pages 16 and 46.)

[117] Tversky, A., 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. (Cited on page 16.)

[118] Tversky, A., Kahneman, D., Oct. 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5 (4), 297–323. (Cited on pages 16 and 27.)

[119] Van Walraven, C., Oake, N., Jennings, A., Forster, A. J., Oct. 2010. The association between continuity of care and outcomes: a systematic and critical review. Journal of Evaluation in Clinical Practice 16 (5), 947–56. (Cited on page 12.)

[120] Von Neumann, J., Morgenstern, O., 1953. The theory of games and economic behavior, 3rd Edition. Princeton University Press, Princeton. (Cited on page 14.)

[121] Watson, A. C., Corrigan, P., Larson, J. E., Sells, M., Nov. 2007. Self-stigma in people with mental illness. Schizophrenia Bulletin 33 (6), 1312–1318. (Cited on page 10.)

[122] Watson, K. K., Platt, M. L., Dec. 2012. Social signals in primate orbitofrontal cortex. Current biology : CB 22 (23), 2268–73. (Cited on page 14.)

[123] Wildman, W. J., Sosis, R., 2011. Stability of Groups with Costly Beliefs and Practices. Journal of Artificial Societies and . . . 14 (3). (Cited on page 13.)

[124] Willis, M. L., Palermo, R., Burke, D., McGrillen, K., Miller, L., Jun. 2010. Orbitofrontal cortex lesions result in abnormal social judgements to emotional faces. Neuropsychologia 48 (7), 2182–7. (Cited on page 14.)

[125] Zahavi, A., Sep. 1975. Mate selection-a selection for a handicap. Journal of Theoretical Biology 53 (1), 205–14. (Cited on page 12.)