Modelling behavioural contagion

Thomas House

Abstract

The last decade has seen much work on quantitative understanding of human behaviour, with online social interaction offering the possibility of more precise measurement of behavioural phenomena than was previously possible. A parsimonious model is proposed that incorporates several observed features of behavioural contagion not seen in existing epidemic model schemes, leading to metastable behavioural dynamics.

1 Introduction

There has been much recent interest in modelling the spread of behaviours in society, particularly health behaviours with respect to infectious disease (Funk et al., 2010b). At the same time, recent empirical work highlights that the complex nature of social contagion makes it very different from 'simple' microparasite contagion (Centola, 2010). Modelling techniques have so far typically involved either explicit stochastic simulation (Centola and Macy, 2005, 2007; Salathé and Bonhoeffer, 2008; Funk et al., 2009), or else application of mathematical models originally developed for other applications, such as the Susceptible-Infectious-Susceptible (*SIS*) epidemic model considered by Kiss et al. (2010) and Funk et al. (2010a). An alternative is to use a discrete-time formalism (Dodds and Watts, 2004, 2005), next-generation arguments (Funk et al., 2009) or methods from statistical physics (de Oliveira, 1992; Pereira and Moreira, 2005) to obtain results about asymptotic behaviour of socially-motivated models, although typically calculating transient features of system dynamics requires Monte-Carlo simulation.

While existing dynamical models have clearly significantly clarified thinking about behavioural spread, and have also motivated important empirical work, they often suffer from lack of mathematical transparency, or are not customised specifically to social contagion. In this paper, a mathematical model that has a small number of easily interpretable parameters is proposed, which reproduces several key features of empirical work and empirically motivated simulation.

2 Methods

2.1 General model

The general model framework is described as follows. Consider a large, closed population, with a proportion $B(t)$ of that population engaging in a behaviour at a given time t. At a given time, each individual is canvasing the opinions of

n other individuals in the population in such a way that the proportion of individuals in the population canvassing m individuals engaging in the behaviour in question is D_m (which depends on $B(t)$ in addition to other static parameters). We assume that individuals with m canvased neighbours who are engaging in the behaviour commence at rate τ_m or cease at rate γ_m as appropriate for their current behaviour state. The dynamical system for behaviour prevalence in the population at time t is then

$$\dot{B}(t) = \sum_{m=0}^{n} D_m(t) \left((1 - B(t))\tau_m - B(t)\gamma_m\right) \ . \tag{1}$$

To specify an integrable system, it is then necessary to define a form for the dynamical parameters τ_m, γ_m and a process for the generation of the proportion D_m.

2.2 Dynamical parameters

We now choose a form for the vectors $(\tau_m), (\gamma_m)$. It is worth noting that the general form above can be specialised to incorporate several other dynamical forms. For example, if $\gamma_m = \gamma$ and $\tau_m = m\tau$ we recover the *SIS* dynamics of Funk et al. (2010b) and Kiss et al. (2010). As another example, the approach of Salathé and Bonhoeffer (2008) takes $\tau_m = m/n, \gamma_m = (n-m)/n$. Importantly, both of these schemes only depend on the mean of D_m and so are unaffected by the different distributions proposed later. Dodds and Watts (2004, 2005) consider generalised 'dose response' behaviour, which in the simplest case is a discrete-time version of the simplest continuous-time model considered here. The more sophisticated models analysed by Dodds and Watts make use of the discrete-time framework to consider agents with memory whilst preserving independent sampling of the population, while here dynamics remain Markovian but the population samples are potentially dependent.

For opinion dynamics, motivated by a comprehensive review of the literature and compelling empirical evidence (Centola and Macy, 2007; Centola, 2010), we expect an S-shaped curve for response of behavioural transmission probability to number of encounters with a behaviour. For simplicity, the limiting case of such a curve is taken so that

$$\tau_m = \begin{cases} \tau & \text{if } m \geq a \ , \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

This complex form for transmission has not yet been included in other dynamical-systems models of behaviour spread, and is the main benefit of the modelling approach considered here. We assume for simplicity that cessation of behaviour happens over time at a rate independent of m, and for convenience work in units of time where $\gamma_m = 1$. Where a is close to $n/2$, then there will be similarities between these transmission dynamics and majority vote models (e.g. de Oliveira, 1992; Pereira and Moreira, 2005) although behaviour cessation will be qualitatively different.

2.3 Canvasing method

To complete our model description, we need a form for the proportion D_m. The simplest assumption is that there are n independent trials with each trial having

probability $B(t)$, meaning that

$$D_m = \mathrm{Bin}(m|n, B(t)) \ , \tag{3}$$

where Bin() is a binomial probability mass function as defined in the Appendix. This is interpreted as each individual canvasing the opinion of n individuals, chosen at random from the whole population. We now consider two different generalisations of the binomial distribution through different models of canvasing.

2.3.1 Clustering

To introduce clustering to the trials, we consider the method of Klotz (1973) (and the parameterisation of Lindqvist (1978)) for generation of D_m. In this construction, the n individuals canvased have states $\{X_i\}_{i=1,\ldots,n}$, which are stochastic variables taking the value 1 for individuals engaging in the behaviour and 0 otherwise. These are chosen sequentially with

$$\begin{aligned} \Pr(X_1 = 1) &= B(t) \ , \\ \Pr(X_i = 1|X_{i-1} = 1) &= B(t) + c(1 - B(t)) \ , \\ \Pr(X_i = 0|X_{i-1} = 0) &= (1 - B(t)) + cB(t) \ . \end{aligned} \tag{4}$$

This introduces one static parameter, the clustering $c \in [0, 1]$. The full distribution D_m for each possible value m of $\sum_i X_i$ that follows from this construction is not reproduced here due to its complexity, but can be found in equation (3.1) of Klotz (1973).

2.3.2 Homophily

Homophily is the social process of "associating with like people," and could be modelled in the framework presented here by stratifying the population as standard epidemic models represent risk groups (Keeling and Rohani, 2007). This would increase the dimensionality of the dynamical system, and remove much of its attractive simplicity. An alternative is to model homophily as a partition of the population into self-loving groups. This means that each individual canvases without replacement from a finite group of size N. A homophily parameter h can then be defined through

$$h := \frac{n}{N} \in [0, 1] \ , \tag{5}$$

so that as $N \to \infty$ individuals canvas the whole population, leading to the minimum homophily value of 0, and where $N = n$ individuals canvas all of their homophily group leading to the maximum $h = 1$. Where M is the largest integer less than $NB(t)$, a well behaved distribution is then

$$D_m = (1+M-NB(t))\mathrm{Hyp}(m|N, M, n)+(NB(t)-M)\mathrm{Hyp}(m|N, M+1, n) \ , \tag{6}$$

where $\mathrm{Hyp}(m|N, M, n)$ is the hypergeometric distribution, representing the probability of m successful trials out of n, drawing without replacement from a population of size N with M individuals in the positive state. Equation (6) assumes that homphily groups are as representative as possible of the prevalence of belief in the population. This assumption therefore represents a limiting case of

the process that generates finite groups. While in practice these groups are likely also to be heterogeneous with respect to behaviour prevalence, such heterogeneity is similar to the clustering introduced above, and so once we have determined the impact of clustering, it makes sense to consider homophily at minimal values of clustering to deliver an unambiguous dynamical signature.

3 Results and Discussion

Having defined an Ansatz for a model of behavioural contagion, Equation (1) becomes a closed system with one dynamical variable $B(t)$, a real transmission parameter τ and an integer threshold for adoption of behaviour, a. We also defined two methods for canvasing of opinion that introduce a neighbourhood size n, and either clustering c or homophily h. Having an ODE-based dynamical system as a model means that critical behaviour, in particular the ability of a behaviour to become established in a sizeable proportion of the population, can be evaluated exactly (meaning at machine precision) and numerical integration is not computationally intensive. At the same time, this model includes the feature of complex contagion as defined by equation (2), meaning that it can capture behaviour not present in, for example, the *SIS* model.

While general analytical results for this model are not obvious, if we consider the case where $n = 2, a = 2$ then there are three fixed points of the system with complex contagion:

$$B_0^* = 0 \ , \quad B_1^* = \tfrac{1}{2}\left(1 - \sqrt{1 - (4\gamma/\tau)}\right) \ , \quad B_2^* = \tfrac{1}{2}\left(1 + \sqrt{1 - (4\gamma/\tau)}\right) \ . \quad (7)$$

When $\tau < 4\gamma$, only the behaviour-free steady state exists, and is stable. When $\tau > 4\gamma$, B_0^* and B_2^* are stable steady states, with B_1^* being an unstable fixed point above which the system evolves towards B_2^* and below which the system evolves towards B_0^*. This is in contrast to *SIS* dynamics where there are only two fixed points: $\tilde{B}_0 = 0$, which is stable when $2\tau < \gamma$, and $\tilde{B}_1 = 1 - (\gamma/(2\tau))$, which is stable when $2\tau > \gamma$.

Figure 1 shows some results from numerical integration of the model. Panel (a) shows three of the distributions considered: binomial; clustered and homophilous. In (b), we see one of the main features of this model that is qualitatively different from *SIS* dynamics: complex contagions are metastable, with both the 'behaviour-free' and 'established behaviour' steady states being absorbing. As Centola and Macy (2005) argued, this is a necessary feature for explaining how initially unpopular norms can become established and maintained through social pressure.

Also in Figure 1, the impact of clustering (Panel (c)) and homophily (Panel (d)) on behavioural dynamics is shown. This provides a mathematical explanation for the results seen in empirical work and simulation (Centola and Macy, 2007; Centola, 2010), namely that clustering enhances behavioural transmission, while homophily (as defined here, subject to caveats about interpretation) reduces behavioural transmission. The non-monotonicity seen in Panel (d) is just an artefact of the discretisation equation (6). These effects are not seen for simple transmission, which only depends on the mean of the distribution D_m and so is unaffected by changes in clustering c or homophily h.

In summary, the mathematical model introduced here complements and develops existing work in three main ways. Firstly, it incorporates many of the

advantages of simple transmission models like the *SIS* model, in that the threshold behaviour, fixed points, transient behaviour and parameter sensitivity can be calculated numerically at machine precision. Secondly, the rates and processes defined implicitly in equation (1) can be used to define a natural stochastic model using the methods of Dangerfield et al. (2009). Since there are relatively few parameters, this opens up the possibility of rigorous statistical fitting of model parameters, although finding a robust method for inference and sufficiently high-quality data is likely to pose a significant challenge. Finally, the mathematical transparency of the model acts as a guide to intuition, meaning that the exact causes of effects seen in more sophisticated simulations and empirical work can be better interpreted.

Acknowledgements

Work supported by the UK Engineering and Physical Sciences Research Council (grant number EP/H016139/1). The author would like to thank Martine Barons and Matt Keeling for helpful comments relating to this work.

References

D. Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–7, Sep 2010.

D. Centola and M. Macy. The Emperor's dilemma: A computational model of self-enforcing norms. *American Journal of Sociology*, 110(4):1009–40, Jan 2005.

D. Centola and M. Macy. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–734, Jul 2007.

C. Dangerfield, J. Ross, and M. Keeling. Integrating stochasticity and network structure into an epidemic model. *Journal of the Royal Society Interface*, 6 (38):761–774, Sep 2009.

M. J. de Oliveira. Isotropic majority-vote model on a square lattice. *Journal of Statistical Physics*, 66(1/2):273–281, 1992.

P. S. Dodds and D. J. Watts. Universal behavior in a generalized model of contagion. *Physical Review Letters*, 92(21):218701, Jan 2004.

P. S. Dodds and D. J. Watts. A generalized model of social and biological contagion. *Journal of Theoretical Biology*, 232(4):587–604, Jan 2005.

S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen. The spread of awareness and its impact on epidemic outbreaks. *PNAS*, 106(16):6872–7, Apr 2009.

S. Funk, E. Gilad, and V. A. A. Jansen. Endemic disease, awareness, and local behavioural response. *Journal of Theoretical Biology*, 264(2):501–509, May 2010a.

S. Funk, M. Salathé, and V. A. A. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of The Royal Society Interface*, 7(50):1247–1256, Sep 2010b.

M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2007.

I. Z. Kiss, J. Cassell, M. Recker, and P. L. Simon. The impact of information transmission on epidemic outbreaks. *Mathematical Biosciences*, 225(1):1–10, May 2010.

J. Klotz. Statistical inference in Bernoulli trials with dependence. *The Annals of Statistics*, 1(2):373–379, Mar 1973.

B. Lindqvist. A note on Bernoulli trials with dependence. *Scandinavian Journal of Statistics*, 5(4):205–208, Jan 1978.

L. F. C. Pereira and F. G. B. Moreira. Majority-vote model on random graphs. *Physical Review E*, 71(1):016123, Jan 2005.

M. Salathé and S. Bonhoeffer. The effect of opinion clustering on disease outbreaks. *Journal of the Royal Society Interface*, 5(29):1505–8, Dec 2008.

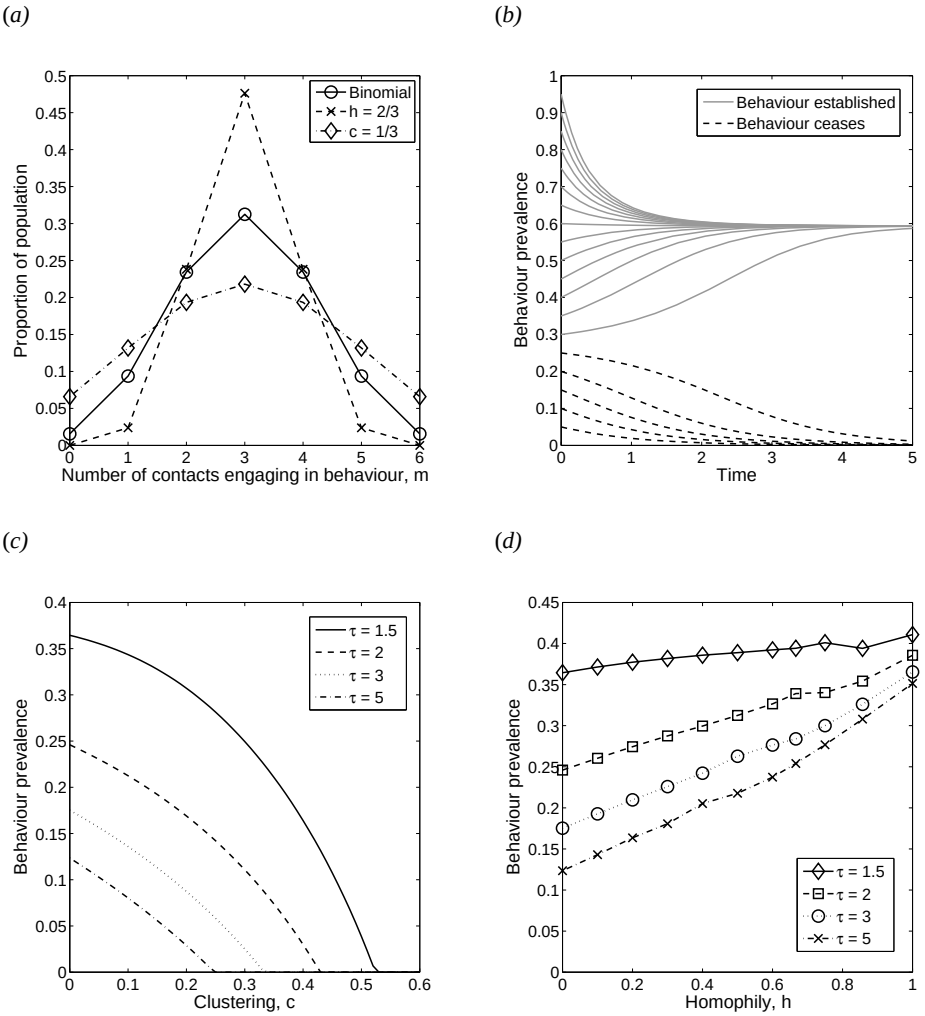

Figure 1: Numerical results for the behavioural model. Parameters $n = 6, a = 3$ are common to all figures. (a) shows the proportion D_m for the baseline, binomial model, in addition to the clustered distribution with $c = 1/3$ and the homophily model with $h = 2/3$. (b) shows temporal dynamics of a complex contagion with $\tau = 1.8, c = h = 0$. Depending on the initial prevalence of behaviour, both the 'behaviour-free' and 'established behaviour' steady states can be reached, and are stable. The phase space of complex contagion is also shown for (c) clustering and (d) homophily. For given values of τ, c and h, different curves show the minimum values of behaviour prevalence that are needed for the system to evolve towards the 'established behaviour' steady state.

A Statistical notation

The binomial coefficients are given by

$$\left(\begin{array}{c} n \\ m \end{array} \right) = \frac{n!}{m!(n-m)!} \ . \tag{8}$$

The binomial probability mass function is, for integer $m \in \{0, \ldots, n\}$,

$$\mathrm{Bin}(m|n,p) = \left(\begin{array}{c} n \\ m \end{array} \right) p^m (1-p)^{n-m} \ . \tag{9}$$

The hypergeometric probability mass function is, for integer $m \in \{0, \ldots, n\}$,

$$\mathrm{Hyp}(m|N,M,n) = \frac{\left(\begin{array}{c} M \\ m \end{array} \right) \left(\begin{array}{c} N-M \\ n-m \end{array} \right)}{\left(\begin{array}{c} N \\ n \end{array} \right)} \ . \tag{10}$$