# Deciding to Disclose

## A Decision Theoretic Agent Model of Pregnancy and Alcohol Misuse

Jonathan Gray, Jakub Bijak, and Seth Bullock

## Abstract

### *Background*

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.

### *Objective*

The primary purpose of this work is to demonstrate the significance of how the decision making process is formulated, by contrasting four distinct decision rules against one another and exploring a simple form of information sharing, the analysis is supported by the use of statistical emulators for a full exploration of the parameter space. We also demonstrate the utility of an approach which it is hoped contributes towards addressing concerns about the ad hoc character of Agent Based Modelling (ABM), by providing a strong theoretical grounding for the reasoning processes of individual agents. To this end we aim to show that these simple rules, operating in an inescapably artificial scenario are nonetheless capable of reproducing trends reported in the literature.

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

## *Methods*

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 information sharing 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.

## *Results*

Selected results showing the ability of all decision rules to reproduce qualitative trends noted in the literature are provided, together with a sensitivity analysis, and comparative heat maps produced using GEMs demonstrating the significance of the precise implementation of the decision making framework.

## *Comments*

We note that the model omits the overwhelming complexity of the real disclosure scenario where hardly any quantitative data are available, and is presented largely in the spirit of a convenient demonstration of the methodology. Clearly a domain where there is sufficient data to permit a more comprehensive approach to validation of model outcomes is desirable, and will form the basis of our future work.

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

# 1 Introduction

The case in favour of Agent Based Modelling (ABM) as a general analytical approach has been made numerously, and elegantly (e.g. Epstein and Axtell (1994); Resnick (1994); Axelrod (1997); Gilbert (1999); Macy and Willer (2002); Silverman et al (2011, 2013); Epstein (2014), 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, tending to be a reflection of the assumptions of the modeller rather than empirically or theoretically grounded (Waldherr and Wijermans, 2013). 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 (Kreps and Cho, 1987) with a basic form of intragroup information sharing. 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, and has been associated with many potentially negative consequences. For example, Andersen et al (2012) report results from a large scale Danish cohort study suggesting that even low levels of consumption in early pregnancy increase the risk of spontaneous abortion, although Savitz (2012) has suggested this may be attributable to a previously known link to absence of mornining sickness. Risk continues to the point of birth - Kesmodel et al (2002) found a heightened risk to the infant - into childhood, with a metastudy by Latino-Martel et al (2010) finding evidence on an increased risk of childhood acute myeloid leukemia (AML). Harm may also extend even further, and a review by Huizink and Mulder (2006) concluded that maternal alcohol consumption could be a contributing factor to Attention Defecit Hyperactivity Disorder (ADHD), and other learning impairments, but note methodological issues in a number of the papers. There is not, however, a clear consensus, with Gray and Henderson (2006) finding no evidence of harm below 1.5 UK units[1] per day. 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 (National Institute for Health and Care Excellence, 2010a), with similar advice from the UK Department of Health (2008).

Turning more specifically to disclosure of alcohol use during pregnancy, research is relatively sparse, although qualitative trends are reported by Phillips et al (2007) and Alvik et al (2006). The former explored factors impacting disclosure through a small case study, highlighting the need to build up rapport 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

[1] Roughly half a pint of beer, a single 25ml measure of spirits, or one 175ml glass of wine.

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 arenas where randomised control trials (RCTs) are not viable for ethical, or financial reasons. With this said, the lack of a strong quantitative evidence base against which to validate the behaviour of the model augers for caution in interpreting the results, and a necessary reminder that in this instance the model is primarily a tool for formalisation of the thought process (Epstein, 2008), rather than a machine for predicting.

A game theoretic approach to generating an abstract form of the problem gives a convenient, and well known framework to reason about the processes involved in the scenario. While scenarios may map to a number of games, exploring one candidate game still allows for a principled comparison between interpretations, and enforces explicit assumptions. 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. Naturally, this does not preclude the incorporation of strategic refinement, since decision rules are to a great extent modular, and as demonstrated in this paper can be exchanged without altering the underpinning decision problem. In addition, rules are agnostic as to where the information used derives from, suggesting room for multi-stage processes. As a corollary, the decision problem agents attempt to answer can change, allowing agents' behaviour in novel problems to be informed by beliefs derived under other conditions.

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

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 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 provide a brief review of the methodological context (section 2), before outlining the model (section 3), and experiments (section 4), with selected results (section 5), followed by a discussion contrasting the decision models (section 6), and conclusions (section 7).

## 2 Previous Research

This section presents a brief overview of previous work relating to the theoretical backdrop to this approach, addressing in turn signalling games, normative decision theory, heuristic decision making, and descriptive decision theory.

### *2.1 Signalling Games*

The majority of classical game theory focuses on strategic decision making, in scenarios where all players have complete information about all aspects of the game. An alternative, perhaps more common situation, is that players have incomplete information, i.e. their knowledge of the state of the world is in some way deficient. Harsanyi (1967) introduced the concept of a Bayesian game, resolving the problems introduced by the incomplete information scenario by allowing the possible variations on the rules to be treated as subgames. This adds an additional player - nature, to the game, where nature takes the first move thereby deciding which subgame is played. Nature is assumed to make their move by lottery, and where the probability distribution governing the lottery is known to all players this permits the game to be formulated as one of complete information.

Here, we are specifically interested in signalling games (Spence, 1973; Kreps and Cho, 1987), where one player holds some private information which may be communicated (or not) by means of a signal. This basic form has been widely applied, with substantial interest in what conditions permit honest signalling as Nash equilibria or Evolutionarily Stable Strategies (ESS). Grafen (1990), following from a suggestion by Zahavi (1975), proposed that if signals intended to indicate mate quality exacted a cost on the signaller (e.g. peacock tail feathers), then honest signalling would constitute an ESS. Similar results have also been demonstrated in a game of job market signalling, where signal cost was differentiated by type (Spence, 1973). Costly signalling has also been suggested as an explanation of behaviour that at first glance appears counter intuitive, for example Godfray (1991) applied the idea to the food solicitation behaviour of chicks, where a stronger signal (i.e. more, or louder chirping) carries a risk of being eaten by predators. Moving beyond animal behaviour, Sosis (2003) considered the implications of ritual behaviour, in the context of religion, representing a costly signal, an idea subsequently extended by Henrich (2009) to include cultural transmission, and Wildman and Sosis (2011) to introduce group differentiation.

Other work augments the signalling game model, for example Austen-Smith and Fryer Jr. (2005) add a second ‘peer group’ audience signalling game to the original

Spence game in an effort to explain poor academic performance in some social groups, with some subsequent empirical support for the idea from Fryer Jr. and Torelli (2010). On a similar tack, Feltovich et al (2002) introduce additional noisy type information, finding that this effectively explained counter-intuitive observed behaviour where actors with every right to boast of their quality fail to do so.

## *2.2 Normative Decision Theory*

While game theory addresses rational *strategic* decision making, decision theory deals instead with rational decision making (Peterson, 2009). Taken literally, this leads to normative decision theory, where the focus is on giving the rational answer to a decision problem. An complementary view - descriptive, or behavioural decision theory, holds that the focus should instead be on giving an account of human decision making performance, complete with observed deviations from perfect rationality, which we address in section 2.4. Finally a third perspective, which to some extent overlaps this division, suggests that decisions are heuristic in nature and rational in ecological context (Gigerenzer and Goldstein, 1996) (section 2.3).

The conceptual underpinning of all of these is the central idea of expected utility, originated by Bernoulli (1954/1738) and later formalised by Von Neumann and Morgenstern (1953), which treats all decisions as gambles defined in terms of payoffs and probabilities.

Recently, several studies have explored biological correlates to aspects of expected utility. The fundamental concept, that all outcomes are comparable in a universal currency has been supported by evidence of neural correlates of decision variables (Platt and Glimcher, 1999), and following from this results from Padoa-Schioppa and Assad (2006, 2008) showing neuronal firing in the orbitofrontal cortex (OFC) corresponding to revealed preferences in monkeys. Additionally, some support for neural representation of value, and risk aversion was found by Christopoulos et al (2009). The model presented in this paper makes an explicit assumption that social decisions utilise the same process, and while this is less well supported there is some evidence to suggest involvement by the same brain region, since damage to the OFC has been shown to impair social judgements in both primates (Watson and Platt, 2012), and humans (Willis et al, 2010).

An alternative normative model of decision making is Bayesian decision theory, proposed by Robbins (1964), which is essentially the application of Bayesian style probabilities to the expected utility model. This allows probabilities used in reasoning to be subjective, which may allow for a better account of decisions from experience (see Hertwig et al (2004); Hau et al (2008) for results elucidating the distinction, and comparing the performance of several non-Bayesian models). This model has seen notable successes in practical problems (McNamara and Houston, 1980; Kristensen, 1997; Dorazio and Johnson, 2003), but suggestions by several authors that it could constitute an effective (top-down) model of learning (Tenenbaum et al, 2006; Griffiths et al, 2010), or induction (Gallistel, 2012) in the brain have attracted substantial criticism[2].

## 2.3 Heuristic Decision Making

As noted, heuristic decision making stems from a contention that Von Neumann and Morgenstern type rationality ignores the context of decision making, and a lack of correspondence between predicted and actual human decisions (see, for example the Allais paradox (Allais, 1953), and subsequent empirical support from Burke et al (1996) and Oliver (2003)). Arguably, this notion begins with Simon (1956), who suggested that humans do not attempt to make optimal choices, but to satisfice and choose the first 'good enough' option. While noting that this will often achieve the same result, the claim is that humans exhibit bounded rationality (Simon, 2000) arising from inherent limits to cognition.

Gigerenzer and Goldstein (1996) take the concept of bounded rationality further, and argue for what they term Fast and Frugal Heuristics (FFHs). This recasts rationality as bound to the context of the behaviour: a rational approach to choosing the right mate might well require checking every possible partner, but given finite time, memory, and so on rapidly becomes unachievable. On this basis, they contend that the rationality of any given decision rule can only be determined in the context of the environment (Todd and Gigerenzer, 2003), which implies that heuristics are task specific. They provide a number of heuristics for varying decision problems, e.g. recognition (Goldstein and Gigerenzer, 2002), cue ordering (Gigerenzer and Goldstein, 1999; Todd and Dieckmann, 2005), and binary decisions (Brandstätter et al, 2006).

## 2.4 Descriptive Decision Theory

While heuristic theories arguably fall under the purview of the descriptive, the wider tendency is towards what are in essence "patches" to normative models. The most influential models in this class derive from Prospect Theory (PT) (Kahneman and Tversky, 1979), which combines a set of heuristics based on observed decision behaviour (Tversky and Kahneman, 1974), with distortions to the perception of probability, and the value of outcomes (Kahneman and Tversky, 1984; Tversky and Kahneman, 1986). Tversky and Kahneman (1992) subsequently addressed issues present in their original formulation by introducing Cumulative Prospect Theory (CPT), which allows for non-binary decisions, but dispenses with the heuristic aspects of the original formulation. The essence then, is that high and low probabilities are treated differently, and the subjective value of a loss differs from the

[2] For example Bowers and Davis (2012) responding to Tenenbaum et al (2006), and Griffiths et al (2010); and Miller (2012) addressing Gallistel (2012).

equivalent gain (losing your shirt is perceived as more of a loss than winning a shirt is a gain). This last effect, known as the framing effect is particularly significant, see for example work by Toll et al (2007) examining the relationship between loss and gain framings and success rates in giving up smoking, and NICE guidance on framing of treatment options (National Institute for Health and Care Excellence, 2007).

CPT has been successful in explaining a number of anomalous results in decision tasks (see Camerer (2000) for a review), and Thaler (2000) comments to the effect that the theory is promising, albeit incomplete, lacking for example any explanation of how frames are constructed. While remaining an effective account of decision behaviour under risk, the theory does not attempt to resolve apparent inconsistencies that arise when outcomes are delayed, i.e. in situations of intertemporal choice. Historically, Discounted Utility (DU) (Samuelson, 1937), which effectively claims that the value of a thing now is exponentially greater than the promise of the same thing at some future date, has been applied to explain this. More recently, Ainslie (1991) has suggested that discounting of future outcomes is hyperbolic, rather than exponential. However neither model is complete, in that both fail to account for results from Thaler (1981) showing differing temporal discounting rates for losses and gains. Loewenstein and Prelec (1992) report additional failings in classic DU models, and propose a modified form of CPT which they suggest is able to handle both immediate, and intertemporal choice.

## 3 Disclosure Game Model

In this section we outline the disclosure game model, and give details of the four decision rules that will be examined in this paper, but begin with a brief sketch of a pregnancy in terms of encounters between a pregnant woman and a midwife.

Typically in the UK, a women will have 12 appointments with a midwife during the antenatal period. Outside of caseloading[3] teams, a woman does not generally have a named midwife, and may see a different practitioner at each appointment. 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 (National Institute for Health and Care Excellence, 2010a,b) 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 con-

[3] Under a caseloading system, women are allocated an individual, or small group of midwives who deal with their care throughout pregnancy, and birth.

sumption is at the time of writing generally at the level of local health authority, hospital trust, or according to the best judgement of the individual midwife, with no guidance provided by NICE. This commonly takes the form of average units per week, but may include Tolerance, Annoyance, Cut down, Eye-opener (T-ACE)[4] (Sokol et al, 1989) and similar measures.

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

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

### *3.1 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 make the simplifying assumption that a woman may have one of only three drinking patterns - light, 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, which has a partially common interest component. Women receive a payoff based on the health of their eventual baby, with a social cost dependent on the signal they sent and the midwife's reaction to it. Midwives receive the same health payoff as the women, but pay a cost for referring to a specialist, mirroring the organisational cost of non-routine care. Table 1 shows the three payoff matrices which together describe the game.

Taken together, this leads to a game tree that is relatively complex (figure 1 shows a simplified representation, with a detailed view of one branch in figure 2) even at the subgame level. Rather than attempt to solve for equilibria, agents treat this two player game as taking place against nature, along the lines of adversarial risk analysis (Insua et al, 2009). This effectively translates the game to a pair of decision problems (figure 3), which agents attempt to resolve at each turn using a simple decision rule, given their prior beliefs and experience of play.

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

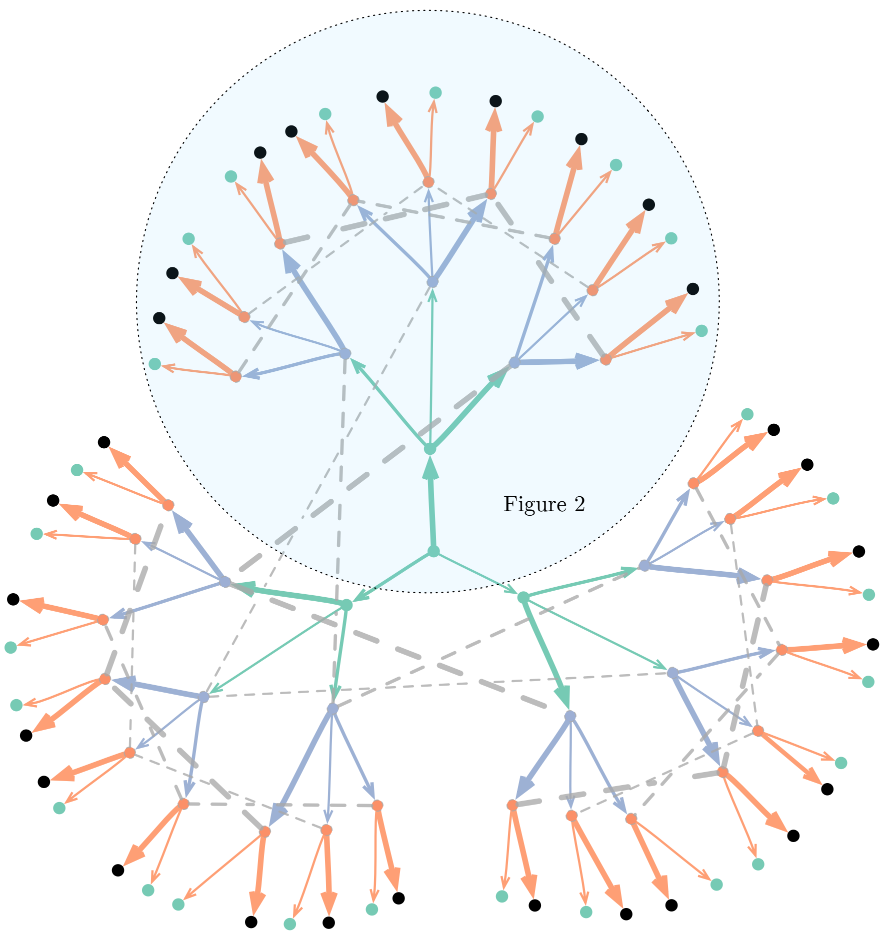



**Fig. 1** Simplified game tree. Arrows correspond to two moves by nature, followed by women, and midwives respectively. Line weight distinguishes between information sets, and possible moves, corresponding to ascending player types. For midwives, referral is shown by a heavier arrow. Light terminal nodes indicate that the game may repeat from that point, and black nodes are exits

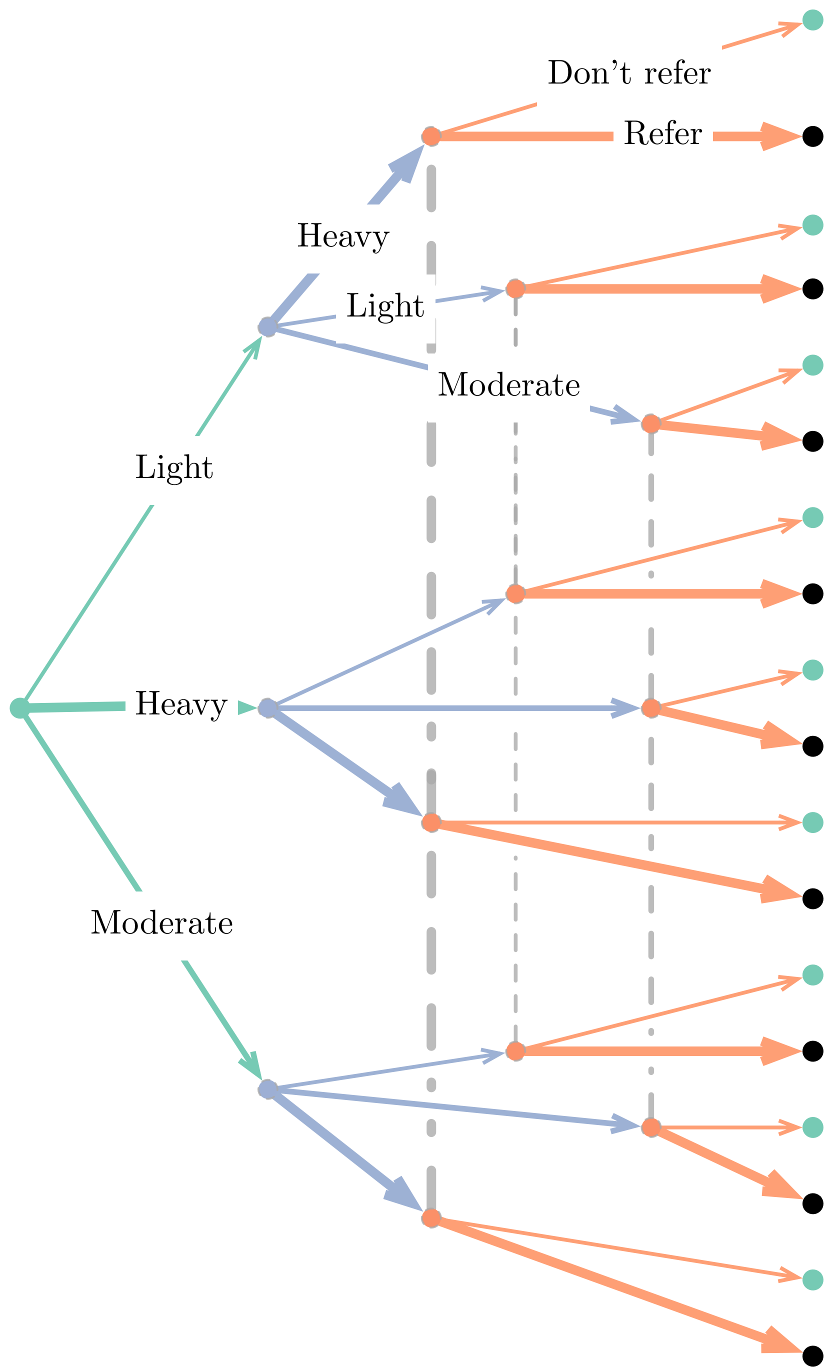
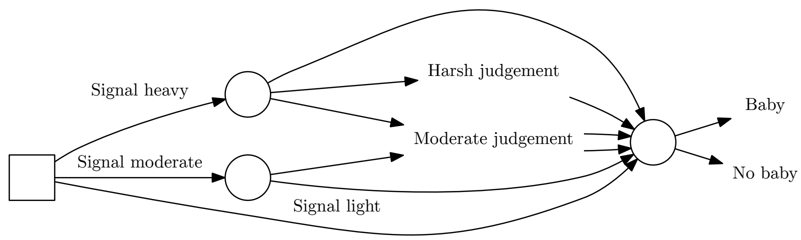
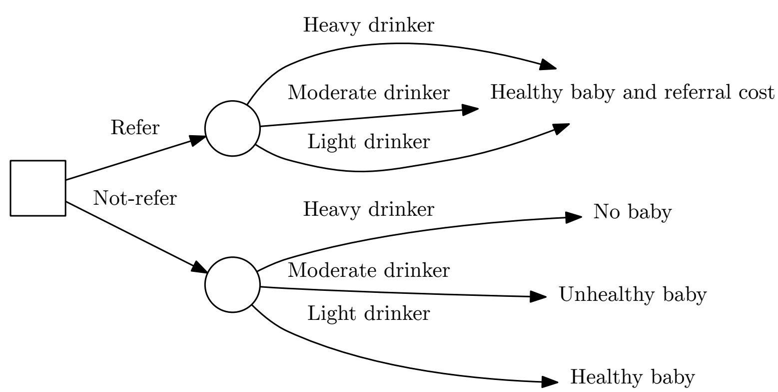



**Fig. 2** Detail view of a single game tree branch, showing the possible moves by each player, with information sets for midwives




**a** Women (heavy drinkers)




**b** Midwives

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

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 (National Institute for Health and Care Excellence, 2010a)) 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 3.2.

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.

| | | 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

**Table 1** Payoff matrices

## *3.2 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 FFH (Gigerenzer, 2004) which uses only information about payoffs given actions; this is followed by a Bayesian risk minimisation rule (2) using the same information; a second Bayesian risk rule (3) which uses information about the underlying lottery; and a two-stage CPT (Hau et al, 2008) agent (4) which is identical to 3, but uses the CPT decision rule from Tversky and Kahneman (1992). Hence, each successive decision model adds a layer of sophistication to the problem representation while retaining the same input-output characteristics.

As noted in section 3.1, agents have perfect recall, and midwives recognise women if they repeatedly encounter them. While agents recall perfectly and make use of the new information for retrospective updates, 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 are heterogeneous in their prior observations, which are assigned stochastically and constrained such that they have encountered each scenario possible for their type at least once, with exactly $k$ encounters overall.

### 3.2.1 Lexicographic Heuristic

The lexicographic heuristic (algorithm 1) follows the form of that used in Hau et al (2008), and assumes a simplified problem representation, where an action is a choice between combined lotteries. Functionally, the heuristic maintains a count of the number of times that each action was followed by a payoff, and chooses the action which most commonly has the best payoff, i.e. one reason decision making. Where there is no clear best action, but one or more is evidently worse, a choice must be made as to whether to discard the poorer action; in this case 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 $\sigma$, which maps to $u_w$ - i.e. $s_w$ is a choice between simple lotteries. The algorithm maintains a count, $n$, of the number of occurrences of each outcome given the choice from $s_w$.

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

**Algorithm 1** Lexicographic heuristic

```
n=1, action=none
while action is 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 = best
    end if
    n = n + 1
end while
return action
```

### 3.2.2 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 probability density function takes the form -

$$D(\Theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k}\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k}\Theta_i^{\alpha_{i-1}} \tag{3}$$

where $\alpha = \{\alpha_1 \dots \alpha_k\}$, $k$ is the number of signal-payoff pairs, $\Theta = \{x_1, \dots, x_{k-1}\}$ all more than zero and summing to less than 1, and $\alpha_i$ is the pseudo-count of prior observations for a pair $i$.

The distribution is particularly convenient, in that to infer the probability of a signal implying a payoff becomes simply -

$$p(x = j|D, \alpha) = \frac{\alpha_j + n_j}{\sum_j(\alpha_j + n_j)} \tag{4}$$

Where $n_j$ is simply the count of occurrences of signal-payoff pair $j$, so that 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 with no particular concern as to their order. 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 -

$$R_w(s_w) = \sum_{x \in X} -xp(x|s_w) \tag{5}$$

$$R_m(s_w, s_m) = \sum_{x \in X} -xp(x|s_w \wedge s_m), \tag{6}$$

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

### 3.2.3 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 $s_w$ -

$$R_w(s_w, \theta_w) = \sum_{i \in A_m} \sum_{j \in \Theta} -u_w(s_w, i, \theta_w, j)p(j)p(i|s_w), \tag{7}$$

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 $s_m$ -

$$R_m(s_w, s_m) = \sum_{i \in \Theta} -u_w(s_w, s_m, i)p(i|s_w) \tag{8}$$

### 3.2.4 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. 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 as in Fox and Tversky (1998). In this instance the Bayesian inference process fills the first stage role.

Rather than the psychologically more interesting PT, the CPT decision rule is used in this instance, because of the requirement for women to evaluate the ‘prospects’[5] of more than two actions. 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 (1992), which treats probabilities differently for gains (eqn 9) and losses (eqn 10) -

[5] Where a prospect is a payoff-probability pair, with the set of prospects for an action defining the possible outcomes for it.

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

$$w^{-}(p) = \frac{p^{\delta}}{(p^{\delta}+(1-p)^{\delta})^{\frac{1}{\delta}}}, \tag{10}$$

where $p$ is the unweighted probability, and $\gamma$ and $\delta$ 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 -

$$v(u_i) = \begin{cases} f(u_i), & \text{if}\, u_i > 0 \\ 0, & \text{if}\, u_i = 0 \\ \lambda g(u_i), & \text{if}\, x < 0 \end{cases}, \tag{11}$$

where,

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

$$g(u_i) = \begin{cases} -(-u_i)^{\beta}, & \text{if}\, \beta > 0 \\ -\ln(-u_i), & \text{if}\, \beta = 0 \\ (1-u_i)^{\beta}-1, & \text{if}\, \beta < 0 \end{cases}, \tag{13}$$

and $\alpha$, and $\beta$ are respectively the power of a gain, and a loss, and $\lambda$ is a multiplier giving the aversion to loss. The CPT value of outcome $u_i$, occurring with probability $p$ is $v(u_i)w^{+}(p)$ if $u_i \geq 0$, and $v(u_i)w^{-}(p)$ otherwise. For any given action the CPT value 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.

### 3.3 Information Sharing

It would seem unreasonable to suppose that neither party recounts their experiences to their peers, and to explore the impact of this we also modify the game to introduce a simple form of information sharing within agent groups. This takes the form of having each agent share their memories with 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.[6]

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

# 4 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 (2006), and Phillips et al (2007); as well as a global sensitivity analysis and construction of statistical emulators to explore, and contrast the response surfaces of the four decision rules.

## *4.1 Qualitative Trends*

Throughout this paper, parameters for the CPT model were as used in Tversky and Kahneman (1992) (table 2). While there has been significant work on determining appropriate parametrisation for the model (e.g. Neilson and Stowe (2002); Nilsson et al (2011); Glöckner and Pachur (2012), and particularly Byrnes et al (1999) and Booij et al (2009) addressing risk aversion and gender), a full exploration of the impact of these parameters, or heterogeneous values within populations is beyond the scope of this work. For simplicity, it was assumed that all three drinking types are equally prevalent within the population, although results derived from the Avon Longitudinal Study of Parents and Children (ALSPAC) suggest that the reality is far more positive (Humphriss et al, 2013). The scenario was biased towards disclosure as the better option by presuming a distribution of midwives strongly skewed towards non-judgemental types, with beliefs initially favouring honesty. Payoffs were as in table 1, which ensure that it is always strictly preferable to refer drinkers, and

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

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.

| 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 |
| $\gamma$ | Probability weighting for gains | 0.61 |
| $\delta$ | Probability weighting for losses | 0.69 |
| $\alpha$ | Power for gains | 0.88 |
| $\beta$ | Power for losses | 0.88 |
| $\lambda$ | Loss aversion | 2.25 |

**Table 2** Model parameters.

In addition to assessing the adequacy of the rules in capturing qualitative trends, we also examined the impact of simple information sharing within the population of women (section 3.3) 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.

### 4.2 Global Sensitivity Analysis

In general, we have followed the example of Bijak et al (2013) for global sensitivity analysis of stochastic agent based models, although see Thiele et al (2014) for a review of alternative techniques. For this purpose the Gaussian Emulation Machine for Sensitivity Analysis (GEM-SA) software (Kennedy, 2004) was used, which implements the Bayesian Analysis of Computer Code Outputs (BACCO) method developed by Oakley and O'Hagan (Oakley and O'Hagan, 2002, 2004; Oakley et al, 2006). 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 (R Core Team, 2014) using Latin Hypercube Sampling (Carnell, 2012) over the space of parameters given in table 3, giving eleven free parameters which were treated as uniformly distributed in the range given. Initially a unit hypercube was generated, then the margins transformed appropriately to cover those regions where the inputs are not bounded between 0 and 1, and for proportions of agent types which necessarily sum to one across the three parameters. 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[7] (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 interquartile range 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 four decision models. These emulators were used to conduct a probabilistic sensitivity analysis using GEM-SA to assess the impact of parameters individually, and in combination.

In addition to the sensitivity analysis, we also employed the resulting emulators to rapidly[8] 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 5.3.

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

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

| 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 |

**Table 3** Parameter ranges.

# 5 Results

## *5.1 Qualitative Trends*

As shown in figure 4, all four decision rules were able to reproduce both qualitative trends towards more disclosure as women experience more appointments (Phillips et al, 2007), and a greater tendency towards underreporting of consumption by heavier drinkers (Alvik et al, 2006). Trends for all four rules are broadly similar, exhibiting a gradual increase across appointments which subsequently levels off. This levelling can in part be explained by the referral results (figure 5), 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.

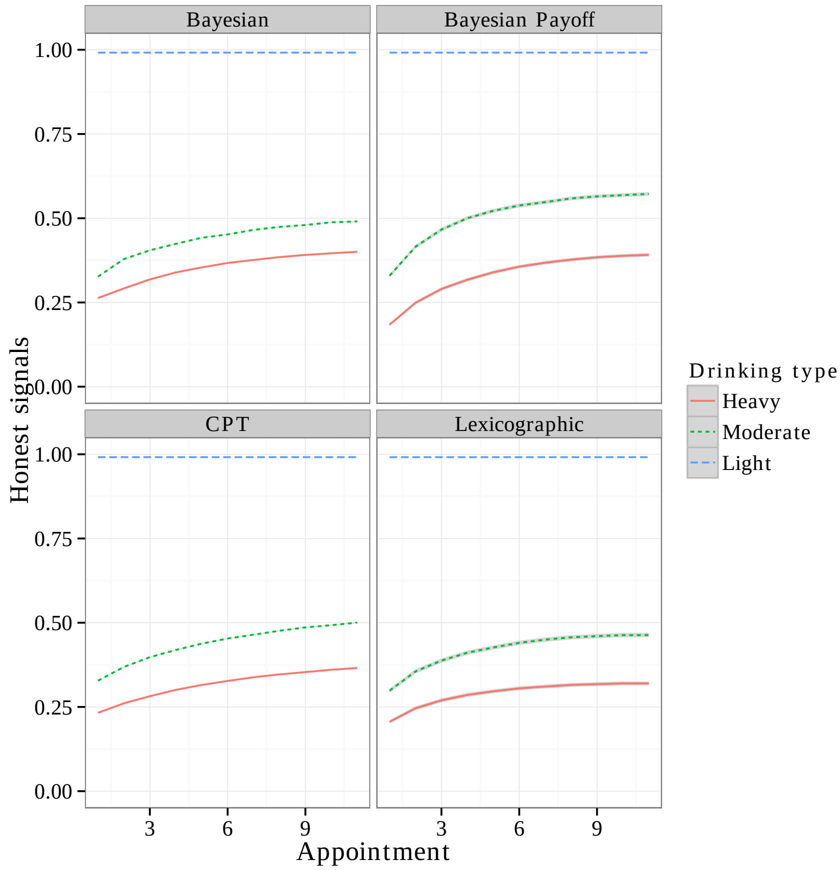



**Fig. 4** Average fraction of population ever signalled honestly after 1000 rounds, mean with 95% confidence limit over 1000 runs

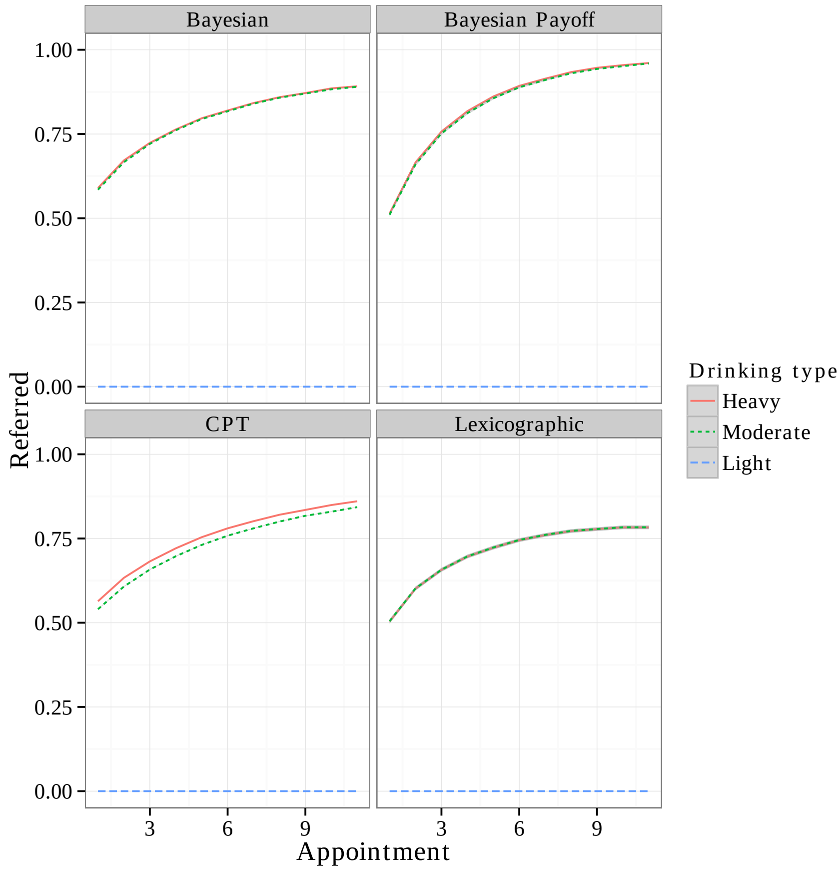



**Fig. 5** Average fraction of population referred after 1000 rounds, mean with 95% confidence limit over 1000 runs

## *5.2 Information Sharing*

Introducing information sharing amongst women leads the behaviour of the decision rules to diverge markedly. Figure 6 shows the proportion of women who have signalled honestly at least once by their final appointment, under four sharing conditions. Aside from the lexicographic decision rule, the general tendency is towards less honest signalling by heavy drinkers, but figure 7 shows a slight increase in referrals for the Bayesian, and CPT rules. For these decision models, this is because information sharing 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 $q_w$ & $w_w$, and the associated increase in false positives. This arises because of the lack of homophily in information sharing, 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 information sharing 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 information sharing, 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 information sharing, for the heuristic style decision rules introduces considerable uncertainty to the model, which is explored further in the sensitivity analysis below.

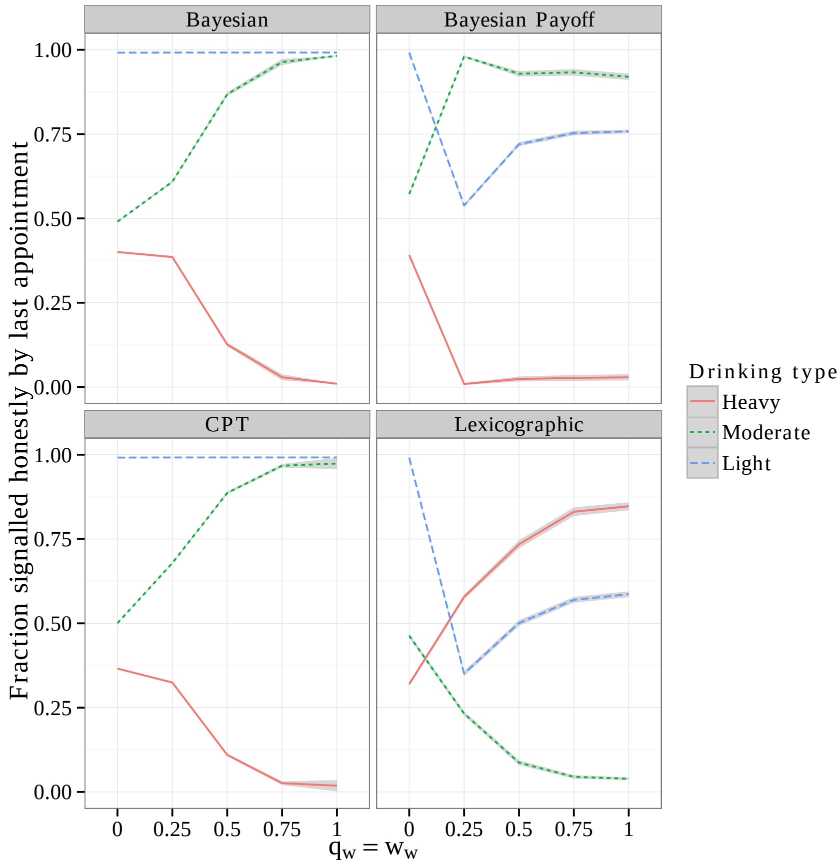



**Fig. 6** Impact of information sharing on trends in honest signalling after 1000 rounds, mean with 95% confidence interval over 100 runs

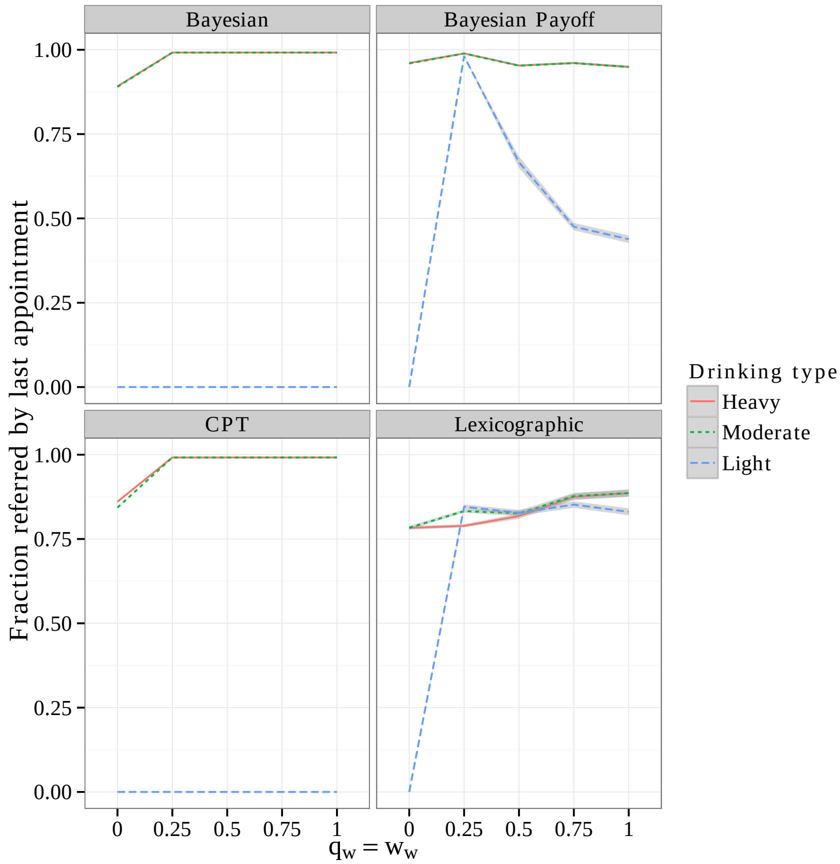



**Fig. 7** Impact of information sharing on referral trends after 1000 rounds, mean with 95% confidence interval over 100 runs

### *5.3 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 information sharing within women. The full results for the sensitivity analysis covering all sixteen emulators are available in appendix 8.

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 information sharing 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 7.

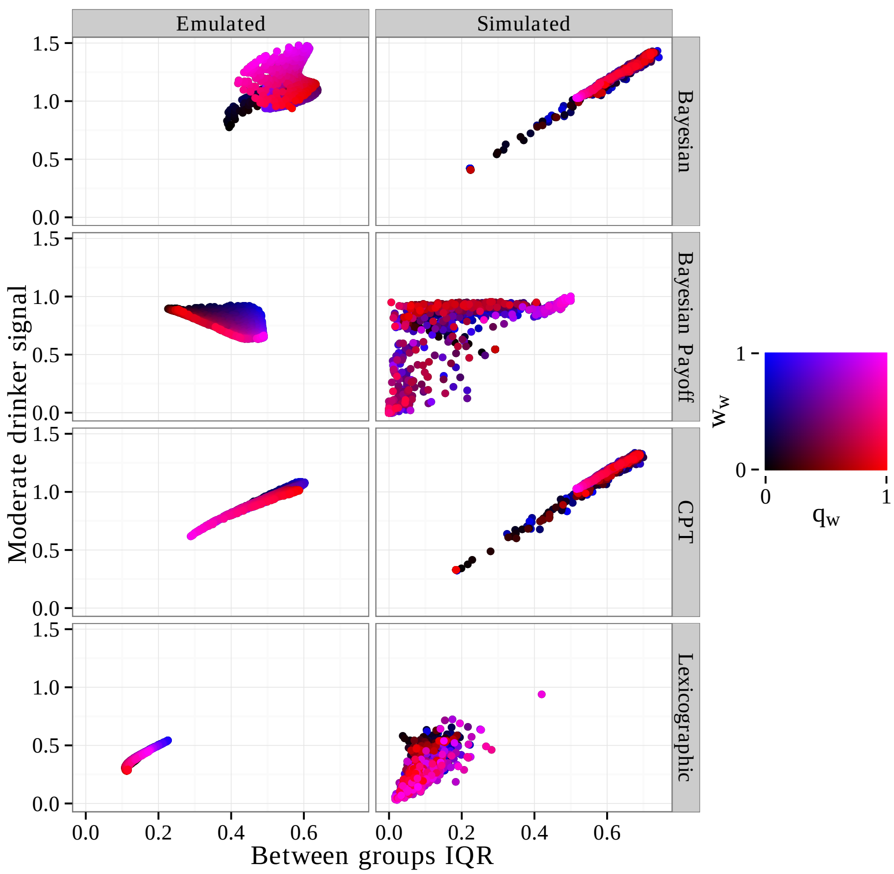



**Fig. 8** Median moderate drinker signal vs median between drinking type IQR for all decision rules

Figure 8 gives a qualitative picture of both emulator quality, and the divergent response surfaces of the decision rules in response to variations in information sharing parameters. Emulator fit is clearly imperfect, but overall behaviour is qualitatively similar, with both emulated and simulated plots demonstrating separation in outcome space for the decision rules.

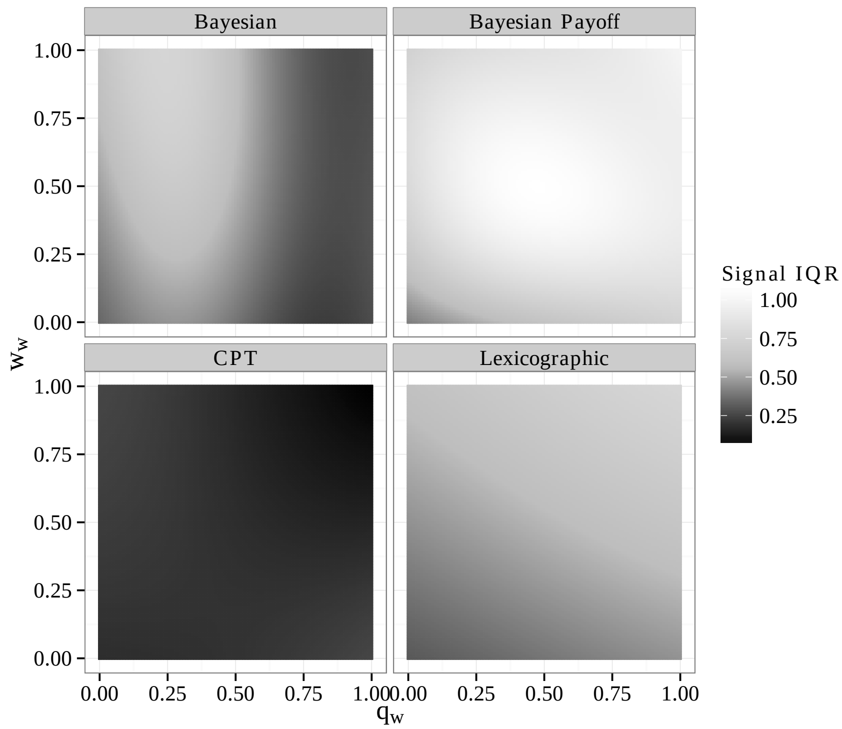



**Fig. 9** Emulated moderate drinker signal IQR in response to varying $q_w$, and $w_w$

Following from the suggestive results for information sharing introducing uncertainty (section 3.3), figure 9 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 that, 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 10 shows the effects, and also highlights the tendency towards poor separability of groups for both the heuristic type decision rules. The response surface of the Bayesian payoff rule is slightly more nuanced than the simple Lexicographic rule. Figure 10 shows better separation, close to partial pooling[9] 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[10] 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 $x_h$.

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

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

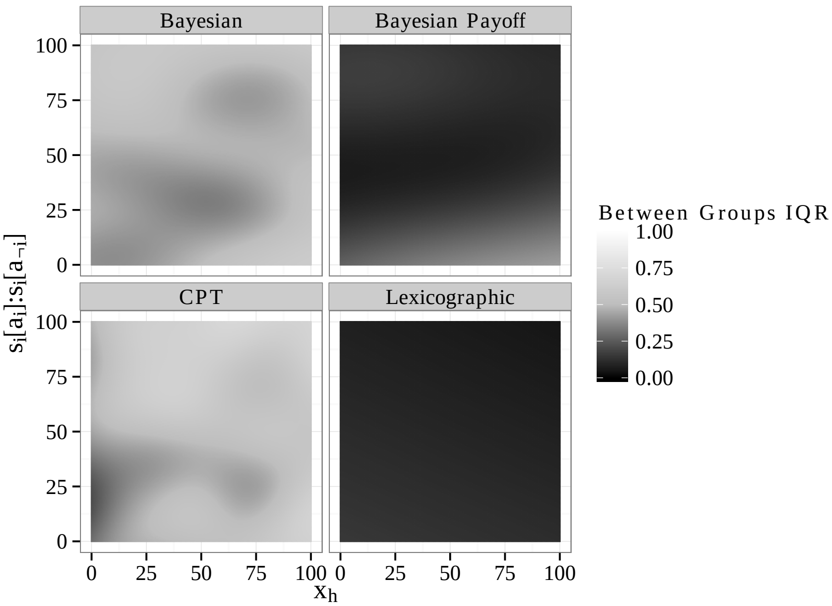



**Fig. 10** Emulated between groups IQR in response to varying $s_i[a_i] : s_i[a_{\neg i}]$, and $x_h$

## 6 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 relatively 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 to a degree. This is an important consideration, given that part of the argument in favour of a decision theoretic approach to agent building 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 (2012) 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 (1981) suggests that incorporating some form of temporal discounting (e.g. exponential (Samuelson, 1937), or hyperbolic (Ainslie, 1991)), or a decision model which explicitly treats intertemporal choice, such as the CPT like model of Loewenstein and Prelec (1992), is warranted.

As noted in section 4.2, the impact of information sharing between 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 so completely, and would pass on concerns about specific women to their colleagues, or that particularly dramatic stories would persist and be passed. This might be addressed by incorporating noisy type information (Feltovich et al, 2002), capturing the unintentional information transmitted during appointments, together with a relaxation of the assumptions about when information may be shared and more sophisticated model of information transmission in general. This also highlights an advantage of the BACCO approach, in that it aids 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.

## 7 Conclusion

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

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/H021698/1 Care Life Cycle, funded within the Complexity Science in the Real World theme. The authors also gratefully acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

## References

Ainslie G (1991) Derivation of “Rational” economic behavior from hyperbolic discount curves. The American Economic Review 81(2):334–340 (Cited on pages 8 and 31.)

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 7.)

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 (Cited on pages 3, 18, 21, and 32.)

Andersen AMN, Andersen PK, Olsen Jr, Grø nbæk M, Strandberg-Larsen K (2012) Moderate alcohol intake during pregnancy and risk of fetal death. International Journal of Epidemiology 41(2):405–13 (Cited on page 3.)

Austen-Smith D, Fryer Jr RG (2005) An economic analysis of "acting white". Quarterly Journal of Economics 120(2):551–583 (Cited on page 5.)

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

Bernoulli D (1954/1738) Exposition of a new theory on the measurement of risk. Econometrica 22(1):23–36 (Cited on page 6.)

Bijak J, Hilton J, Silverman E, Cao VD (2013) Reforging the wedding ring:. Demographic Research 29:729–766 (Cited on page 19.)

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 (Cited on page 18.)

Bowers JS, Davis CJ (2012) Bayesian just-so stories in psychology and neuroscience. Psychological Bulletin 138(3):389–414 (Cited on page 7.)

Brandstätter E, Gigerenzer G, Hertwig R (2006) The priority heuristic: Making choices without trade-offs. Psychological Review 113(2):409–32 (Cited on page 7.)

Burke MS, Carter JR, Gominiak RD, Ohl DF (1996) An experimental note on the allais paradox and monetary incentives. Empirical Economics 21(4):617–632 (Cited on page 7.)

Byrnes JP, Miller DC, Schafer WD (1999) Gender differences in risk taking: A meta-analysis. Psychological Bulletin 125(3):367–383 (Cited on page 18.)

Camerer CF (2000) Prospect theory in the wild: Evidence from the field. In: Kahneman D, Tversky A (eds) Choices, Values, and Frames, Cambridge University Press, pp 288–300 (Cited on page 8.)

Carnell R (2012) LHS: Latin Hypercube Samples. R package version 0.10 (Cited on page 20.)

Christopoulos GI, Tobler PN, Bossaerts P, Dolan RJ, Schultz W (2009) Neural correlates of value, risk, and risk aversion contributing to decision making under risk. The Journal of Neuroscience 29(40):12,574–83 (Cited on page 6.)

Department of Health (2008) Pregnancy and alcohol. Department of Health, London (Cited on page 3.)

Dorazio RM, Johnson FA (2003) Bayesian inference and decision theory-a framework for decision making in natural resource management. Ecological Applications 13(2):556–563 (Cited on page 6.)

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` (Cited on page 4.)

Epstein JM (2014) Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science. Princeton University Press (Cited on page 3.)

Epstein JM, Axtell RL (1994) Growing Artificial Societies: Social Science from the Bottom Up. Complex Adaptive Systems, Brookings Institution Press (Cited on page 3.)

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 6 and 31.)

Fox CR, Tversky A (1998) A belief-based account of decision under uncertainty. Management Science 44(7):879–895 (Cited on page 16.)

Fryer Jr RG, Torelli P (2010) An empirical analysis of 'acting white'. Journal of Public Economics 94(5):380–396 (Cited on page 6.)

Gallistel CR (2012) Extinction from a rationalist perspective. Behavioural Processes 90(1):66–80 (Cited on page 7.)

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 (Cited on page 14.)

Gigerenzer G, Goldstein D (1999) Betting on one good reason: Take the best and its relatives. In: Simple Heuristics that Make Us Smart, Oxford University Press, New York (Cited on page 7.)

Gigerenzer G, Goldstein DG (1996) Reasoning the fast and frugal way: models of bounded rationality. Psychological Review 103(4):650–69 (Cited on pages 6 and 7.)

Gilbert N (1999) Simulation a new way of doing social science. American Behavioral Scientist 42(10):1485–1487 (Cited on page 3.)

Glöckner A, Pachur T (2012) Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. Cognition 123(1):21–32 (Cited on pages 18 and 31.)

Godfray HCJ (1991) Signalling of need by offspring to their parents. Nature 352(6333):328–330 (Cited on page 5.)

Goldstein DG, Gigerenzer G (2002) Models of ecological rationality: The recognition heuristic. Psychological Review 109(1):75–90 (Cited on page 7.)

Gomberg ES (1988) Alcoholic women in treatment: The question of stigma and age. Alcohol and Alcoholism 23(6):507–14 (Cited on page 9.)

Grafen A (1990) Biological signals as handicaps. Journal of Theoretical Biology 144:517–546 (Cited on page 5.)

Gray R, Henderson J (2006) Report to the department of healthreview of the fetal effects of prenatal alcohol exposure. Tech. rep., Oxford: National Perinatal Epidemiology Unit (Cited on page 3.)

Griffiths TL, Chater N, Kemp C, Tenenbaum JB, Perfors A (2010) Probabilistic model of cognition: Exploring representations and inductive biases. Trends in Cognitive Sciences 14(8):357–364 (Cited on page 7.)

Harsanyi JC (1967) Games with incomplete information played by "Bayesian" players, i-iii. Management Science 14(3):159–182 (Cited on page 5.)
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 (Cited on pages 6 and 14.)
Henrich J (2009) The evolution of costly displays, cooperation and religion. Evolution and Human Behavior 30(4):244–260 (Cited on page 5.)
Hertwig R, Barron G, Weber EU, Erev I (2004) Decisions from experience and the effect of rare events in risky choice. Psychological Science 15(8):534–9 (Cited on page 6.)
Huizink AC, Mulder EJH (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 3.)
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) (Cited on page 18.)
Insua DR, Rios J, Banks D (2009) Adversarial risk analysis. Journal of the American Statistical Association 104(486):841–854 (Cited on page 9.)
Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. Econometrica 47:263–291 (Cited on page 7.)
Kahneman D, Tversky A (1984) Choices, values, and frames. American Psychologist 39(4):341–350 (Cited on page 7.)
Kennedy MC (2004) CTCD: GEM software. `http://ctcd.group.shef.ac.uk/gem.html`, accessed: 10/7/2014 (Cited on page 19.)
Kesmodel U, Wisborg K, Olsen SF, Henriksen TB, Secher NJr (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 3.)
Kreps DM, Cho I (1987) Signalling games and stable equilibria. The Quarterly Journal of Economics 102(2):179–221 (Cited on pages 3 and 5.)
Kristensen S (1997) Sensor planning with bayesian decision theory. Robotics and Autonomous Systems 19(3-4):273–286 (Cited on page 6.)
Latino-Martel P, Chan DSM, Druesne-Pecollo N, Barrandon E, Hercberg S, Norat T (2010) Maternal alcohol consumption during pregnancy and risk of childhood leukemia: Systematic review and meta-analysis. Cancer Epidemiology, Biomarkers & Prevention 19(5):1238–60 (Cited on page 3.)
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 8 and 31.)
Macy MW, Willer R (2002) From factors to actors: Computational sociology and agent-based modeling. Annual Review of Sociology 28(1):143–166 (Cited on page 3.)
McNamara J, Houston A (1980) The application of statistical decision theory to animal behaviour. Journal of Theoretical Biology 85(4):673–90 (Cited on page 6.)

Miller RR (2012) An assessment of Gallistel's (2012) rationalistic account of extinction phenomena. Behavioural Processes 90(1):81–3; discussion 87–8 (Cited on page 7.)

National Institute for Health and Care Excellence (2007) Antenatal and Postnatal Mental Health. CG45. NICE, London (Cited on page 8.)

National Institute for Health and Care Excellence (2010a) Antenatal care. CG62. NICE, London (Cited on pages 3, 8, and 13.)

National Institute for Health and Care Excellence (2010b) Pregnancy and complex social factors. CG110. NICE, Manchester (Cited on pages 8 and 9.)

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 18.)

Nilsson Hk, Rieskamp J, Wagenmakers EJ (2011) Hierarchical bayesian parameter estimation for cumulative prospect theory. Journal of Mathematical Psychology 55(1):84–93 (Cited on page 18.)

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 (Cited on page 20.)

Oakley J, O'Hagan A (2002) Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika 89(4):769–784 (Cited on page 20.)

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 (Cited on page 20.)

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 7.)

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441(7090):223–6 (Cited on pages 4 and 6.)

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 (Cited on pages 4 and 6.)

Peterson M (2009) An Introduction to Decision Theory. Cambridge University Press, Cambridge, UK (Cited on page 6.)

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 (Cited on pages 3, 18, 21, and 32.)

Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400(6741):233–8 (Cited on page 6.)

R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (Cited on page 20.)

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

Resnick M (1994) Turtles, Termites and Traffic Jams: Explorations in Massively Parallel Microworlds. The MIT Press, Cambridge, MA (Cited on page 3.)

Robbins H (1964) The empirical bayes approach to statistical decision problems. The Annals of Mathematical Statistics 35(1):1–20 (Cited on page 6.)

Rustichini A (2009) Neuroeconomics: what have we found, and what should we search for. Current Opinion in Neurobiology 19(6):672–7 (Cited on page 4.)

Samuelson PA (1937) Note on measurement of utility. The Review of Economic Studies 4(2):155–161 (Cited on pages 8 and 31.)

Savitz DA (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 3.)

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 (Cited on page 3.)

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 (Cited on page 3.)

Simon HA (1956) Rational choice and the structure of the environment. Psychological Review 63(2):129–138 (Cited on page 7.)

Simon HA (2000) Bounded rationality in social science: Today and tomorrow. Mind & Society 1(1):25–39 (Cited on page 7.)

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 (Cited on page 9.)

Sosis R (2003) Why aren't we all hutterites? Human Nature 14(2):91–127 (Cited on page 5.)

Spence M (1973) Job market signaling. The Quarterly Journal of Economics 87(3):355–374 (Cited on pages 5 and 6.)

Tenenbaum JB, Griffiths TL, Kemp C (2006) Theory-based bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences 10(7):309–18 (Cited on pages 6 and 7.)

Thaler R (1981) Some empirical evidence on dynamic inconsistency. Economics Letters 8(3):201–207 (Cited on pages 8 and 31.)

Thaler R (2000) From homo economicus to homo sapiens. The Journal of Economic Perspectives 14(1):133–141 (Cited on page 8.)

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 (Cited on page 19.)

Todd PM, Dieckmann A (2005) Heuristics for ordering cue search in decision making. In: Saul L, Weiss Y, Bottou L (eds) Advances in Neural Information Processing Systems 17, MIT Press, pp 1393–1400 (Cited on page 7.)

Todd PM, Gigerenzer G (2003) Bounding rationality to the world. Journal of Economic Psychology 24(2):143–165 (Cited on page 7.)

Toll BA, O'Malley SS, Katulak NA (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 8.)

Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. Science 185(4157):1124–31 (Cited on page 7.)

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 7.)

Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5(4):297–323 (Cited on pages 7, 14, 16, and 18.)

Von Neumann J, Morgenstern O (1953) The Theory of Games and Economic Behavior, 3rd edn. Princeton University Press, Princeton (Cited on pages 6 and 7.)

Waldherr A, Wijermans N (2013) Communicating social simulation models to sceptical minds. Journal of Artificial Societies and Social Simulation 16(4):13 (Cited on page 3.)

Watson KK, Platt ML (2012) Social signals in primate orbitofrontal cortex. Current Biology 22(23):2268–73 (Cited on page 6.)

Wildman WJ, Sosis R (2011) Stability of groups with costly beliefs and practices. Journal of Artificial Societies and Social Simulation 14(3) (Cited on page 5.)

Willis ML, Palermo R, Burke D, McGrillen K, Miller L (2010) Orbitofrontal cortex lesions result in abnormal social judgements to emotional faces. Neuropsychologia 48(7):2182–7 (Cited on page 6.)

Zahavi A (1975) Mate selection-a selection for a handicap. Journal of Theoretical Biology 53(1):205–14 (Cited on page 5.)

# 8 Sensitivity Analysis

## *8.1 Median Moderate Drinker Signalling*

| 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 A1** Median moderate drinker signalling parameter sensitivity

| Rule | $\sigma^2$ | Nugget $\sigma^2$ | 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 A2** Median moderate drinker signalling emulator statistics

| 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 A3** Top 5 Interactions terms for CPT 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 A4** Top 5 Interactions terms for Bayesian 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 A5** Top 5 Interactions terms for Lexicographic 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 |

**Table A6** Top 5 Interactions terms for Bayesian Payoff decision rule.

### 8.2 Median Between Groups IQR

| 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 A7** Median between groups IQR parameter sensitivity

| Rule | $\sigma^2$ | Nugget $\sigma^2$ | 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 A8** Median between groups IQR emulator statistics

| 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 A9** Top 5 Interactions terms for CPT 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 A10** Top 5 Interactions terms for Bayesian 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 A11** Top 5 Interactions terms for Lexicographic 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 |

**Table A12** Top 5 Interactions terms for Bayesian Payoff decision rule.

## *8.3 Median Moderate Drinker Signalling IQR*

| 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 A13** IQR of median moderate drinker signalling parameter sensitivity

| Rule | $\sigma^2$ | Nugget $\sigma^2$ | 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 A14** IQR of median between groups IQR emulator statistics

| 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 A15** Top 5 Interactions terms for CPT 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 A16** Top 5 Interactions terms for Bayesian 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 A17** Top 5 Interactions terms for Lexicographic 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 |

**Table A18** Top 5 Interactions terms for Bayesian Payoff decision rule.

## 8.4 Between Groups IQR IQR

| 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 A19** IQR of median between groups IQR parameter sensitivity

| Rule | $\sigma^2$ | Nugget $\sigma^2$ | 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 A20** IQR of median between groups IQR emulator statistics

| 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 A21** Top 5 Interactions terms for CPT 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 A22** Top 5 Interactions terms for Bayesian decision rule.