PHILOSOPHICAL TRANSACTIONS B

Evaluating reproductive decisions as discrete choices under social influence

Journal:	*Philosophical Transactions B*
Manuscript ID	RSTB-2015-0154.R2
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Bentley, R. Alexander; University of Houston, Dept of Comparative Cultural Studies Brock, William A.; University of Missouri, Department of Economics; University of Wisconsin, Department of Economics Caiado, Camila; Durham University, Department of Mathematical Sciences O'Brien, Michael; University of Missouri, Anthropology
Issue Code: Click here to find the code for your issue.:	FERTILITY
Subject:	Behaviour < BIOLOGY, Health and Disease and Epidemiology < BIOLOGY
Keywords:	Bangladesh, discrete choice, fertility, sexual health, contraception, social influence

Evaluating reproductive decisions as discrete choices under social influence

R. Alexander Bentley[1, *], William A. Brock[2, 3], Camila C. S. Caiado[4], and Michael J. O'Brien[5]

Discrete choice, coupled with social influence, plays a significant role in evolutionary studies of human fertility, as investigators explore how and why reproductive decisions are made. We have previously proposed that the relative magnitude of social influence can be compared against the transparency of payoff, also known as the transparency of a decision, through a heuristic diagram that maps decision making along two axes. The horizontal axis represents the degree to which an agent makes a decision individually versus one that is socially influenced, and the vertical axis represents the degree to which there is transparency in the payoffs and risks associated with the decision the agent makes. Having previously parameterised the functions that underlie the diagram, we detail here how our estimation methods can be applied to real-world data sets concerning sexual health and contraception.

[1]Department of Comparative Cultural Studies, University of Houston, Houston TX 77204, USA

[2]Department of Economics, University of Missouri, Columbia, MO 65211, USA

[3]Department of Economics, University of Wisconsin, Madison, WI 53706, USA

[4]Department of Mathematical Sciences, Durham University, Durham DH1 3LE, UK

[5]Department of Anthropology, University of Missouri, Columbia, MO 65211, USA

[*]Author for correspondence (rabentley@central.uh.edu)

Keywords: Bangladesh, contraception, discrete choice, fertility, sexual health, social influence

1. Introduction

Evolutionary theory is becoming more prominent in studies of changing fertility [1], especially those that examine changes in fertility norms over historical time scales [2]. Central to most theories of the demographic transition—a decline in human fertility and mortality in a society [3]—is *individual choice*, often structured with the assumptions of rational-choice theory [4], including the tradeoff between quality and quantity of children [5]. This is broadly consistent with *human behavioural ecology*, in which fertility (and other) decisions are viewed as being motivated by proximate goals, which are ultimately driven by an intrinsic utility function that accounts for energy balance, environmental resources and mortality risk of the socio-ecological environment [6]. This perspective likewise is consistent with a positive correlation between fertility and wealth and a negative correlation between fertility and child-rearing costs [7].

Shenk *et al.* [8] describe three models that apply within this framework, one of which, the *cultural-transmission model*, is our primary focus here. It assumes that people make health decisions based on their knowledge of risks, costs and benefits but also that it matters how they *learn* about these factors. For much of human prehistory, social learning within smaller communities was strategically directed toward familiar experts or best-informed members. In these long-term contexts, the benefits of social learning and friendship are substantial enough to have been a factor in human evolution [9–11]. In a typical collective-action problem, such as changing fertility norms, the transparency about what others are doing is essential [12]. Conversely, secrecy reduces transparency in the sense that it adds 'noise' (e.g. [13]), the effect that we later quantify as μ.

Broadly consistent with these underlying principles, community medicine often emphasises two drivers of behavioural change—information and supply—although in some models (e.g. [14]), those variables are seen as insufficient to change behaviour without the addition of *social influence* [15]. For example, in Poland poor, uneducated women living in a wealthy, educated group tend to adopt the low-fertility level of the group rather than the higher fertility that otherwise would be associated with their low income and low education as individuals [16].

In characterising a case study in terms of information, supply, and social influence, we might use population-scale data to estimate the relative magnitude of socially influenced learning versus individual learning. We might focus on the function of heterogeneity in how people make decisions, particularly the balance of individual learning, which produces information, and social learning, which diffuses and exploits that information [10,17,18]. All too often, however, observational data alone fail to distinguish not only between social learning and individual learning but also between social learning and homophily—the tendency for people with similar traits to co-associate [19,20]. Sorting this out may require explicit temporal ordering of socially influenced events in observational data [21,22].

There is another important dimension in addition to the continuum between individual learning and social learning: the transparency of payoff (and risk) of a particular decision. Transparency, sometimes referred to as the intensity of choice, can have an explicit social dimension in terms of how information spreads—from one to many or from many to many—and the rate at which it spreads. With respect to the successful and decades-long family-planning programme in Matlab, Bangladesh, Phillips *et al.* [23] attributed the success of the 1978 programme to the training of community health workers, in contrast to the less-successful 1975 programme. Simmons *et al.* [24, p. 32], for example, observed a community health worker in a Matlab household easing the fears of a woman concerning a tubectomy by pointing toward a nearby house and saying, "All of them had the operation. I took them for the operation. Did any of them get pregnant? Did you hear it?" This reveals both expertise and direct social influence concerning an important decision for the woman. Because the inherent decision was an intense one—a frightening surgical operation resulting in permanent sterility—the transparency of social influence arguably needed to be high as well.

2. A model of decision making

These principal components motivated us to consider social influence versus transparency of choice as two continuous, intersecting dimensions of decision making (figure 1), one that captures how transparent a decision is in terms of benefits and risks (vertical axis) and the other how socially influenced the decision is (horizontal axis). We have proposed that this modelling framework can be used to extract, from popularity statistics, estimates of social learning and the transparency of decisions at a population scale [25,26]. If we think of the diagram as a map, at its extreme corners are rational choice at the northwest, well-informed social learning at the northeast, indiscriminate copying at the southeast and isolated guesswork at the southwest.

Figure 1

This decision map, while aimed at sociological or anthropological phenomena, has a sound biological basis in the study of social animals. A study of GPS-tracked wild olive baboons in Kenya [27], for example, reduced decision making (primarily about where to go) to two dimensions similar to our map: directional agreement and number of initiators. We might say directional agreement, which is lowest when individuals move in opposing directions and highest when all individuals move together, is analogous to transparency. Number of imitators would then be analogous to social influence in our model, in which the social payoff is proportional to the product of the social-learning parameter and the difference in popularity between the two choices (Appendix). As shown in figure 2, we see a correspondence between Strandburg-Peshkin *et al.'s* [27] *directional agreement* and our *transparency* and between their *number of initiators* and our *strength of social influence*. Both approaches find the highest predictability of following in the northeast corner, where leadership is transparent [28]. Less intuitively, when low on the vertical axis, increasing social influence (moving eastward) may actually decrease the predictability of followers (figure 2). In colloquial terms, if social transparency is low, adding more leaders may lead to the proverbial 'too many cooks spoiling the broth'.

Figure 2

(a) The map quadrants

Let's see how the map works with respect to fertility. In the northwest, the payoffs of having a certain number of children are transparent to a mother, and social influence is minimal. To put it another way, the northwest quadrant is where the net benefits of having *n* children (as *n* varies) are clearly understood, and both local and global sources of social influence have little influence on fertility choices. Costs of giving birth to another child include the physical demands for giving birth, such as adequate body fat, nutrition and energy [29]. This type of cost should be quite transparent to the mother. Also in the northwest are decisions influenced by the net benefits governed by many of the economic realities of giving birth [29]. If children are net assets in terms of labour (agricultural work, for example), then fertility will be high, but as children require increased parental investment (in education, for example), then fertility will be lower. This is why, broadly speaking, fertility declines as wealth increases. Inside this broad trend across populations, however, are subpopulations within which the correlation is positive between wealth and fertility but on a smaller scale. This too is modelled after individual payoffs to fertility, in the simple economic reality that wealth enables a mother to have more children [30].

The northeast quadrant contains fertility outcomes that are maintained by transparent social influence and governed by norms of kinship, religion, gender roles and community expectations. Transparent social influence might come from an acknowledged village expert in a particular category [31], whereas less-intense influence might diffuse among passive members of village social networks, who might not necessarily adopt the behaviour themselves [32]. Among over 20 high-fertility communities in rural Poland, for example, Colleran *et al.* [16] found that the low-fertility norms of educated women were learned socially by lesser-educated women within the same social networks who were using the more-educated women as models.

The two quadrants in the southern half of the map are less well explored in terms of fertility. In wealthy, urban, media-saturated societies, however, these quadrants seem increasingly relevant. A potential mother in the modern West, for example, might read news stories that debate the benefits of breastfeeding versus bottle feeding. She may see vigorous debates in the popular media contesting the 'joys' of having children versus the 'pains' of raising a child—she might, for example, see bumper stickers on expensive cars that say 'DINK' (dual-income, no kids). This bombardment by the media about the costs and benefits of having children might cause enough confusion that transparency is reduced about the net benefits of having *n* children as *n* varies.

Homophily might have an influence on the level of transparency in decision making. Homophily is clearly a reality as parents often socialise with other parents, either through shared membership in schools and clubs or simply so their children can play together. Similarly, nonparents often seek each other out and steer clear of people who have annoying children. All these social effects have ambiguous transparency: Is fertility simply a matter of the fertility of those around you? If so, how accurately does it track ambient fertility? Can fertility be subject to fashion, at least within broader economic constraints (e.g. having one child instead of two)? These are issues that reside in the southeast quadrant.

In the southwest quadrant, fertility is not transparent in terms of individual payoffs or social norms. This is probably uncommon because fertility is so strongly directed by physiological, economic and social influences. The expectations of this quadrant, however, can serve as a benchmark to test whether the individual payoffs of fertility may be becoming less transparent, or, equivalently, 'noisier'. As a thought experiment, what if a mother decides to have just one child in order to invest more in the child's education, but by the time the child is of college age most people are receiving free, online degrees? Could she have had three children instead, to better the chances that free education would make at least one of her children wealthy? The fact that this is not a totally far-fetched scenario in itself shows that calculating the economic costs and benefits of children may be becoming less transparent.

(b) Generalized data patterns

Bentley *et al.* [25,33] describe generalized data patterns that help to locate a phenomenon on the map in figure 1, given time-series data on the relative popularities of different options. The aim is to extract as much information as possible from patterns, such as the distribution of popularity for a single time phase, or ranked popularity lists and/or time series in the popularity of each option over a number of successive time phases. Consider, for example, the effect of a targeted intervention in the last decade by the U.K. National Health Service (NHS). In 2003 the NHS instigated a campaign to promote more sexual-health screenings in an effort to reduce the incidence of sexually transmitted infections (STIs). The programme was successful, increasing the number of screenings from about 17,000 in the first year to over 340,000 in year five [34]. The NHS strategy was to increase information through a dedicated website, advertising and outreach and to direct participants to particular options, including community contraceptive services, general practice and educational settings. In addition, it promoted various private means of performing tests individually, such as using personal testing kits one could purchase at the pharmacy or through the mail. NHS had 24 different types of offerings by 2008 (NCSP 2008).

One question we might ask is how the decision environment of sexual-health screenings (at the regional or national scale) moved around on the map as the health department increased the amount of information, number of options and overall participation. Social influence may or may not have increased with more participation, and decisions may have become (a) more transparent as the NHS communicated cost/benefit information about screenings or (b) less transparent with the increase in options (and associated experts). Ideally, we would like to have popularity data on each option over the course of the programme, at finely resolved time intervals [25,33], such as monthly or quarterly data on the number of STI tests done, for each of the half-dozen options, over a decade and subdivided by gender, county, ethnicity and the like. In reality, the data available in this case include the popularity of the different choices, by gender, for two time periods: one at the start of the programme in 2003 and another for the end of the first phase of the programme in 2007–2008 [34]. Another data set offered by the NHS presents quarterly statistics on the popularity of each option but for only two years (2012 and 2013) and without data on the frequencies of less-popular options, which are subsumed under the category ‘other’.

Consider the two distributions of choice popularity shown in figure 3, one from 2003–2004 and the other from 2007–2008. The number of options listed in the data tables [34] was 10 in 2003–2004 and increased to 14 in 2007–2008. Even with these numbers we can find that the distribution was probably log-normal, which is a type of long-tailed probability-distribution function that we expect in the eastern part of the map, where social influence contributes to the multiplicative growth, or proportional advantage, in the popularity of choices. Here is a simple model that generates a log-normal distribution at each date *t* as well as a limiting distribution that is log-normal with finite mean and variance. Assume the popularity, n_t, of a choice at time *t* is given by the stochastic process $n_t = n_{t-1}^{\rho} e^{g_0 + \sigma e_t}$, $0 < \rho < 1$, where the process $\{e_t\}$ is independently and identically distributed normally with mean zero and variance one. It follows by taking logs that $\ln(n_t)$ conditional on n_{t-1} is normally distributed, with mean g_0 and variance σ^2. The stationary distribution of log(n(t)) is normal, with a finite mean and variance.

Figure 3

So far, this would place the phenomenon in the eastern part of the map. Figure 3 shows that the peak of the log-normal distribution did not change much between 2003–2004 and 2007–2008, except that the distribution became narrower for the decisions of both women and men. This implies that the variance in the rate of multiplicative growth, σ, became smaller. If that variance were correlated with transparency of social learning—a doctor's recommendation carries more weight than an advertisement by a pharmacy—we could hypothesize that transparency actually increased between 2003 and

2008, despite the increase in number of options. Note also that the peak occurs at a higher popularity value among men's choices.

With this hypothesis in mind, we can turn to a second set of information, consisting of time-series data on the popularity of options in 2012 and 2013 [35]. These data have quarterly resolution, but unfortunately, frequencies of the least-popular options are subsumed under 'other'. Nevertheless, we gain an additional, complementary insight by looking at these timelines (figure 4). As we proposed previously [25], generally the farther south on the map one goes, the noisier and less predictable the time series are for the different options, with flat, parallel time series generally indicating high transparency, except when a discovery is made and adopted. From the 2012–2013 U.K. screening data, we can see that the timelines from different regions are generally very smooth and parallel, but with some regions (e.g. Avon, Gloucester and Wiltshire) more so than others (e.g. North East). In the North East, choices were much more volatile than in the other regions—a pattern that needs to be explained through further work. The upper right panel in figure 4 illustrates the effect of the supply of a new health service (a discovery) in Cumbria and Lancashire in late 2013. There, contraception and sexual health services almost overnight became the most popular options.

Figure 4

---

3. Parameterizing the map

In anticipation of future applications that use much larger data sets, we parameterised the vertical axis of the map as b (or b_t if it changes through time), which represents the transparency of decision payoffs, from absolutely transparent along the upper edge ($b_t = \infty$) to completely opaque at the lower edge ($b_t = 0$). The learning dimension—the horizontal axis—is parameterised by J_t, which represents the continuum from decisions made individually at the far left edge ($J_t = 0$) to pure social decision making, or copying, at the far right edge ($J_t = \infty$).

Our approach is derived from the discrete-choice random-utility model,

$$\tilde{U}_k = U_k + Jp_k + \mu\tilde{\varepsilon}_k, k = 1,2,...,N \tag{1}$$

where the random variables $\{\tilde{\varepsilon}_k, k = 1,2,...,N\}$ are independently and identically distributed with an extreme-value distribution [36]. Here, the parameter μ in equation (1) scales the amount of noise in choice making. If $\mu = 0$, the decision maker just chooses with probability one the choice k^*, where $\tilde{U}_k = U_k + Jp_k$ is the largest. Our intensity-of-choice parameter, b, is proportional to $1/\mu$. If $b = 0$, i.e. $\mu = \infty$, each choice k is made with probability $1/N$. We define an increase in transparency as an increase in the parameter b. Any context in which b increases, i.e. reduces μ, is an increase in transparency for us. There are many different reasons why μ decreases (or increases) when a group of decision makers are making a choice over N alternatives. If transparency becomes very small, then μ becomes very large, and the contribution of the intrinsic utilities, U_k, of the different options becomes negligible. For example, increasing easily available information about the benefits and costs of different contraceptive methods should increase b, i.e. decrease μ. At the same time, if transparency is small, a large degree of social influence, J, and/or conformity of practice (high p_k for the consensus-choice k) could offset the effects of the noise component, $\mu\tilde{\varepsilon}_k$. This gives the model scope to account for cases where human social learning occurs over many generations with no explicit understanding of problems (e.g. [37]).

In resolving both axes together, the map calls for traditional and novel forms of time-series analysis of the form and temporal dynamics of popularity distributions. In the Appendix we describe the details of the multinomial logit framework for the general case of N choices, in which $k = 1, 2, \ldots N$ indexes the N different choices and U_k is the inherent utility of choice k. We have defined the map space analytically as functions of observable co-variates and estimated parameters [26,36]. In expressing probability, P_k, of choice k (vs. choice probability, P_1), we parameterise b and J in terms of observable variables:

$$\ln\frac{P_{igt}(k)}{P_{igt}(1)} = \underbrace{b(\theta, z_{igt})}_{\text{transparency}}[\underbrace{\varphi_1(x_{ikgt} - x_{i1gt})}_{\text{individual}} + \underbrace{J(\varphi_2, y_{igt})(P_{tkg} - P_{t1g})}_{\text{social}}] \qquad (2)$$

Here, i, g, t denote a person or subgroup of group g at date t. The quantity $\varphi_1(x_{ikgt} - x_{i1gt})$ represents the difference in the personal component of payoff of the individual choice k over choice 1, excluding the social component of payoff, which is the second term in (2).

The key to applying this framework to a real-world case study is the proposal that the covariate vectors x, y, z represent observable quantities that can be constructed from data [36]. If we index the covariates by agent i, group g and time t, then the covariates for each agent can include those that predict the propensity of the behavior, denoted by x_{itg}, those associated with the presence of social influence, y_{itg} , and those that relate how variable the choices were through time, z_{itg}. Parameter vectors θ, φ_1, and φ_2 operate on covariates denoted by vectors *x, y* and *z*, respectively. Parameter vector φ_1 represents an individual's sensitivity to differences in choice, acting on the payoff difference between options, and parameter vector φ_2 represents the transparency of social influence, which acts on the popularity of the option. Estimating the parameter vector φ_1 determines the individual sensitivity to differences in choice $(x_k - x_1)$.

In studies of fertility choice, as discussed above, the social-choice transparency, φ_2, could reflect the tendency to have the same number of children as other mothers, whose success and/or education has become more socially visible. In this formulation of equation (2), social learning is part of the payoff, which corresponds to the fitness of the chosen behaviour. On a fitness landscape, a group might be on a local peak but not necessarily the globally optimal peak among multiple equilibria. In our parameterised map above, there can be a unique set of optima for each pair of *b* and *J*, each number of options *N* and each set of intrinsic utility values *U*. For a binary decision, such as whether or not to have a child, we would define $N = 2$, but for other decisions, such as which form of contraception to use (potentially including none), we would have $N > 2$ (e.g. pill, IUD, condom, surgery and so on). Because the social part of the fitness function depends on the relative popularity of each choice taken, the fitness landscape becomes more rugged moving to the right (figure 1), as the sensitivity to social influence, *J*, increases. We describe this in more detail in the Appendix. It actually becomes difficult to define the fitness landscape for multiple different options, even for $N > 3$

options [28].

4. Methods of estimating coordinates on *b* and *J*

In order to estimate coordinates on the dimensions b and J, we can use nonlinear least squares and the parameterisation above through a log-odds regression [26]. As described in the Appendix, we can estimate $b_{itg} = b(\theta, z_{itg})$ as well as the other parameters of the choice dynamics once we have adequate time-series data on a vector of covariates, x_{itg}, and have parameterised the transparency-of-choice function, $b_{itg} = b(\theta, z_{itg})$. Estimating the parameter vector φ_2 empirically determines the social-influence function, $J(.)$. Brock *et al.* [26] generated artificial data to see how these estimates can be used to describe a path through the map for each agent i in each group g for which we have data at date t. For example, data for a binary choice ($N = 2$) might have $\varphi_2 = 1$, $\varphi_1 = 1$ and $\theta = 5$, with group size $G = 100$ and $M = 100$ agents per group. In figure 5, the red dots are simulated data, and the blue dots are our estimates. We can see that the largest discrepancies are in the vertical dimension, which is why we aim to improve the estimation of transparency of choice.

Figure 5

Having demonstrated its potential for simulated data, our objective is to apply the diagram to population-scale data on fertility choices and/or contraception use, sampled frequently (e.g. monthly) over a period of many years. A good data set to which we could apply our framework is the 33-year record (1977–2010) of contraceptive choices in Matlab, an exceptional long-term data set on about 250,000 people, comprising information collected since 1977 in a region where many health and family-planning interventions have been made [8,24,38]. The available data are records of decisions, together with associated (anonymous) details of the individuals making those decisions, including total fertility, surviving children, age at marriage, household income, education and media exposure. Additionally, Shenk *et al.* [8] surveyed 944 women from two areas, in three adult age categories, with recorded variables similar to those above.

The Matlab data were recorded fortnightly, so T in this case would be on the order of 850. There were 80 different catchments of community health workers, so $G = 80$. We can approximate the number of people per village, M, at about 1200 initially in 1977, when the Matlab area included 180,000 people in 149 villages [39]. Among the five main options for contraception, we might initially rank the inherent payoff, x, hypothetically as tubectomy < injection < IUD < pill < condom. In other words, tubectomy has a higher cost/benefit ratio than asking a partner to use a condom. In terms of social visibility, y, the ranking could be considerably different and probably more variable through time, depending on region-wide conversations that take place and even potentially on individual preferences of the individual community health workers.

What we need to determine from the data for the example of having k children (for different $k > 0$) compared to none are (a) the covariates x that influence the individual component of the difference in payoff from having k children to having no children, (b) the covariates y that influence the social component of payoff from having k children to having none and (c) the covariates *z* that influence the function, *b*(.), which governs the precision of choice between *k* children and none. These covariates might be derived from statistics such as total fertility, religion, surviving children, age at marriage, household income or education, as in the data used by Shenk *et al.* [8].

Data sets such as the Matlab data may be rich enough to estimate our model framework. The vector of parameters representing a person's sensitivity to inherent payoff, φ_1 in the term $\varphi_1(x_{ikgt} - x_{i1gt})$, could potentially be estimated using covariates such as a wife's education, a husband's primary occupation, family land ownership and whether a family is engaged in agriculture [8]. In this case, φ_1 would be a four-dimensional vector with one parameter to be estimated for each covariate. In order to attempt to separate social influences from the endogenous social effects that have a social multiplier [40], we could add a term $\psi_1(\bar{x}_{kgt} - \bar{x}_{1gt})$, where $\bar{x}_{kgt}$ denotes the averages of relevant covariates over the relevant reference group, *g*.

If the *bari*—a patrilineal group often consisting of 2–6 houses arranged around a courtyard [41], though frequently larger in more densely settled villages—is too small for use in calculating an average measure, the next largest reference group, for example, the village, would be the best choice. We expect the strength of social interactions to be strongest at the bari level and somewhat weaker at the village level. Based on the work of Shenk *et al.* [8], child deaths in each bari or village, infant mortality rate and the number of women in each health worker's intervention area appear to be reasonable candidates for contextual variables.

The vector of parameters φ_2 might be estimated for well-chosen specifications of the function $J(\varphi_2, y_{igt})$ for a vector of covariates such as average number of children born to women in the same reference group as person or subgroup i of that reference group. As already mentioned, however, it is difficult to prove social influence from the data, as opposed to contextual effects [40] and/or selection biases sorting unobservable covariates that influence fertility. It usually is the convention in empirical work using discrete-choice models [42] to normalize the intensity-of-choice function $b(.)$ to a constant, i.e. unity. Brock *et al.* [26] discuss how one might estimate $b(.)$ up to scale by using data over time, but they also show through simulation that it is more difficult to estimate $b(.)$ with as much precision as other parameters such as $\varphi_1, \psi_1, \varphi_2$.

One possible approach to estimating the intensity-of-choice function is to normalize $b(.)$ to unity for a first time segment of the data (t = 1) and to estimate the parameters $\varphi_1, \psi_1, \varphi_2$ for that time segment. Then, estimate those same three parameters of b for each subsequent segment. If the estimates scale approximately proportional to those parameters at t = 1, then we have a means of estimating change in $b(.)$ over time as a simple function of those parameters. If there are approximate scale factors $b_{[t_1+1,t_2]}, b_{[t_2+1,t_3]}, \ldots$, they could serve as estimates of the function $b(.)$. Even with ideal data available, attempting to extract approximate proportional scale factors may often be the best way to estimate $b(.)$ [26].

5. Conclusions

What contribution does our model bring to the table? Other models (e.g. [43]) allow for errors in payoff estimation—our transparency dimension—but usually with no social influence. A few models, although they consider social influence (e.g. [44]), do not explicitly consider the econometric problem of identifying and estimating the strength of endogenous social interactions [36,45,46]. In addition, they do not employ elements of discrete-choice econometrics [47], nor, with two exceptions [26,42], do they estimate the intensity-of-choice (transparency) function, $b(.)$, using observational data.

Our contribution here is combining the tools for these estimations and suggesting how to apply the tools to fertility-related choices. The toolbox is valuable because endogenous social interactions possess a 'social multiplier', which can be exploited to promote desired target behaviour by policy (e.g. a lowering of fertility or increased adoption of pit toilets, birth-control methods and the like). Estimating these models without correction for group-selection effects, correlated effects, correlated unobservables and general contextual effects easily leads to spurious findings of endogenous social interactions [36,40,46]. This problem occurs in the presence of many other social influences that do not possess a social multiplier [40,45].

Our framework was designed specifically to estimate social influence and transparency of choice from aggregated, anonymous decision data, even when we lack survey data about the motivations of individual decision makers. The next step will be to apply our parameterised model to real data, such as the contraception and fertility data from Matlab, potentially to map a trajectory in terms of the parameters b_t and J_t and determine how they change through time. Recall that each time-dependent variable is a vector, or list, not just a value. The variable x is a list of covariates that determine differences in payoff to different choices; the variable y is a list of covariates that one thinks will play a role in payoffs to social visibility; and the variable z is a list of covariates that one thinks will play a role in the precision of choice making over time.

In contrast to ancient contexts, major challenges today, such as global health, inequality and environmental change, are not transparent to many people either socially or factually. Often people make decisions based on emotions or hearsay. If our null hypothesis is to assume individuals make decisions with clarity of payoffs interpreted through an evolved psychological utility function, then departures from that null should include decisions made without clarity of payoff or through social influence, which may underlie substantial intercommunity variation in social norms and perceived costs and benefits [48]. Classic studies of social conformity [49], as well as more recent experiments [50,51], confirm that social interaction can substantially undermine the independence of the judgments of participants. Conversely, although individual 'rational' learners may be in the minority, certain studies indicate that if their decisions are consistent or intense, they will guide a majority of social learners to a common goal [52–54].

As decisions are increasingly mediated by online technologies [55], our framework may help with questions about change in the magnitude of social influence and the transparency (intensity) of its effect. In an age deluged by information, options and influences, transparency of social learning becomes a vital issue. More research is needed to characterize the strength of potential causal pathways over time and to document the potential importance of social influences on fertility behaviour. Endogenous social interactions are of special interest because the potential social-multiplier effect can be exploited by policy to induce a ripple effect on fertility behaviour across communities. We believe the map shown in figure 1 is a useful organizing framework once it is operationalised within a precise statistical framework.

Authors' contributions. R.A.B., W.A.B., and C.C.S.C performed the analysis. R.A.B., W.A.B., C.C.S.C. and M.J.O wrote the article.

Competing interests. We have no competing interests.

Acknowledgements. We are grateful to Mary K. Shenk and two anonymous reviewers for helpful comments on earlier versions of the manuscript.

References

1. Sear R. 2015 Evolutionary contributions to the study of human fertility. *Pop. Stud.* **69**, S39–S55.

2. Milot E, Mayer FM, Nussey DH, Boisvert M, Pelletier F, Réale D. 2011 Evidence for evolution in response to natural selection in a contemporary human population. *Proc. Nat. Acad. Sci.* **108**, 17040–17045.

3. Chesnais J-C. 1993 *The demographic transition: stages, patterns, and economic implications: a longitudinal study of sixty-seven countries covering the period 1720–1984.* Oxford: Oxford University Press.

4. Johnson-Hanks J. 2008 Demographic transitions and modernity. *Ann. Rev. Anthropol.* **37**, 301–315.

5. Doepke, M. 2015 Gary Becker on the quantity and quality of children. *J. Demog. Econ.* **81**, 59–66.

6. Gibson, MA, Lawson DW. 2015 Applying evolutionary anthropology. *Evol. Anthropol.* **24**, 3–14.

7. Nettle D, Gibson MA, Lawson DW, Sear R. 2013 Human behavioral ecology: current research and future prospects. *Behav. Ecol.* **24**, 1031–1040.

8. Shenk MK, Towner MC, Kress HC, Alam N. 2013 A model comparison approach shows stronger support for economic models of fertility decline. *Proc. Nat. Acad. Sci.* **110**, 8045–8050.

9. Christakis NA, Fowler JH. 2014 Friendship and natural selection. *Proc. Nat. Acad. Sci.* **111**, 10796–10801.

10. Hoppitt W, Laland KN. 2013 *Social learning: an introduction to mechanisms, methods, and models.* Princeton, NJ: Princeton University Press.

11. Hruschka DJ. 2010 *Friendship: development, ecology, and evolution of a relationship.* Berkeley: University of California Press.

12. Hauser OP, Rand DG, Peysakhovich A, Nowak MA. 2014. Cooperating with the future. *Nature* **511**, 220–223.

13. Li J, Lee L. 2009 Binary choice under social interactions: an empirical study with and without subjective data on expectations. *Applied Econom.* **24**, 257–281.

14. Christakis NA, Fowler JH. 2013 Social contagion theory: examining dynamic social networks and human behavior. *Stat. Med.* **32**, 556–577.

15. Guiteras R, Levinsohn J, Mobarak AM. 2015 Encouraging sanitation investment in the

developing world: a cluster-randomized trial. *Science* in press (doi: 10.1126/science.aaa0491).

16. Colleran H, Jasienska G, Galbarczyk A, Mace R. 2014 Community-level education accelerates the cultural evolution of fertility decline. *Proc. Biol. Sci.* **281**, 20132732.

17. Mesoudi A, Whiten A. 2008 The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Phil. Trans. R. Soc. B* **363**, 3489–3501.

18. Rendell L, Boyd R, Enquist M, Feldman MW, Fogarty L, Laland KN. 2011 How copying affects the amount, evenness and persistence of cultural knowledge: insights from the Social Learning Strategies Tournament. *Phil. Trans. R. Soc. B* **366**, 1118–1128.

19. Shalizi CRS, Thomas AC. 2010 Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* **40**, 211–239.

20. Thomas AC. 2013 The social contagion hypothesis: comment on 'Social contagion theory: examining dynamic social networks and human behavior'. *Stat. Med.* **32**, 581–590.

21. Aral S, Muchnik L, Sundararajan A. 2009 Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Nat. Acad. Sci.* **106**, 21544–21549.

22. Hobaiter C, Poisot T, Zuberbühler K, Hoppit W, Gruber T. 2014 Social network analysis shows direct evidence for social transmission of tool use in wild

chimpanzees. *PLOS ONE* **12**(9), e1001960.

23. Phillips JF, Simmons R, Koenig MA, Chakraborty J. 1988 The determinants of reproductive change in a traditional society: evidence from Matlab, Bangladesh. *Stud. Fam. Plann.* **19**, 313–334.

24. Simmons R, Baqee L, Koenig MA, Phillips JF. 1988 Beyond supply: the importance of female family planning workers in rural Bangladesh. *Stud. Fam. Plann.* **19**, 29–38.

25. Bentley RA, O'Brien MJ, Brock WA. 2014 Mapping collective behavior in the big-data era. *Behav. Brain Sci.* **37**, 63–119.

26. Brock WA, Bentley RA, O'Brien MJ, Caiado CCS. 2014 Estimating a path through a map of decision making. *PLoS ONE* **9**(11), e111022.

27. Strandburg-Peshkin A, Farine DR, Couzin ID, Crofoot MC. 2015 Shared decision-making drives collective movement in wild baboons. *Science* **348**, 1358–1361.

28. Caiado CCS, Brock WA, Bentley RA, O'Brien MJ. 2016 Fitness landscapes among many options under social influence. *J. Theor. Biol.* __, ___–___.

29. Low BS. 2005. Women's lives there, here, then, now: a review of women's ecological and demographic constraints cross-culturally. *Evol. Hum. Behav*. **26**, 64–87.

30. Mace R. 2008 Reproducing in cities. *Science* **319**, 764–766.

31. Henrich J, Broesch J. 2011 On the nature of cultural transmission networks: evidence

from Fijian villages for adaptive learning biases. *Phil. Trans. R. Soc. B* **366**, 1139–1148.

32. Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO. 2013 The diffusion of microfinance. *Science* **341**, 1236498.

33. Bentley RA, O'Brien MJ, Ormerod P. 2011 Quality versus mere popularity: a conceptual map for understanding human behavior. *Mind Soc.* **10**, 181–191.

34. National Chlamydia Screening Programme. 2008 *Five years: the fifth annual report of the National Chlamydia Screening Programme 2007/08*. London. www.chlamydiascreening.nhs.uk/ps/resources/annual-reports/NCSPa-rprt-07_08.pdf.

35. Chlamydia Testing Activity Dataset 2015. Public Health England data on the National Chlamydia Screening Programme, available at www.chlamydiascreening.nhs.uk/ps/data.asp.

36. Brock WA, Durlauf SN. 2001 Interactions-based models. In *Handbook of econometrics* (ed JJ Heckman, E Leamer), pp. 3297–3380. Dortrecht, Netherlands: Elsevier.

37. Boyd R, Richerson PJ, Henrich J. 2011 The cultural niche: why social learning is essential for human adaptation. *Proc. Nat. Acad. Sci.* **108**, 10918–10925.

38. International Centre for Diarrhoeal Disease Research, Bangladesh. 2012 *Health and Demographic Surveillance System—Matlab, Vol. 44, Registration of Health and Demographic Events 2010*. Dhaka, Bangladesh: Centre for Population, Urbanization

and Climate Change.

39. Gribble J, Voss M-L. 2009 *Family Planning and Economic Well-being: New Evidence from Bangladesh*. Policy Brief, Population Reference Bureau, May 2009. Washington DC. www.prb.org/pdf09/fp-econ-bangladesh.pdf.

40. Manski CF. 1993 Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.* **60**, 531–542.

41. Shrestha NR. 2002 *Nepal and Bangladesh: a global studies handbook.* Santa Barbara, CA: ABC–CLIO.

42. Goldbaum D, Mizrach B. 2008 Estimating the intensity of choice in a dynamic mutual fund allocation decision. *J. Econ Dyn. Control* **32**, 3866–3876.

43. McKelvey RD, Palfrey TA. 1998. Quantal response equilibria for extensive form games. Exper. Econ. **1**, 9–41.

44. Choi S, Gale D, Kariv S. 2012 Social learning in networks: a quantal response equilibrium analysis of experimental data. *Rev. Econ. Design,* **16**, 135–157.

45. Blume LE, Brock WA, Durlauf, SN, Jayaraman R. 2015 Linear social interactions models. *J. Pol. Econ. 123*, 444–496.

46. Brock WA, Durlauf SN. 2007 Identification of binary choice models with social interactions. *J. Econometrics 140*, 52–75.

47. McFadden D. 1984 Econometric analysis of qualitative response models. *In Handbook of econometrics: volume II* (ed Z Griliches, M Intriligator), pp. 1395–

1497. Amsterdam: North-Holland.

48. Miller S, Yardley L, Little P, PRIMIT team. 2012 Development of an intervention to reduce transmission of respiratory infections and pandemic flu: measuring and predicting hand-washing intentions. *Psychol. Health Med.* **17**, 59–81.

49. Asch SE. 1955 Opinions and social pressure. *Sci. Am.* **193**, 31–35.

50. Lorenz J, Rauhut H, Schweitzer F, Helbing D. 2011 How social influence can undermine the wisdom of crowd effect. *Proc. Nat. Acad. Sci.* **108**, 9020–9025.

51. Salganik MJ, Dodds PS, Watts DJ. 2006 Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856.

52. Couzin ID, Ioannou CC, Demirel G, Gross T, Torney CJ, Hartnett A, Conradt L, Levin SA, Leonard NE. 2011 Uninformed individuals promote democratic consensus in animal groups. *Science* **334**, 1578–1580.

53. Dyer JRG, Johansson A, Helbing D, Couzin ID, Krause J. 2009 Leadership, consensus decision making and collective behaviour in human crowds. *Phil. Trans. R. Soc. B* **364**, 781–789.

54. Kurvers RHJM, Krause J, Croft DP, Wilson ADM, Wolf M. 2014 The evolutionary and ecological consequences of animal social networks: emerging issues. *Trends Ecol. Evol.* **29**, 326–335.

55. Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, Fowler JH. 2012 A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298.

Figure Captions:

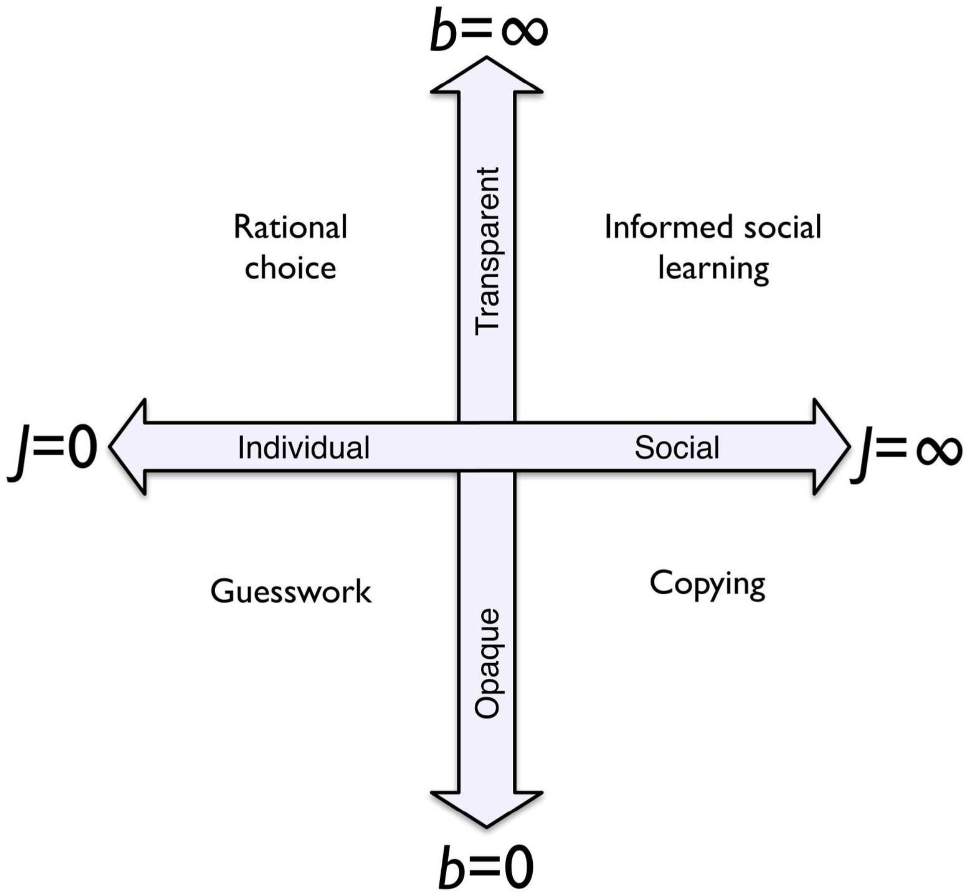
Figure 1. A diagram depicting domains of human decision making, based on a continuum from individual learning to social learning on the horizontal axis (J) and the transparency of payoffs informing a decision (intensity of choice) on the vertical axis (b). After [25].

Figure 2. A comparison between a plot for understanding the group movement among social animals and the outcome of the decision heuristic in figure 1. The figure on the left refers to field data among wild baboons [27] showing *directional agreement* among initiators of movement on the vertical axis and the *number of initiators* on the horizontal axis. On the right is a contour plot in the framework of figure 1, showing the maximum fitness for three choices when that fitness is defined as a function of both the inherent utility of the choice and its social popularity [28]. The plot on the right shows the maximum fitness at each coordinate because multiple equilibria may be possible at each point.

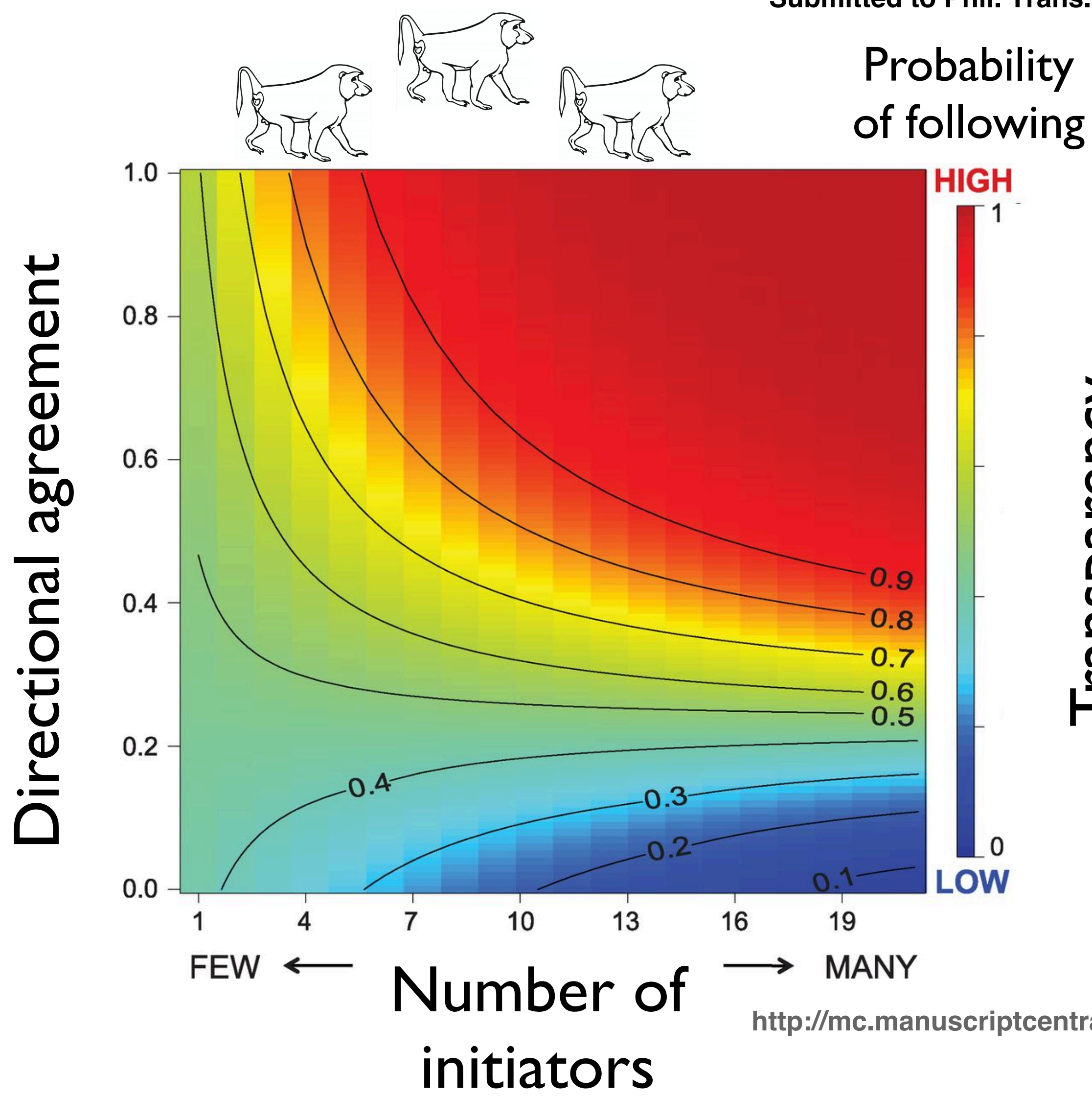
Figure 3. Distributions of choice popularity for 2003–2004 (left) and 2007–2008 (right). Data shown for females in red and males in blue. The number of options listed in the data tables [34] was 10 in 2003–2004 and increased to 14 in 2007–2008. For each distribution the line shows the maximum likelihood estimate for a log-normal function described by log-normal mean, g_0, and standard deviation, σ, as follows: for 2003–2004, (g_0, σ) = (5.82, 2.29) for women and (4.01, 1.85) for men; for 2007–2008, (g_0, σ) = (9.03, 1.33) for women and (8.10, 1.74) for men.

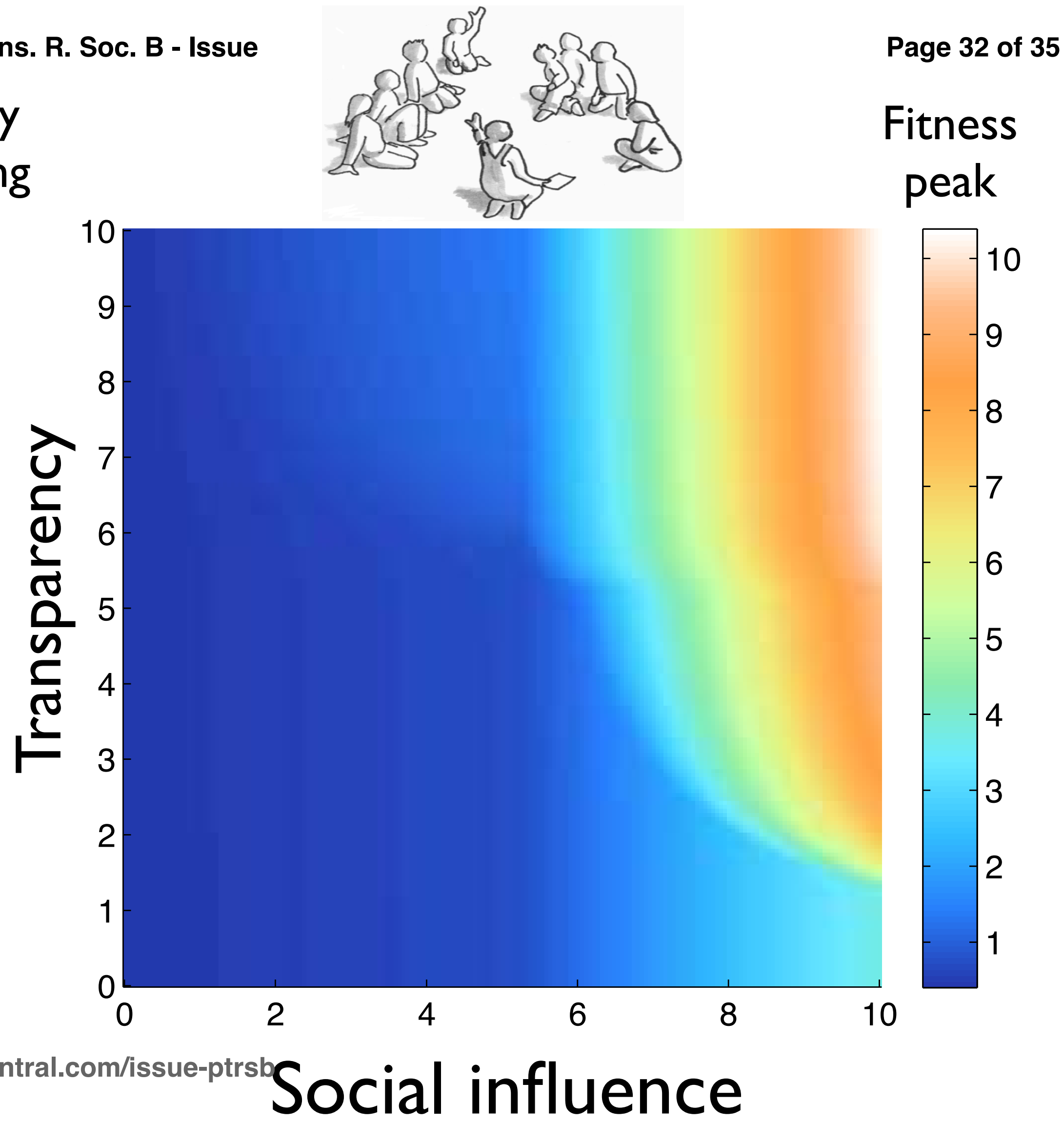
Figure 4. Representative timelines for the popularity of different STI screening options in the United Kingdom for each quarter of the years 2012 and 2013 [35]. The publicly available data are specifically the numbers of chlamydia tests for 15–24 year olds, grouped by testing-service type and by region (designated by "Upper Local Authority" in terms of the divisions of the U.K. National Health Service). Abbreviations: GUM—genito-urinary medicine; GP—general practitioner (local doctor); CSHS—contraception and sexual health services; TOP—termination of pregnancy.

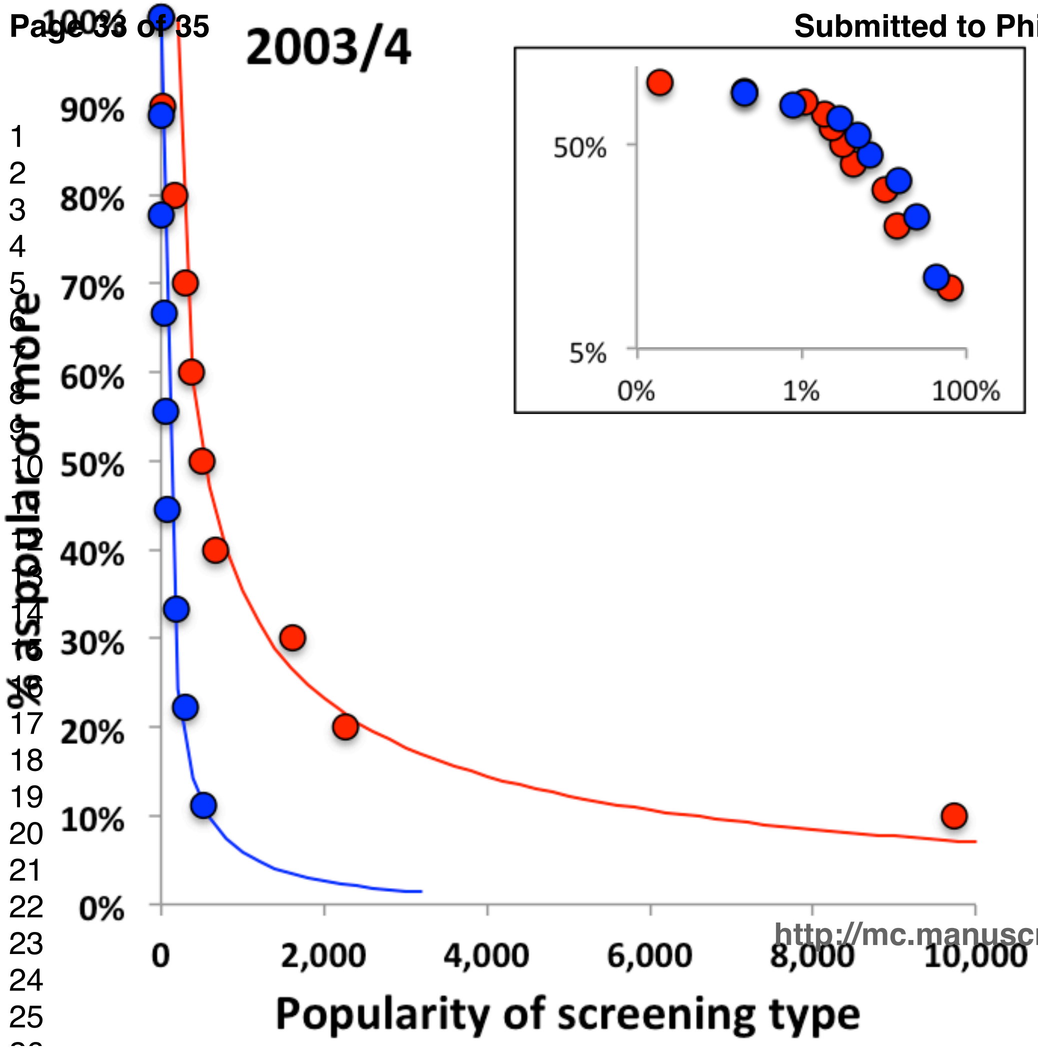
Figure 5. Diagram comparing nonlinear least-squares estimates (blue dots) against simulations (red dots) of the model defined by equations A4 and A5 in the Appendix. After [24]. The simulations (red dots) used two choices, and the axes are measured in units of the parameter values (from equation A5) for choice intensity, θ, on the vertical axis and social-influence intensity, φ_2, on the horizontal axis.

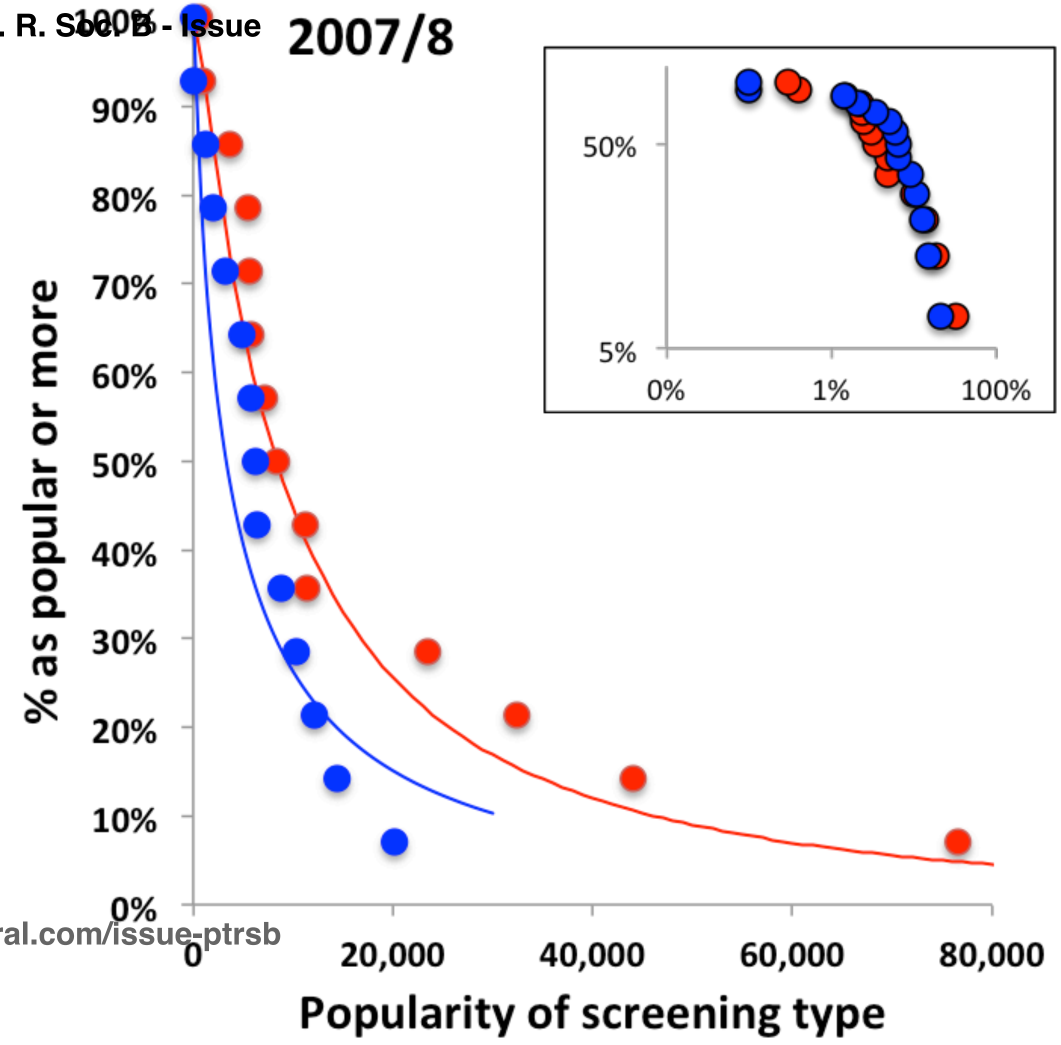


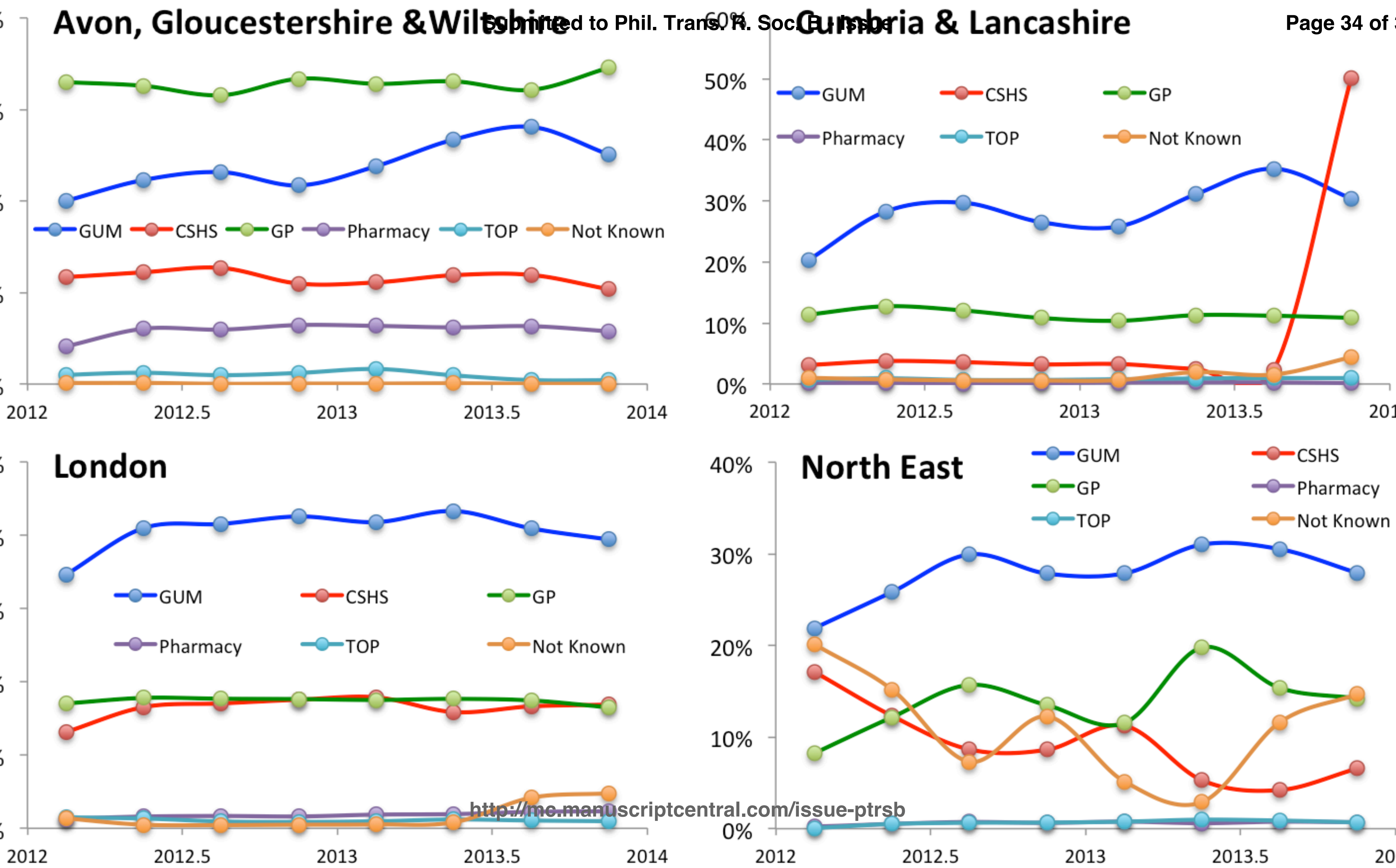
A diagram depicting domains of human decision making, based on a continuum from individual learning to social learning on the horizontal axis (J) and the transparency of payoffs informing a decision (transparency of choice) on the vertical axis (b) (after Bentley et al. 2014).
187x180mm (300 x 300 DPI)

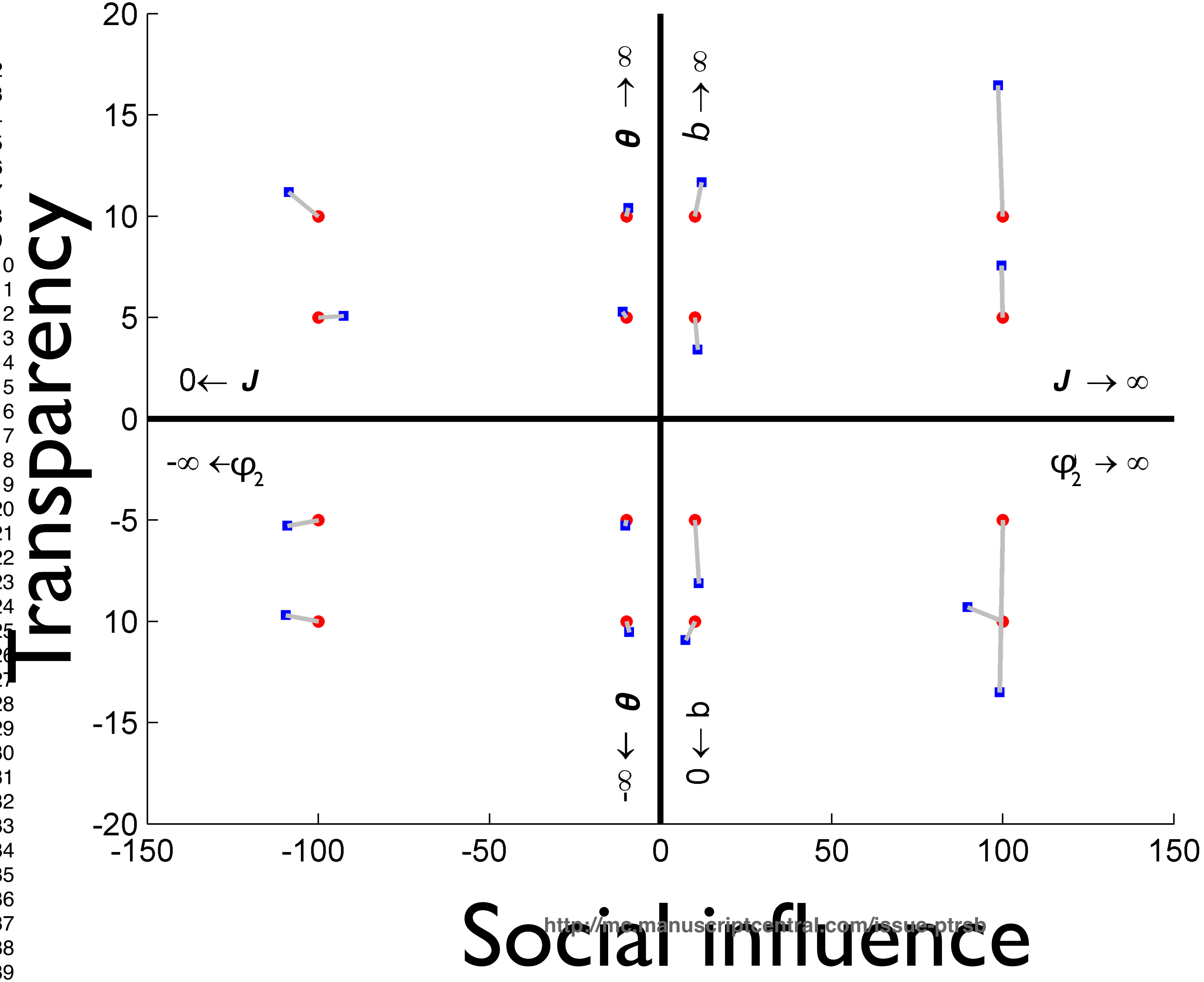

Probability
of following
Directional agreement
1.0
0.8
0.6
0.4
0.2
0.0
HIGH
LOW
1
0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1
4
7
10
13
16
19
FEW
MANY
Number of
initiators


Fitness
peak
Transparency
Social influence
0
2
4
6
8
10
1
2
3
4
5
6
7
8
9
10


2003/4
% as popular or more
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
2,000
4,000
6,000
8,000
10,000
Popularity of screening type
50%
5%
0%
1%
100%


2007/8
% as popular or more
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
20,000
40,000
60,000
80,000
Popularity of screening type
50%
5%
0%
1%
100%


Avon, Gloucestershire & Wiltshire
GUM
CSHS
GP
Pharmacy
TOP
Not Known
40%
30%
20%
10%
0%
2012
2012.5
2013
2013.5
2014
Cumbria & Lancashire
GUM
CSHS
GP
Pharmacy
TOP
Not Known
60%
50%
40%
30%
20%
10%
0%
2012
2012.5
2013
2013.5
2014
London
GUM
CSHS
GP
Pharmacy
TOP
Not Known
50%
40%
30%
20%
10%
0%
2012
2012.5
2013
2013.5
2014
North East
GUM
CSHS
GP
Pharmacy
TOP
Not Known
40%
30%
20%
10%
0%
2012
2012.5
2013
2013.5
2014


Transparency
Social influence
20
15
10
5
0
-5
10
-15
-20
-150
-100
-50
0
50
100
150
θ → ∞
b → ∞
0← J
J → ∞
-∞ ←φ2
φ2 → ∞
θ ← -∞
0 ← b