Deciding to Disclose

A Decision Theoretic Agent Model of Pregnancy and Alcohol Misuse

Jonathan Gray, Jakub Bijak, and Seth Bullock

Abstract

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

We employ a game theoretic framework to define a signalling game. The game represents a scenario where pregnant women decide the extent to which they disclose their drinking behaviours to their midwives, and midwives employ the information provided to decide whether to refer their patients for costly specialist treatment. This game is then recast as two games played against "nature", to permit the use of a decision theoretic approach where both classes of agent use simple rules to decide their moves. Four decision rules are explored - a lexicographic heuristic which considers only the link between moves and payoffs, a Bayesian risk minimisation agent that uses the same information, a more complex Bayesian risk minimiser with full access to the structure of the decision problem, and a Cumulative Prospect Theory (CPT) rule.

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

―――――――――――――――

Jonathan Gray
University of Southampton, Southampton, United Kingdom e-mail: `j.gray@soton.ac.uk`

Jakub Bijak
University of Southampton, Southampton, United Kingdom e-mail: `j.bijak@soton.ac.uk`

Seth Bullock
University of Southampton, Southampton, United Kingdom e-mail: `sgb@ecs.soton.ac.uk`

1 Introduction

The case in favour of Agent Based Modelling (ABM) as a general analytical approach has been made numerously, and elegantly (e.g. [11; 38; 3; 15; 24; 41; 42; 10], amongst others). As such we will not belabour the point, and instead turn to addressing some of the concerns expressed about the approach. In this instance we focus on the perception of ABM as ad hoc in nature, reflecting the assumptions of the modeller rather than empirically or theoretically grounded [47]. To ameliorate this concern, we draw on decision theory to produce simple rule based, learning, decision making agents and show that they are able to play a form of signalling game[1] [22] with a basic form of intragroup social learning. Four decision models of varying complexity, and behavioural plausibility are contrasted, by way of demonstrating the significance of the operationalisation of decision making in ABM.

This exercise is framed in the context of disclosure decisions, taking drinking patterns in pregnant women as a motivating example. Alcohol consumption in the antenatal period is a significant issue in itself, although there is not a clear consensus on associated risk. In terms of official guidance in the UK, National Institute for Health and Care Excellence (NICE) acknowledge that evidence of harm to the fetus is less than conclusive, but advise not drinking at all, or significant moderation [25], with similar advice from the UK [8].

Turning more specifically to disclosure of alcohol use by women to healthcare professionals during their pregnancy, research is relatively sparse, although qualitative trends are reported by Phillips et al [34] and Alvik et al [2]. The former explored factors impacting disclosure through a small case study, highlighting the need to build up rapport between woman and midwife over several appointments; the latter compared post partum reports of consumption with contemporaneous accounts, finding apparent underreporting during pregnancy which was amplified by increased drinking. The simulation model described in this paper is able to replicate both qualitative trends, i.e. an increase in disclosure over appointments, and more honest behaviour by moderate as compared to heavier drinkers.

The resulting scenario is of substantial independent interest, and shows the potential utility of a simulation approach in domains where process is opaque, here both because of the interest in concealment, and obvious ethical concerns. With this said, the lack of a strong quantitative evidence base against which to validate the behaviour of the model augers for caution in interpreting the results, and a necessary reminder that in this instance the model is primarily a tool for formalisation of the thought process [9], rather than a machine for predicting.

A game theoretic approach to generating an abstract form of the problem gives a convenient, and well known framework to reason about the processes involved in the scenario. While scenarios may map to a number of games, exploring one candidate game still allows for a principled comparison between interpretations, and enforces

---

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

explicit assumptions. Relating this to decision theory shifts the emphasis away from analytical equilibrium-seeking, and heightens the importance of behaviour change. This is clearly desirable where the focus is on the behavioural processes driving a system in motion, and how they change in response to that movement. Fundamentally, the shift of emphasis is from the process of acquiring and inferring the information needed to make choices, to the process of decision. This does not preclude a strategic dimension, since decision rules are to a great extent modular, and as demonstrated in this paper can be exchanged without altering the decision problem. In addition, rules are agnostic as to where the source of information, suggesting room for multi-stage processes. As a corollary, the decision problem agents attempt to answer can change, allowing behaviour in novel problems to be informed by beliefs derived under other conditions. This is also indicative of the broader benefits to ABM as an approach. Embedding these abstract rules in a simulated environment allows for mechanics which cannot be readily explored using purely analytic, or predictive approaches, for example, the social learning dynamic of the disclosure game model.

While there is no universal theory of human behaviour to sit at the centre of ABM as a method, a key motivation for decision rules is their claim to provide an account of decision making that is behaviourally and cognitively plausible. Their mooted capability in this regard is to some extent supported by work from neuroeconomics, which aims to empirically test theories of decision making [39]. Many key aspects common to decision rules, for example the idea that a common currency is used by the brain to compare outcomes [32; 33], are supported by neurological findings. In addition, a single decision rule represents a parsimonious alternative to numerous case specific production rules.

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

The remainder of this paper proceeds to outline the proposed approach to model development (section 2), and experiments (section 3), with selected results (section 4), followed by a discussion contrasting the decision models (section 5), and conclusions (section 6).

2 Disclosure Game Model

In this section we sketch[2] the process of moving from a real world scenario to a minimal game which sufficiently captures reality, expressing the result as a decision

---

[2] A complete example of this for the alcohol misuse in pregnancy model is given in appendix 8, with a schedule of simulation provided in appendix 9

problem representation, and translating this to a simulation model. We then outline four possible decision rules, and as an example of additional flexibility of process models and simulation in contrast to purely predictive or analytical approaches, extend the model to allow a simple form of social learning.

2.1 Modelling Approach

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

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

Transitioning from the resulting game, to a set of decision problems is a relatively simple task. We treat the n player game as n one-player games [20], where the moves of other players are drawn from a probability distribution - nature, in game theoretic parlance. As with the game, the decision problem representation admits a degree of variation, and may need to be adjusted to reflect the decision rules that will be used.

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

2.2 Scenario

Typically in the UK, a women will have 12 appointments with a midwife during the antenatal period, and in the majority of cases will encounter several different midwives [37] in the course of their care. In the UK, and unlike most healthcare scenarios, maternity notes are held by the patient, so midwives do not have extensive information prior to an appointment unless they have encountered the woman previously. Maternity notes are not generally linked to extra-departmental records, meaning that a history of alcohol related admissions to another service may remain unknown unless revealed by the woman.

According to NICE guidance [25; 26] the issue of substance misuse should be raised at the initial booking appointment, followed by subsequent action if a concern is raised is at the discretion of the midwife. This may take the form of specific

guidance to reduce intake, or if deemed necessary a referral to a specialist midwife and relevant interdisciplinary team. On alcohol consumption, policy regarding how to determine the level of consumption is at the time of writing generally at the level of local health authority, hospital trust, or according to the best judgement of the individual midwife, with no guidance provided by NICE. This commonly takes the form of average units per week, but may include Tolerance, Annoyance, Cut down, Eye-opener (T-ACE)[3] [43] and similar measures.

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

In an ideal world, all interactions with healthcare providers would be immediately and fully disclosive, with no repercussions for the patient. However, alcohol misuse by women is known to attract stigma [17], and is a recognised barrier to appropriate treatment in the maternity context [26; 36].

2.3 Disclosure Game

In order to translate the scenario sketched above into a more abstract, tractable form, we cast it as a signalling game, and assume that women's disclosures (or not), are signals. We also impose a discretisation on the continuum of alcohol use, and use three types of behaviour - light[4], moderate, or heavy. Correspondingly, they are limited in what signals they may send when claiming to be one of these three types.

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

At the end of a game, each player receives a payoff dependent on the actions and types of both players. Because both women and midwives have an interest in the outcome of the pregnancy, and would prefer a healthy baby, the payoff has a common interest component. Hence, both players receive a payoff based on the outcome of pregnancy, but women bear a social cost dependent on the signal they sent and the midwife's reaction to it. Similarly, midwives pay a cost if they refer to a specialist, mirroring the organisational cost of non-routine care. Table 1 shows the three payoff matrices which together describe the game.

As an example, consider the challenge faced by an agent of the heavy drinking type. In order to get the best health outcome, they must be referred and would ideally achieve this without paying any social cost at all. The best move depends on the type,

---

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

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

Table 1 Payoff matrices

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

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

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

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

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

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

and beliefs of the midwife. For example, a particularly unlucky scenario might be for the midwife to not only be of a harshly judgemental disposition, but to believe that no women really need to be referred. Even a relatively weak belief in this possibility can make the honest signal look like an unwarranted risk.

To formally define the game, let $N = \{m, w\}$ be the set of players each with a private type $\theta_i \in \Theta$, and a set of types $\Theta = \{l, m, h\}$, with pure strategies $A_m = \{r, n\}, A_w = \{l, m, h\}$. Here, $\{l, m, h\}$ correspond to light, moderate, and heavy alcohol consumption for women, and non-judgemental, moderately judgemental, and harshly judgemental for midwives. Midwives' pure strategies $\{r, n\}$ are to refer, or do nothing, and those for women are to signal that they have one of the possible drinking patterns. Additionally define two utility functions -

$$u_w(s_w, s_m, \theta_w, \theta_m) = X_{s, s_w, \theta_m} + X_{h, \theta_w, s_m} \tag{1}$$

$$u_m(s_w, s_m, \theta_w) = X_{h, \theta_w, s_m} + X_{c, \theta_w, s_m}, \tag{2}$$

with X_c, X_h, and X_s being the payoff matrices as in table 1, s_w and s_m denoting a specific signal by a woman, and referral response by a midwife. Lastly let $p_w(l, m, h)$, $p_m(l, m, h)$ be distributions over types of women, and midwives respectively.

As noted, rather than solve the game, we allow populations of agents to play it, and hence stipulate further that women are drawn in order from a queue of n_w women (where $n_w = 1000$ in all simulations), and play against a midwife chosen at

random from a population of n_m ($n_m = 100$). They play for a maximum of r_w rounds ($r_w = 12$ following the routine number of ante-natal appointments in the UK [25]) or until they are referred, and a new player is drawn from the same distribution that produced the original players to replace them. If they are not referred, they rejoin the back of the queue after their appointment. In either case, they are informed of their payoff after each round and update their beliefs accordingly using one of the rules described in section 2.5.

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

2.4 Social Learning

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

In the disclosure game model, this takes the form of having each agent recount their play history to their colleagues with some probability q. Individuals then incorporate shared information into their beliefs using weighted updates, such that a shared observation of a low type signal contributes to their beliefs by w, and $0 \leq w \leq 1$ (i.e. $n_j = n_j + w$). Women share only when they have finished play, and provide their complete history of games, because they have accurate information about the outcomes. By the same rationale, midwives share only their history with the most recent woman they referred. Sharing occurs simultaneously for all players at the end of each round, and all memories are either shared immediately or discarded.[5]

Because of their differing problem representations, the simple payoff reasoners and their more complex counterparts incorporate this exogenous information differently. The simple payoff based rule relies on a belief structure relating actions directly to rewards which is essentially model free. Because payoffs differ by the agent's private type, the information shared may not correspond to the experience of the listening agent in the same scenario. As a result, payoff reasoners have a belief bias towards the most common player type, and can believe in outcomes that are, for them, impossible.

By contrast, representing the problem in terms of the probabilities of the individual lotteries imposes a model that abstracts the new information from payoffs, and allows the agent to discard implausible outcomes. This stronger assumption as to

---

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

the static and known qualities of payoffs does however reduce the flexibility of the decision rule.

2.5 Agent Models

While in principle a wide variety of agent models are possible, given that decision rules operate on essentially the same information, and produce the same outputs, we limit ourselves here to four. The simplest is a lexicographic rule (1), in the spirit of a Fast and Frugal Heuristic [14] which uses only information about payoffs given actions; this is followed by a Bayesian risk minimisation rule (2) using the same information; a second Bayesian risk rule (3) which uses information about the underlying lottery; and a two-stage Cumulative Prospect Theory (CPT) [18] agent (4) which is identical to 3, but uses the CPT decision rule from [46]. Hence, each successive decision model adds a layer of sophistication to the problem representation while retaining the same input-output characteristics.

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

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

2.5.1 Lexicographic Heuristic

The lexicographic heuristic (algorithm 1) follows the form of that used in [18], and assumes a simplified problem representation, where an action is a choice between combined lotteries. Functionally, the heuristic maintains a count of the number of times that each action was followed by a payoff, and chooses the action which most commonly has the best payoff, i.e. one reason decision making. Where there is no clear best action, but one or more is evidently worse, a choice must be made as to whether to discard the poorer action; in this case we have elected to retain it. This approach requires minimal computation, and does not assume that u_i is static, or known.

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

Midwives solve a slightly different problem with more information, where s_w is known, and s_m is the lottery choice - $f(s_w, s_m, \sigma)$. This is resolved by maintain-

ing a separate count for each signal (i.e. n_{s_w,s_m}), and otherwise following the same algorithm (1).

Algorithm 1 Lexicographic heuristic

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

2.5.2 Bayesian Payoff

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

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

$$p(x = j|D, \alpha) = \frac{\alpha_j + n_j}{\sum_j(\alpha_j + n_j)} \tag{3}$$

Where n_j is simply the count of occurrences of signal-payoff pair j, and α_i is the pseudo-count of prior observations[6] for a pair i. Hence, the belief that a signal will lead to a payoff is the number of times that pairing has been observed (including the pseudo-count), over the total number of observations thus far. This makes computation of beliefs fast and simple, since all that must be maintained is a count of observations. As before, midwives follow a similar pattern but maintain n_{s_w} independent counts of pairings between referral choice and payoff, updating their beliefs about the relationship between the choice to refer and payoff given the signal they have received.

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

$$R_w(s_w) = \sum_{x \in X} -xp(x|s_w) \tag{4}$$

---

[6] Psuedo-counts are related to, but distinct from prior beliefs. Here, the psuedo-count is a parameter to the prior belief distribution and is nothing more than a hypothetical count of prior observations.

$$R_m(s_w, s_m) = \sum_{x \in X} -xp(x|s_w \wedge s_m),\qquad(5)$$

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

2.5.3 Bayesian Risk Minimisation

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

$$R_w(s_w, \theta_w) = \sum_{i \in A_m} \sum_{j \in \Theta} -u_w(s_w, i, \theta_w, j) p(j) p(i|s_w),\qquad(6)$$

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

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

$$R_m(s_w, s_m) = \sum_{i \in \Theta} -u_m(s_w, s_m, i) p(i|s_w)\qquad(7)$$

2.5.4 Descriptive Decision Theory

The most complex decision rule used is CPT, which attempts to reproduce a number of systematic deviations from rationality observed in humans. Rather than risk, 'prospects', the sequence of payoff-probability pairings in ascending order of payoff, associated with an action are used as decision criteria. While CPT has primarily been applied in the context of decisions from description, it has been modified to deal with decisions from experience by incorporating a first stage where probabilities are estimates from observations [13]. In this instance the Bayesian inference process fills the first stage role.

CPT uses transformed probabilities, underweighting small probabilities, and overweighting large ones. This is intended to reflect the observed behaviour of humans, where sufficiently high likelihoods are treated as certain, and contrastingly low probabilities as impossible. The correct weighting function is subject to some debate, but here we have used that of Tversky and Kahneman [46], which treats probabilities differently for gains (eqn 8) and losses (eqn 9) -

$$w^+(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}} \tag{8}$$

$$w^-(p) = \frac{p^\delta}{(p^\delta + (1-p)^\delta)^{\frac{1}{\delta}}}, \tag{9}$$

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

$$v(u_i) = \begin{cases} f(u_i), & \text{if } u_i > 0 \\ 0, & \text{if } u_i = 0 \\ \lambda g(u_i), & \text{if } x < 0 \end{cases} \tag{10}$$

where,

$$f(u_i) = \begin{cases} u_i^\alpha, & \text{if } \alpha > 0 \\ \ln(u_i), & \text{if } \alpha = 0 \\ 1 - (1 + u_i)^\alpha, & \text{if } \alpha < 0 \end{cases} \tag{11}$$

$$g(u_i) = \begin{cases} -(-u_i)^\beta, & \text{if } \beta > 0 \\ -\ln(-u_i), & \text{if } \beta = 0 \\ (1 - u_i)^\beta - 1, & \text{if } \beta < 0 \end{cases} \tag{12}$$

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

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

$$\pi_n^+ = w^+(p_n) \tag{13}$$

$$\pi_{-m}^- = w^-(p_{-m}) \tag{14}$$

$$\pi_i^+ = w^+(p_i + \ldots + p_n) - w^+(p_{i+1} + \ldots + p_n), 0 \le i \le n-1 \tag{15}$$

$$\pi_i^- = w^-(p_{-m} + \ldots + p_i) - w^-(p_{-m} + \ldots + p_{i-1}), 1 - m \le i \le 0 \tag{16}$$

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

3 Method

This section provides details of experiments conducted to examine the ability of the model to reproduce qualitative trends reported in the midwifery literature by Alvik et al [2], and Phillips et al [34]; as well as a global sensitivity analysis and construction of statistical emulators to explore, and contrast the response surfaces of the four decision rules.

3.1 Qualitative Trends

Throughout this paper, parameters for the CPT model were as used in [46] (table 2). While there has been significant work on determining appropriate parametrisation for the model (e.g. [27; 28; 16], and particularly Byrnes et al [6] and Booij et al [5] addressing risk aversion and gender), a full exploration of the impact of these parameters, or heterogeneous values within populations is beyond the scope of this work. For simplicity, it was assumed that all three drinking types are equally prevalent within the population, although results derived from the Avon Longitudinal Study of Parents and Children suggest that the reality is far more positive[7] [19]. The scenario was biased towards disclosure as the better option by presuming a distribution of midwives strongly skewed towards non-judgemental types, with beliefs initially favouring honesty. Payoffs were as in table 1, which ensure that it is always strictly preferable to refer drinkers, and together with the initial belief that signals will be honest, not refer those claiming otherwise.

Two key measures were used: the fraction of the subpopulation who had ever signalled honestly, and the proportion referred. Both measures were taken after every round of play, and were taken relative to the agent's position in their sequence of appointments giving the probability of signalling honestly, or being referred having had a given number of appointments.

In addition to assessing the adequacy of the rules in capturing qualitative trends, we also examined the impact of simple social learning within the population of women (section 2.4) on the robustness of these trends. The original experiment was repeated at $q_w, w_w \in \{0.25, 0.5, 0.75, 1 | q_w = w_w\}$, with 100 runs under each condition.

3.2 Global Sensitivity Analysis

In general, we have followed the example of [4] for global sensitivity analysis of stochastic agent based models, although see Thiele et al [45] for a review of alter-

---

[7] 95.5% of women in the sample reported consumption at, or below, NICE recommended safe levels.

Table 2 Model parameters.

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

native techniques. For this purpose the Gaussian Emulation Machine for Sensitivity Analysis (GEM-SA) software [21] was used, which implements the Bayesian Analysis of Computer Code Outputs (BACCO) method developed by Oakley and O'Hagan [30; 31; 29]. This is a form of variance-based sensitivity analysis, which assumes that the model output is an unknown, smooth function of the inputs. The unknown function can then be approximated as a Gaussian Process, which is fitted to the training data using Bayes' Theorem and then serves as an emulator for the simulator. The emulator is then able to provide an indication about the extent to which uncertainty in a parameter propagates to uncertainty in the output, and how sharply the output responds to change in each parameter.

Parameters for training were generated in R [35] using and appropriately transformed Latin Hypercube Sample [7] over the space of parameters given in table 3, giving eleven free parameters which were treated as uniformly distributed in the range given. Given the limitation of 400 design points for the GEM-SA software, we produced exactly that many parameter combinations and collected results for 100 runs of each, with emulator quality assessed by leave-one-out cross validation. A fixed set of 100 random seeds was used, such that each parameter set was run once with each seed, for every decision rule.

To capture the response characteristics for the model, we measured four outcome variables - (1) the interquartile range of the average signal sent by each type of agent in a run, (2) the average signal of moderate drinking agents in a run, and (3, 4) the interquartile range of the average signal, and interquartile range (IQR) between

simulation runs. Together these four metrics give an indication of how far women are separable by their signalling behaviour (1), the behaviour of the at risk drinking groups[8] (2), and finally the variability in the system in response to changes to the parameters (3 & 4).

Measurements were taken at the end of 1000 rounds of play, and emulators were built against the median, and IQR of each measure to assess both the overall trend, and the extent to which the parameters contribute to variance between runs.

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

In addition to the sensitivity analysis, we also employed the resulting emulators to rapidly[9] explore the parameter space. While emulated results are subject to inaccuracy, they do provide an indication of the interest, and plausibility, of regions of the parameter space. Results for the outcomes of the interactions of $s_i[a_i] : s_i[a_{\neg i}]$ with x_h, and q_w with w_w are given in section 4.3.

Table 3 Parameter ranges.

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

4 Results

4.1 Qualitative Trends

As shown in figure 1, all four decision rules were able to reproduce both qualitative trends towards more disclosure as women experience more appointments [34], and

---

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

[9] Once constructed, the emulator has an analytical solution conditional on the roughness parameters, which obviates the need to use MCMC.

a greater tendency towards underreporting of consumption by heavier drinkers [2]. Trends for all four rules are broadly similar, exhibiting a gradual increase across appointments which subsequently levels off. This levelling can in part be explained by the referral results, which show that the majority of drinkers are referred, even with substantial concealment. Referrals continue to occur, in the absence of honest signals, because drinkers are able to achieve a referral by masquerading as higher or lower types, dependent on how their initial beliefs are biased. Despite this the results suggest that a minority of risky drinkers will evade detection altogether, with no notable distinction between heavy and moderate types. Under these parameters, light drinkers always signal honestly and are never referred since there is no perceived advantage in doing so, and the evidence of deceptive signalling is insufficient to outweigh the biased priors of the midwives.
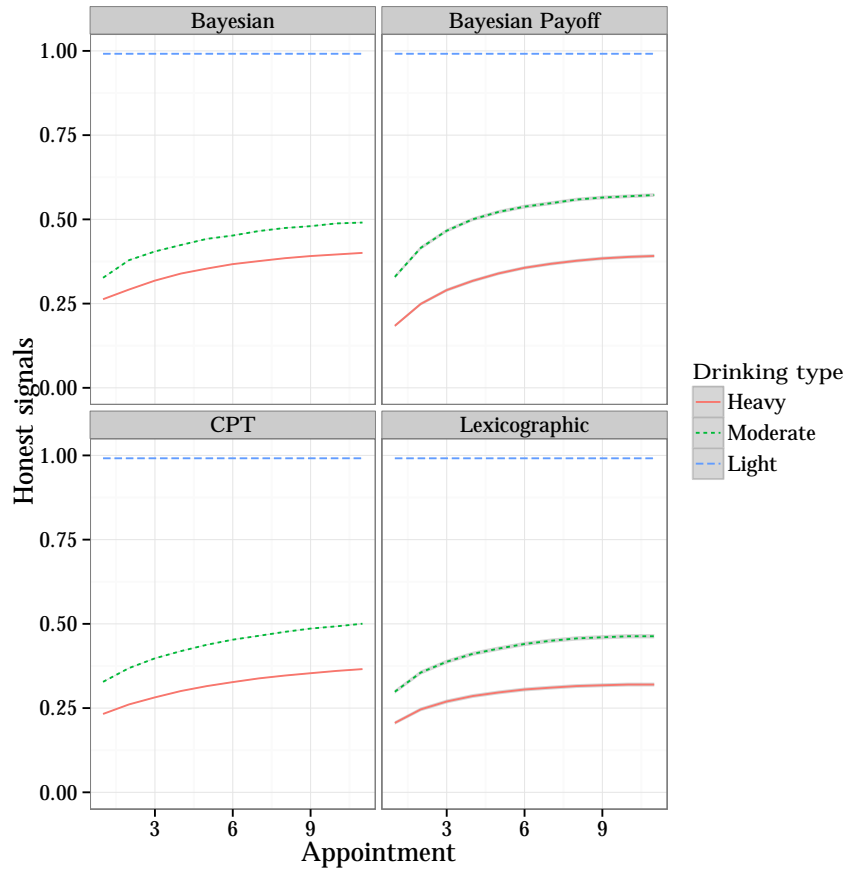

Fig. 1 Average fraction of population ever signalled honestly after 1000 rounds, mean with 95% confidence limit over 1000 runs

4.2 Social Learning

Introducing social learning amongst women leads the behaviour of the decision rules to diverge markedly. Figure 2 shows the proportion of women who have signalled honestly at least once by their final appointment, under four sharing conditions.
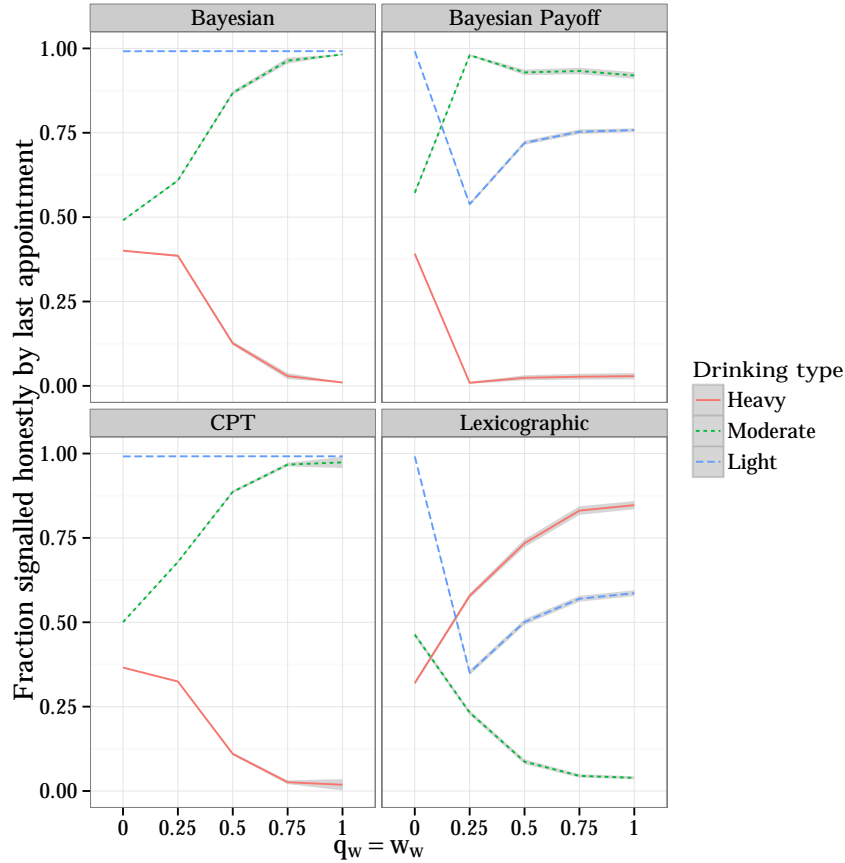

Fig. 2 Impact of social learning on trends in honest signalling after 1000 rounds, mean with 95% confidence interval over 100 runs

Aside from the lexicographic decision rule, the general tendency is towards less honest signalling by heavy drinkers, which is accompanied by a slight increase in referrals for the Bayesian, and CPT rules. For these decision models, this is because social learning exacerbates the existing tendency of heavy drinkers to impersonate moderate drinkers, who behave more honestly as heavy drinkers become less so.

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

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

4.3 Sensitivity Analysis

In this section we present a brief overview of the sensitivity analysis, followed by selected results highlighting the global effect of changes to perceived payoffs and degree of bias towards honesty, as well as social learning within women. The full results for the sensitivity analysis covering all sixteen emulators are available in appendix 7.

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

Perhaps surprisingly, variance attributable to social learning between midwives is relatively low, with neither the weight nor probability accounting for more than 5% of variance in any measure. While there are small contributions to variance in interaction with other parameters (e.g. 4% to between groups IQR, for the interaction with the proportion of light drinkers under the Bayesian rule), this may suggest that the implementation is lacking, which we touch on in section 6.

Figure 3 gives a qualitative picture of both emulator quality, and the divergent response surfaces of the decision rules in response to variations in social learning parameters. Emulator fit is clearly imperfect, but overall behaviour is qualitatively
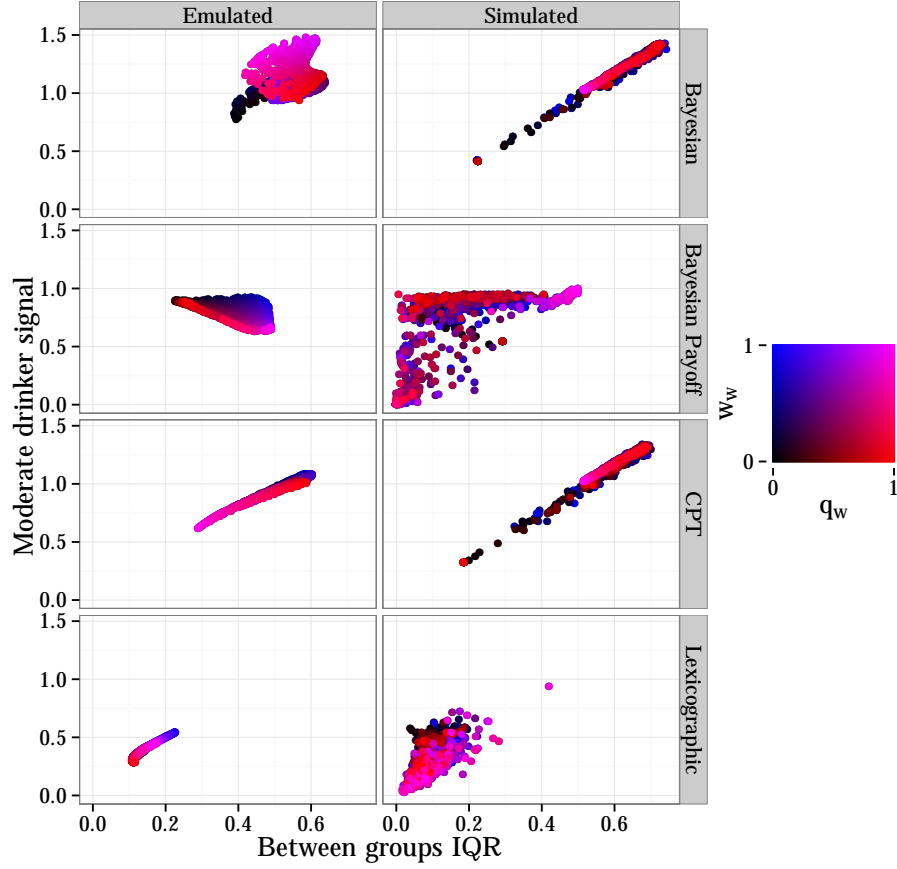
Fig. 3 Median moderate drinker signal vs median between drinking type IQR for all decision rules

similar, with both emulated and simulated plots demonstrating separation in outcome space for the decision rules.

Following from the suggestive results for social learning introducing uncertainty (section 4.2), figure 4 shows emulated points covering the parameter space in high resolution. These plots reflect the increase in uncertainty of outcome shown for the heuristic type rules, which is especially severe for the Bayesian payoff rule. They also suggest that the Bayesian decision rule is less stable under conditions where the weight of shared information is substantially higher than the probability of sharing. This suggests, perhaps unsurprisingly, that what information is shared is significant to outcomes.

For the CPT, and Bayesian decision models, the interaction of bias towards honesty, and distinction between payoffs has a significant, and non-linear effect on instability, and separability of groups. Figure 5 shows the effects, and also highlights
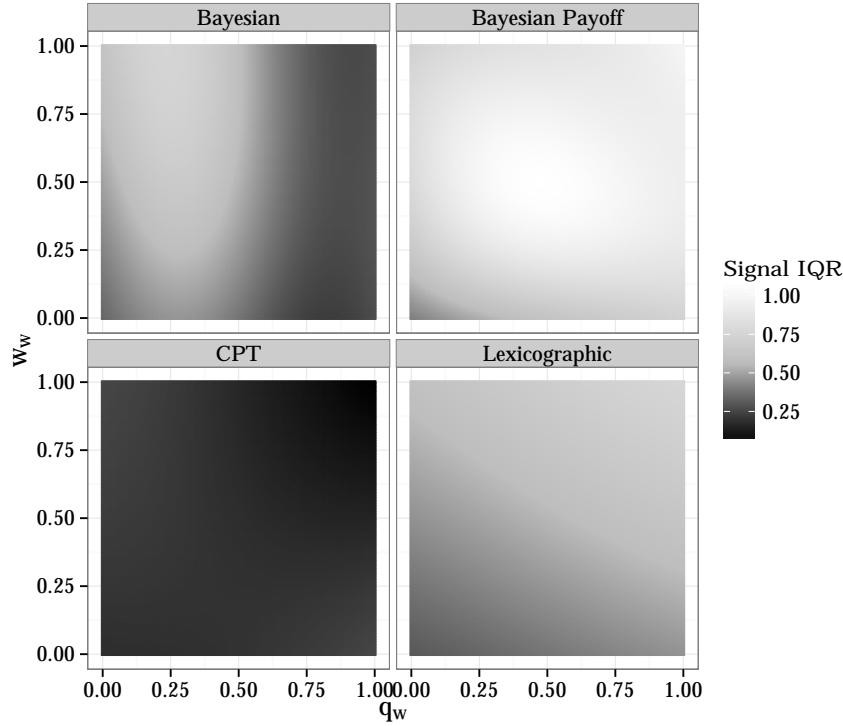
Fig. 4 Emulated moderate drinker signal IQR in response to varying q_w, and w_w

the tendency towards poor separability of groups for both the heuristic type decision rules. The response surface of the Bayesian payoff rule is slightly more nuanced than the simple Lexicographic rule. Figure 5 shows better separation, close to partial pooling[10] at high payoff distinction, with relatively modest honesty bias, which is reflected by the variance contributions of 11% and 8% respectively. For the more complex rules, the general tendency is towards less pooling for higher values of both, but with pockets where full pooling[11] occurs. The plots also suggest that the sensitivity of the CPT rule is marginally greater, which is supported by the significant contribution to variance of close to 15% for all measures of x_h.

---

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

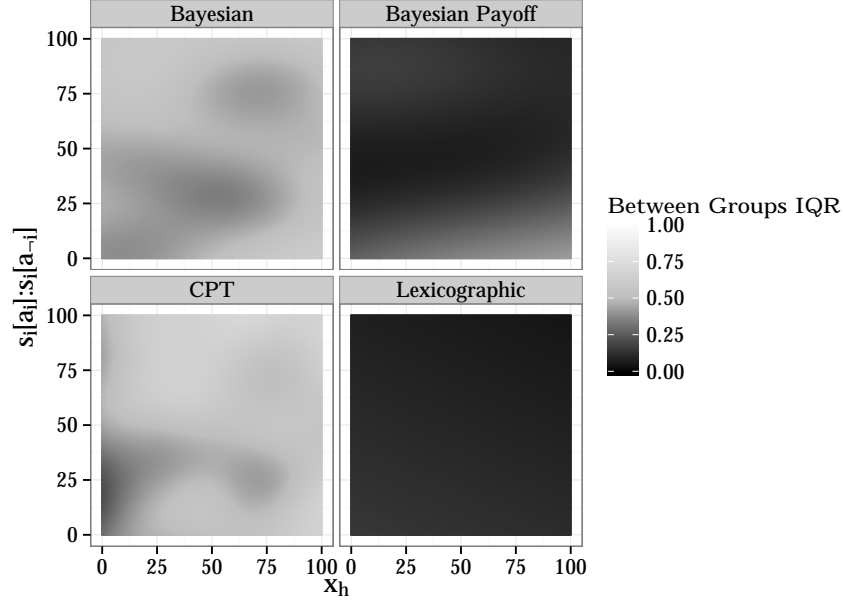[11] Indicating that all signaller types are using a single signal.

Fig. 5 Emulated between groups IQR in response to varying $s_i[a_i] : s_i[a_{\neg i}]$, and x_h

5 Discussion

From a pragmatic perspective, the differing response characteristics of the decision rules are substantial and significant. The lack of a mechanism of homophily in the very simple rules leads to a high level of uncertainty in the overall dynamics, since even irrelevant information is taken at face value and to be equally valuable. By contrast, the belief structure of the CPT, and Bayesian models represents a useful abstraction which allows the use of the relevant aspects of the information. Naturally, incorporating homophily, by, for example weighting shared information by the type of the sharer, would represent a trivial modification to the heuristic models. While to some extent this highlights the flexibility of the decision rule approach, it

would of course sacrifice the parsimony of the model. This is an important consideration, given that part of the argument in favour of a decision theoretic approach lies in the minimal nature of the behavioural rules, which can be seen as occupying a middle ground between parameter and model.

One of the notable features of the results is the similarity in behaviour of the two classes of decision rule. To some extent this reflects poorly on the most complex rule, CPT, which diverges only minimally in behaviour from the Bayesian model. This might be to a degree anticipated, in that the payoffs are unrealistic, which limits the utility of the CPT approach. Additionally, work by Glöckner and Pachur [16] has shown that there is considerable variation in individual parameters for the decision model, whereas we have let them remain homogeneous here. In the same vein, utility functions should arguably vary between individual agents, which could potentially be addressed by replacing the fixed payoffs used here with a distribution. With this said, the significant increase in complexity, which entails both additional parameters and increases to simulation time may auger for a middle ground, particularly where elicitation of payoffs is impractical. This, together with the variability associated with the heuristic type decision rules speaks to a trade off between capturing reality, and replicating it.

Continuing the discussion of the issues raised by the representation of payoffs, the temporal aspect is significant, in that there is a timing difference in payoffs, since while the potential social pain of disclosure is immediate, the health outcome comes only later. In light of this, that there is a known impact of time on perceived utility Thaler [44] suggests that incorporating some form of temporal discounting (e.g. exponential [40], or hyperbolic [1]), or a decision model which explicitly treats intertemporal choice, such as the CPT-like model of Loewenstein and Prelec [23], is warranted.

As noted in section 3.2, the impact of social learning in midwives is surprisingly minimal, where it might be expected to play a more significant role in reality. A possible explanation for this lies in the implementation, which may place an excessive constraint on how much information midwives can share. The restriction to sharing only after a referral, together with the disparity in population sizes, and random allocation of appointments leads to midwives rarely having more than a single interaction with woman to pass on to their colleagues. Furthermore, because midwives are only informed of the true type if a referral occurs, they have an inherent myopia since until they have evidence of deception they will not refer, with said evidence difficult to obtain without a referral.

In reality it might be anticipated that midwives would not withhold judgement, and would pass on concerns about specific women to their colleagues, or that particularly dramatic stories would persist and be passed. This might be addressed by incorporating noisy type information [12], capturing the unintentional information transmitted during appointments, together with a relaxation of the assumptions about when information may be shared and more sophisticated model of information transmission in general. This also highlights an advantage of the BACCO approach, in diagnosing issues with model design by giving insight into parameters which are contributing inappropriately to variance in output. Coupled with the ability of

emulators to rapidly explore parameter space, this clearly suggests that statistical emulation is a powerful tool to support simulation based approaches.

6 Conclusion

The conclusions that can be drawn about the behaviours of real life women, and their midwives, are necessarily limited by the paucity of data available to validate the model. While qualitative trends offer some indication, they are limited in scope, and do not permit strong claims about the drivers of disclosure. As such, further work will focus on applying the model to domains where validation data is more available, which will support a more comprehensive evaluation of the model discrepancy. With this said, the trends reported by Alvik et al [2], and Phillips et al [34] are borne out by the model, and predictions from the two more complex rules suggest that encouraging information sharing between women may encourage disclosure, but at the expense of reducing accuracy. By contrast, if one takes the view that a Lexicographic model is a better approximation of real behaviour, then outcomes can best be influenced by controlling how far midwives punish their women socially. We would however suggest that there are better reasons than the outputs of a simulation for doing so.

More broadly, the results demonstrate the logistical feasibility, and its utility as a 'tool for thinking', of an agent model grounded in decision theory. The results also make clear that deciding the operationalisation of the decision making is of key significance.

Acknowledgements

References

[1] Ainslie G (1991) Derivation of "Rational" economic behavior from hyperbolic discount curves. The American Economic Review 81(2):334–340
[2] Alvik A, Haldorsen T, Groholt B, Lindemann R (2006) Alcohol consumption before and during pregnancy comparing concurrent and retrospective reports. Alcoholism: Clinical & Experimental Research 30(3):510–5

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

[4] Bijak J, Hilton J, Silverman E, Cao VD (2013) Reforging the wedding ring:. Demographic Research 29:729–766

[5] Booij AS, Praag BMS, Kuilen G (2009) A parametric analysis of prospect theorys functionals for the general population. Theory and Decision 68(1-2):115–148

[6] Byrnes JP, Miller DC, Schafer WD (1999) Gender differences in risk taking: A meta-analysis. Psychological Bulletin 125(3):367–383

[7] Carnell R (2012) LHS: Latin Hypercube Samples. R package version 0.10

[8] Department of Health (2008) Pregnancy and alcohol. Department of Health, London

[9] Epstein JM (2008) Why model? Journal of Artificial Societies and Social Simulation 11(4):12, URL `http://jasss.soc.surrey.ac.uk/11/4/12.html`

[10] Epstein JM (2014) Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science. Princeton University Press

[11] Epstein JM, Axtell RL (1994) Growing Artificial Societies: Social Science from the Bottom Up. Complex Adaptive Systems, Brookings Institution Press

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

[13] Fox CR, Tversky A (1998) A belief-based account of decision under uncertainty. Management Science 44(7):879–895

[14] Gigerenzer G (2004) Fast and frugal heuristics: The tools of bounded rationality. In: Koehler DJ, Harvey N (eds) Blackwell Handbook of Judgment and Decision Making, Blackwell Publishing, pp 62–88

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

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

[17] Gomberg ES (1988) Alcoholic women in treatment: The question of stigma and age. Alcohol and Alcoholism 23(6):507–14

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

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

[20] Insua DR, Rios J, Banks D (2009) Adversarial risk analysis. Journal of the American Statistical Association 104(486):841–854

[21] Kennedy MC (2004) CTCD: GEM software. `http://ctcd.group.shef.ac.uk/gem.html`, accessed: 10/7/2014

[22] Kreps DM, Cho I (1987) Signalling games and stable equilibria. The Quarterly Journal of Economics 102(2):179–221

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

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

[25] National Institute for Health and Care Excellence (2010) Antenatal care. CG62. NICE, London

[26] National Institute for Health and Care Excellence (2010) Pregnancy and complex social factors. CG110. NICE, Manchester

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

[28] Nilsson H, Rieskamp J, Wagenmakers EJ (2011) Hierarchical bayesian parameter estimation for cumulative prospect theory. Journal of Mathematical Psychology 55(1):84–93

[29] Oakley A, Strange V, Bonell C, Allen E, Stephenson J (2006) Process evaluation in randomised controlled trials of complex interventions. British Medical Journal 332(7538):413–6

[30] Oakley J, O'Hagan A (2002) Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika 89(4):769–784

[31] Oakley JE, O'Hagan A (2004) Probabilistic sensitivity analysis of complex models: a bayesian approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66(3):751–769

[32] Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441(7090):223–6

[33] Padoa-Schioppa C, Assad JA (2008) The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nature Neuroscience 11(1):95–102

[34] Phillips D, Thomas K, Cox H, Ricciardelli LA, 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

[35] R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna

[36] Radcliffe P (2011) Substance-misusing women: Stigma in the maternity setting. British Journal of Midwifery 19(8):497–506

[37] Redshaw M, Henderson J (2014) Safely delivered: A national survey of womens experiences of maternity care 2014. The National Perinatal Epidemiology Unit, Oxford

[38] Resnick M (1994) Turtles, Termites and Traffic Jams: Explorations in Massively Parallel Microworlds. The MIT Press, Cambridge, MA

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

[40] Samuelson PA (1937) Note on measurement of utility. The Review of Economic Studies 4(2):155–161

[41] Silverman E, Bijak J, Noble J (2011) Feeding the beast: Can computational demographic models free us from the tyranny of data? In: Lenaerts T, Giacobini

M, Bersini H, Bourgine P, Dorigo M, Doursat R (eds) Advances in Artificial Life, ECAL 2011, MIT Press, vol 276, pp 747–754

[42] Silverman E, Bijak J, Hilton J, Cao VD, Noble J (2013) When demography met social simulation: A tale of two modelling approaches. Journal of Artificial Societies and Social Simulation 16(4):9

[43] Sokol RJ, Martier SS, Ager JW (1989) The t-ace questions: Practical prenatal detection of risk-drinking. American Journal of Obstetrics and Gynecology 160(4):863 – 870

[44] Thaler R (1981) Some empirical evidence on dynamic inconsistency. Economics Letters 8(3):201–207

[45] Thiele JC, Kurth W, Grimm V (2014) Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. Journal of Artificial Societies and Simulation 17(3):11

[46] Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5(4):297–323

[47] Waldherr A, Wijermans N (2013) Communicating social simulation models to sceptical minds. Journal of Artificial Societies and Social Simulation 16(4):13

7 Sensitivity Analysis

This section provides complete variance based sensitivity analysis results for the disclosure game model. Each subsection gives results for one simulation output under all four decision rules, with tables providing the percentage of overall variance attributable to the individual parameters, emulator quality statistics, and the 5 most significant interaction contributions to variance in the output.

7.1 Median Moderate Drinker Signalling

Table 4 Median moderate drinker signalling parameter sensitivity

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

Table 5 Median moderate drinker signalling emulator statistics

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

Table 6 Top 5 Interactions terms for CPT decision rule.

Parameter	Variance
$x_h * s_i[a_i] : s_i[a_{\neg i}]$	20.814
$p_w(l) * s_i[a_i] : s_i[a_{\neg i}]$	5.698
$p_w(l) * x_h$	2.895
$s_i[a_i] : s_i[a_{\neg i}] * w_w$	2.799
$s_i[a_i] : s_i[a_{\neg i}] * q_m$	1.686

Table 7 Top 5 Interactions terms for Bayesian decision rule.

Parameter	Variance
$p_w(l)*s_i[a_i] : s_i[a_{\neg i}]$	17.270
$p_w(l)*q_m$	6.054
q_m*w_w	3.814
$s_i[a_i] : s_i[a_{\neg i}]*q_m$	3.538
$s_i[a_i] : s_i[a_{\neg i}]*w_w$	3.084

Table 8 Top 5 Interactions terms for Lexicographic decision rule.

Parameter	Variance
$p_m(l)*p_m(m)$	15.331
$p_m(m)*p_w(l)$	3.682
$p_m(l)*p_w(l)$	3.581
$p_m(m)*q_m$	0.349
$p_m(l)*q_m$	0.279

Table 9 Top 5 Interactions terms for Bayesian Payoff decision rule.

Parameter	Variance
$p_w(l)*w_w$	4.045
$x_h*s_i[a_i] : s_i[a_{\neg i}]$	1.856
$p_w(l)*q_m$	1.231
$s_i[a_i] : s_i[a_{\neg i}]*q_m$	0.997
q_m*w_w	0.929

7.2 Median Between Groups IQR

Table 10 Median between groups IQR parameter sensitivity

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

Table 11 Median between groups IQR emulator statistics

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

Table 12 Top 5 Interactions terms for CPT decision rule.

Parameter	Variance
$x_h * s_i[a_i] : s_i[a_{\neg i}]$	19.551
$p_w(l) * s_i[a_i] : s_i[a_{\neg i}]$	3.838
$s_i[a_i] : s_i[a_{\neg i}] * w_w$	2.450
$p_w(l) * x_h$	2.337
$s_i[a_i] : s_i[a_{\neg i}] * q_m$	2.046

Table 13 Top 5 Interactions terms for Bayesian decision rule.

Parameter	Variance
$s_i[a_i] : s_i[a_{\neg i}]*q_m$	4.284
$p_w(l)*q_m$	3.866
$p_w(l)*s_i[a_i] : s_i[a_{\neg i}]$	2.943
$x_h*s_i[a_i] : s_i[a_{\neg i}]$	2.680
q_m*w_w	2.282

Table 14 Top 5 Interactions terms for Lexicographic decision rule.

Parameter	Variance
$p_m(l)*p_m(m)$	12.046
$p_m(m)*p_w(l)$	5.054
$p_m(l)*p_w(l)$	3.005
$p_m(l)*w_w$	0.819
$p_m(m)*w_w$	0.757

Table 15 Top 5 Interactions terms for Bayesian Payoff decision rule.

Parameter	Variance
$p_w(l)*s_i[a_i] : s_i[a_{\neg i}]$	12.883
$p_w(l)*w_w$	5.667
$p_w(l)*x_h$	2.447
$x_h*s_i[a_i] : s_i[a_{\neg i}]$	2.360
$p_m(m)*p_w(l)$	1.919

7.3 Median Moderate Drinker Signalling IQR

Table 16 IQR of median moderate drinker signalling parameter sensitivity

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

Table 17 IQR of median between groups IQR emulator statistics

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

Table 18 Top 5 Interactions terms for CPT decision rule.

Parameter	Variance
$x_h * s_i[a_i] : s_i[a_{\neg i}]$	17.377
$p_w(m) * s_i[a_i] : s_i[a_{\neg i}]$	3.356
$s_i[a_i] : s_i[a_{\neg i}] * w_w$	3.036
$s_i[a_i] : s_i[a_{\neg i}] * q_m$	2.067
$p_w(l) * s_i[a_i] : s_i[a_{\neg i}]$	1.721

Table 19 Top 5 Interactions terms for Bayesian decision rule.

Parameter	Variance
$p_w(m)*s_i[a_i] : s_i[a_{\neg i}]$	4.188
$x_h*s_i[a_i] : s_i[a_{\neg i}]$	3.120
$s_i[a_i] : s_i[a_{\neg i}]*q_m$	2.423
$p_w(l)*s_i[a_i] : s_i[a_{\neg i}]$	2.279
$p_w(l)*q_m$	1.489

Table 20 Top 5 Interactions terms for Lexicographic decision rule.

Parameter	Variance
$p_m(m)*p_w(l)$	12.068
$p_m(l)*p_w(l)$	6.794
$p_m(l)*p_m(m)$	5.567
$p_m(m)*q_m$	0.886
$p_m(l)*q_m$	0.692

Table 21 Top 5 Interactions terms for Bayesian Payoff decision rule.

Parameter	Variance
$p_w(l)*p_w(m)$	8.357
$p_w(l)*w_w$	7.431
$p_w(l)*s_i[a_i] : s_i[a_{\neg i}]$	4.882
$p_w(l)*q_m$	2.346
$p_w(l)*w_m$	2.025

7.4 Between Groups IQR IQR

Table 22 IQR of median between groups IQR parameter sensitivity

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

Table 23 IQR of median between groups IQR emulator statistics

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

Table 24 Top 5 Interactions terms for CPT decision rule.

Parameter	Variance
$x_h * s_i[a_i] : s_i[a_{\neg i}]$	18.626
$p_w(l) * s_i[a_i] : s_i[a_{\neg i}]$	3.312
$s_i[a_i] : s_i[a_{\neg i}] * w_w$	2.823
$s_i[a_i] : s_i[a_{\neg i}] * q_m$	2.694
$x_h * q_m$	1.022

Table 25 Top 5 Interactions terms for Bayesian decision rule.

Parameter	Variance
$p_w(l)*s_i[a_i] : s_i[a_{\neg i}]$	7.947
$x_h*s_i[a_i] : s_i[a_{\neg i}]$	4.048
$s_i[a_i] : s_i[a_{\neg i}]*q_m$	3.134
$p_w(l)*q_m$	2.307
$p_m(m)*s_i[a_i] : s_i[a_{\neg i}]$	2.232

Table 26 Top 5 Interactions terms for Lexicographic decision rule.

Parameter	Variance
$p_m(l)*p_m(m)$	8.659
$p_m(m)*p_w(l)$	3.726
$p_m(l)*p_w(l)$	3.237
$p_m(l)*w_w$	1.564
$p_m(m)*w_w$	1.558

Table 27 Top 5 Interactions terms for Bayesian Payoff decision rule.

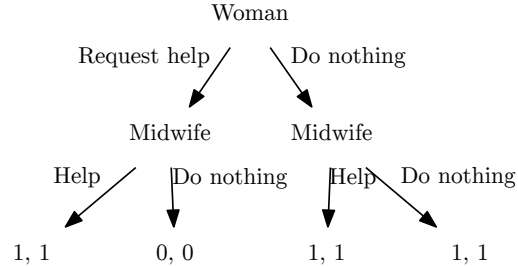| Parameter	Variance
q_m*w_w	3.830
$p_w(l)*p_w(m)$	3.808
$p_w(l)*s_i[a_i] : s_i[a_{\neg i}]$	3.401
$p_w(l)*q_m$	2.385
$x_h*s_i[a_i] : s_i[a_{\neg i}]$	2.294

8 Disclosure Game Model Development

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

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

A very simple way to represent this would be a game with two players, who both have two possible moves - ask for help, or not; and refer, or not (fig 6). Since both

parties are invested in the outcome of the pregnancy, we might allow them to share the same payoff if everything ends well.

Fig. 6 A very simple two player game. The only time things in this very restricted world obviously end poorly, is if the woman asks for help but does not get any. This implies that a rational player would always refer if asked for help, and is indifferent otherwise - in other words, there are three possible Nash equilibriums.

Woman

Request help /   \ Do nothing

Midwife   Midwife

Help / | Do nothing   Help \ Do nothing

1, 1   0, 0   1, 1   1, 1

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

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

To resolve this, we can do exactly that, and cast it as a signalling game (fig 7), with three types of player, corresponding to categories of drinking behaviour (light, moderate, and heavy). Each of these types of player, will play a different game. This also introduces a third player, who we will call nature. Nature takes the first move, and decides the type of the woman according to some probability distribution; in this case we will allow the probability of types to be uniform. This changes the dynamics of play substantially, since the midwife can no longer be certain of which game they are playing, and hence which move yields the best outcome. We must also amend the moves, and payoffs slightly. The woman now claims to be one of the types, and may send a signal to say, for example, that they are a heavy drinker. We will also modify the common payoffs to allow light drinkers to get the best outcome no matter what, and moderate and heavy types to get the best outcome only if referred. We can also differentiate between the consequences of not getting help for these types by letting heavy drinkers have a very negative outcome, and moderate drinkers a slight one.
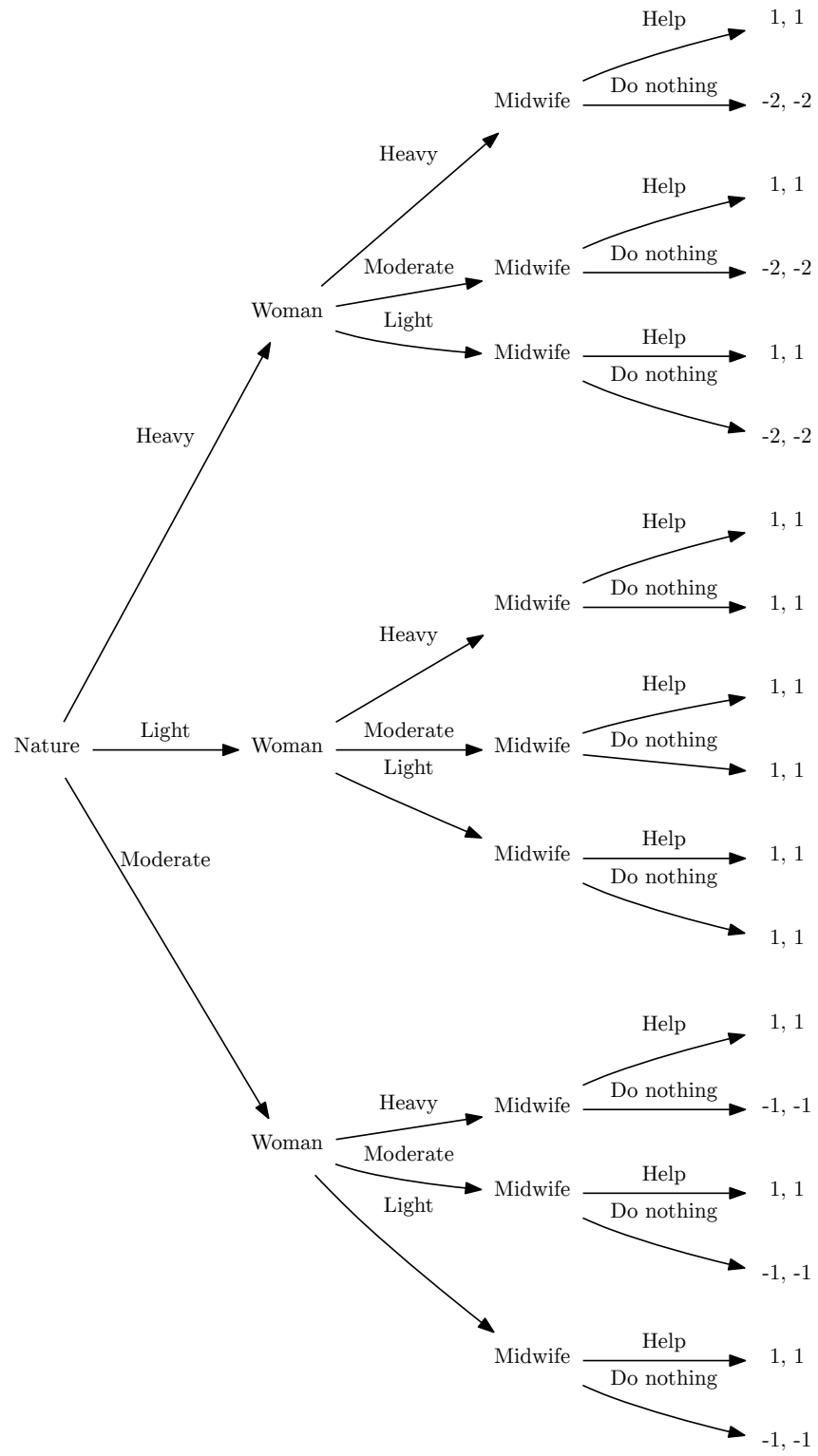
Fig. 7 A less simple two player signalling game.

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

At this juncture, we do not gain much further from the game representation, and instead separate it into multiple decision problems.

Breaking the game down into separate decision problems can be achieved by treating the moves of the other players as a chance node, and omitting moves by nature that are known to them. For women, there are two such nodes, corresponding to the move by nature determining the type of midwife they play against, and the midwife's action. Midwives have simpler problem with only a single chance node, because the woman's move is known to them. Figure 8 shows the structure of the resulting decision problems. Note that there are in fact three distinct decision problems for the three types of woman, since the move by nature determining their type is known to them.
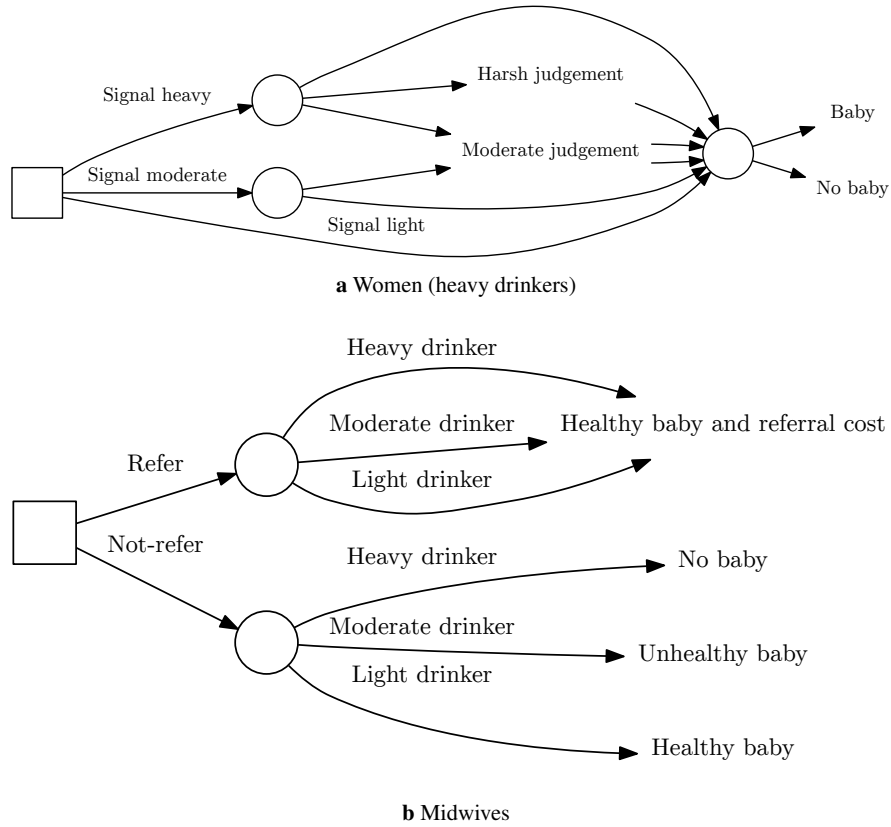
a Women (heavy drinkers)

b Midwives

Fig. 8 Influence diagrams, showing the game broken into two decision problems. Squares indicate a decision node, while circles are (from the perspective of the agent) chance nodes

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

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

9 Simulation Schedule

In this section we give the step by step process for a single run of the disclosure game simulation.

1. Generate 1000 women, and place them in a queue
2. Generate 100 midwives.
3. For each round of the game

    a. Take 100 women from the queue
    b. Pair each one with a random midwife
    c. For each pair
        i. The woman sends a signal
        ii. The midwife refers or not based on the signal
        iii. The woman is informed of their payoff, the midwife's type, and whether they are referred
        iv. The woman updates her beliefs
        v. The midwife stores the game in their memory
        vi. If the woman is referred
            A. The midwife is informed of the woman's true type
            B. The midwife retrospectively updates their beliefs using the true type, and memories of any games with this woman
            C. The midwife is now eligible to share their memories of games with this woman
    d. Women who have not been referred or had their baby, join the back of the queue
    e. New women are generated to replace those referred, or delivered
    f. The new women are added to the back of the queue
    g. For each referred or birthed woman
        i. With probability p, their memory of games is shared with the active women
        ii. They are removed from simulation
    h. The active women update their beliefs
    i. For each midwife with information to share
        i. With probability p, their memory of games with the referred woman is shared
        ii. The memory is no longer eligible to be shared
    j. The midwives update their beliefs

10 Agent Examples

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

10.1 Lexicographic Heuristic

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

10.2 Bayesian Payoff

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

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

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

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

10.3 Bayesian Risk Minimisation

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

$$R_w(l,l) = \sum_{i \in A_m} \sum_{j \in \Theta} -u_w(l,i,l,j)p(j)p(i|l)$$

$$R_w(l,l) = -u_w(l,r,l,l)p(l)p(r|l) - u_w(l,n,l,l)p(l)p(n|l)$$
$$- u_w(l,r,l,m)p(m)p(r|l) - u_w(l,n,l,m)p(m)p(n|l)$$
$$- u_w(l,r,l,h)p(h)p(r|l) - u_w(l,n,l,h)p(h)p(n|l)$$

$$u_w(l,i,l,j) = 10$$

$$R_w(l,l) = -10p(l)p(r|l) - 10p(l)p(n|l) - 10p(m)p(r|l) - 10p(m)p(n|l)$$
$$- 10p(h)p(r|l) - 10p(h)p(n|l)$$

$$p(l) = \frac{1+0}{1+1+1+3} = \frac{1}{6}$$

$$p(m) = \frac{1+0}{1+1+1+3} = \frac{1}{6}$$

$$p(h) = \frac{1+3}{1+1+1+3} = \frac{2}{3}$$

$$p(r|l) = \frac{1+0}{1+1+2} = \frac{1}{4}$$

$$p(n|l) = \frac{1+2}{1+1+2} = \frac{3}{4}$$

$$R_w(l,l) = -10 \cdot \frac{1}{6} \cdot \frac{1}{4} - 10 \cdot \frac{1}{6} \cdot \frac{3}{4} - 10 \cdot \frac{1}{6} \cdot \frac{1}{4} - 10 \cdot \frac{1}{6} \cdot \frac{3}{4} - 10 \cdot \frac{2}{3} \cdot \frac{1}{4} - 10 \cdot \frac{2}{3} \cdot \frac{3}{4}$$

$$= -10$$

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

10.4 Descriptive Decision Theory

Once again, we return to the light drinker example. The inferential aspects are identical with the more complex Bayesian risk minimisation algorithm, hence $p(j)p(i|l)$, and $u_w(l,i,l,j)$ remain the same, but the agent additionally calculates $v(u_w(l,i,l,j))w^+(p(j))w^+(p(i|l))$. For the CPT parameters, the values are those originally given by Tversky and Kahneman [46] and used in the actual simulations which are given in table 28.

Table 28 CPT parameters.

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

$$\alpha = 0.88$$

$$\gamma = 0.61$$

$$p(l) = \frac{1}{6}$$

$$p(m) = \frac{1}{6}$$

$$p(h) = \frac{2}{3}$$

$$p(r|l) = \frac{1}{4}$$

$$p(n|l) = \frac{3}{4}$$

$$u_w(l,i,l,j) = 10$$

$$f = (10; \frac{1}{24}, 10; \frac{1}{8}, 10; \frac{1}{24}, 10; \frac{1}{8}, 10; \frac{1}{6}, 10; \frac{1}{2})$$

$$f^+ = f, f^- = ()$$

$$n = 5$$

$$v(u_w) = f(u_w) = u_w^\alpha$$

$$v(u_w) = 10^{0.88} = 7.59$$

$$\pi_0^+ = w^+(\frac{1}{24} + \frac{1}{8} + \frac{1}{24} + \frac{1}{8} + \frac{1}{6} + \frac{1}{2}) - w^+(\frac{1}{8} + \frac{1}{24} + \frac{1}{8} + \frac{1}{6} + \frac{1}{2})$$

$$= w^+(1) - w^+(\frac{23}{24})$$

$$= 0.19$$

$$\pi_1^+ = w^+(\frac{1}{8} + \frac{1}{24} + \frac{1}{8} + \frac{1}{6} + \frac{1}{2}) - w^+(\frac{1}{24} + \frac{1}{8} + \frac{1}{6} + \frac{1}{2})$$

$$= w^+(\frac{23}{24}) - w^+(\frac{5}{6})$$

$$= 0.17$$

$$\pi_2^+ = w^+(\frac{1}{24} + \frac{1}{8} + \frac{1}{6} + \frac{1}{2}) - w^+(\frac{1}{8} + \frac{1}{6} + \frac{1}{2}) = w^+(\frac{5}{6}) - w^+(\frac{19}{24})$$

$$= 0.04$$

$$\pi_3^+ = w^+(\frac{1}{8} + \frac{1}{6} + \frac{1}{2}) - w^+(\frac{1}{6} + \frac{1}{2}) = w^+(\frac{19}{24}) - w^+(\frac{2}{3})$$

$$= 0.09$$

$$\pi_4^+ = w^+(\frac{1}{6} + \frac{1}{2}) - w^+(\frac{1}{2}) = w^+(\frac{2}{3}) - w^+(\frac{1}{2})$$

$$= 0.09$$

$$\pi_5^+ = w^+(\frac{1}{2})$$

$$= 0.42$$

$$V(f) = V(f^+) + V(f^-) = V(f^+) + 0$$

$$V(f^+) = \sum_i^n \pi_i^+(f^+) v_i^+(f^+) = 7.59$$