Conformity or Diversity: Social Implications of Transparency in Personal Data Processing

WORKING PAPER

Rainer Böhme

Technische Universität Dresden
Institute of Systems Architecture
01062 Dresden, Germany

rainer.boehme@tu-dresden.de

Abstract. Consider the hypothetical situation of a society with virtually unconstrained storage and exchange of personal information and shameless exploitation thereof for decision making, for example in contract negotiation. In this paper we develop a stylised formal model to tackle the question if public knowledge about *how exactly* personal information is used in decision making changes aggregate behaviour. Simulation results suggest a slightly positive relationship between transparency and conformity. This has implications on the common conjecture that collection and processing of personal information is tolerable as long as transparency is warranted.

Keywords: Economics of Privacy, Transparency, Privacy-Enhancing Technologies (PET), Transparency-Enhancing Technologies (TET)

1 Introduction

Individuals, in participating in social interaction, share information about themselves with others. The advent of information and communication technologies as tools and means for social interaction reduces the cost to collect, store, combine, and process such information. It is well understood that accumulated personal information from past transactions can create information asymmetries in future transactions between the same agents [1] and, if information is traded, even for transaction between agents who have never interacted before [2, 3]. Hence, data collection has attracted criticism from consumer and civil rights organisations, which reinforced a debate on privacy rights and informational self-determination. As a result, since the 1970s, most countries have passed legislation to deal with privacy concerns in state-to-individual and business-to-individual (consumer) interactions.

Since the 1980s, following the earlier vision of Baran [4], computer scientists have increasingly researched into technical solutions to combat the privacy problems caused by technological progress. Technologies such as anonymous commu-

7th Workshop on the Economics of Information Security 2008
Tuck School of Business, Dartmouth College, Hanover, NH, June 25–27
Last revision: April 24, 2008

nication infrastructures, formalisation of privacy policies (e.g., P3P [5]), automatic enforcement (access control [6]), as well as protocols for pseudonymous but accountable transactions are nowadays subsumed as *privacy-enhancing technologies* (PETs) [7–9]. Most PETs are designed to support data avoidance, which allows the construction of systems that are secure against relatively strong adversaries by relying on distributed architectures. The objective is to minimise the amount of trust required in individual transaction partners. Although some PETs can be designed very securely in theory, their principle of data avoidance/reduction is deemed impractical for many applications and the prospects for a wide adoption of PETs in the near future remain dim. PETs are typically designed for $1 : 1$ or $1 : n$ interactions in which each partner has full control over his or her devices and the signals they emit. We are not aware of practical solutions for privacy-preserving $n : m$ interactions (although problem descriptions can be found in the literature, e.g. [10]) beyond very specific protocols for transactions with clearly defined semantics (for instance, cryptographic voting schemes or private multi-party auction protocols [11]).

1.1 From PETs to TETs

In the light of online social networking sites, sensor networks, ambient intelligence and behavioural biometrics, where $n : m$ interactions and untrusted devices (sensors) are the rule rather than exceptions, it becomes evident that data avoidance most likely will not offer a solution for privacy threats in general social interactions. Data avoidance cannot be enforced at all by individuals alone, and only at unacceptably high costs by regulation (i.e., in the last consequence, only through restrictions on the ownership of freely programmable devices or sensors). Therefore, operable alternatives are sought.

Transparency-enhancing technologies[1] (TETs) are believed to be more viable options in these situations [12, 13]. The idea is to inform people in detail how personal attributes (might) affect decisions concerning themselves. Consider an example where personal information is used for insurance red-lining or credit scoring. If affected individuals cannot escape the data collection, then they should at least know how exactly a certain data disclosure, such as moving in a statistically more 'risky' area, will affect their future premium or credit conditions. One can argue that transparency limits excessive discrimination on the base of personal information through three channels: First, on an individual level, pre-emptive transparency-enhancing technologies assist people in making decisions which do not affect their personal 'score' adversely. Second, on a mechanism level, scoring procedures that are not strategy proof, or the effectiveness of which depends on the scoring details to remain obscure, become less useful and would thus be avoided. Third, on a social level, if public scrutiny reveals that a particular scoring function is arbitrarily discriminating and as such incompatible with the society's values, the risk of public uproar and reputation damage might

[1] The notion of 'technology' is rather broad. For example, a sign informing pedestrians about video surveillance in public places can be seen as a (low-tech) TET.

put social pressure on data controllers not to implement abusive practices in the first place. Note that all these outcomes depend on the optimistic assumption that the TET is honest about the true data processing habits, a requirement that is difficult to verify and enforce. So TETs, like PETs, are no panaceas that solve all privacy concerns of a modern society.

1.2 TETs and individual behaviour

The topic of this paper is to study the effects of TETs on social behaviour, more precisely on its impact to diversity in behaviour. Diversity between individuals, i.e. the extent to which individuals live their own life-style, is considered as valuable precondition in political and economic theory, where diversity is liked to concepts of pluralism and competition, respectively.

At the first sight, two conflicting hypotheses on the relation between transparency and diversity can be formulated intuitively.

- **Transparency supports conformity** because, in the absence of information asymmetries and strategic interaction with others, the optimal path is obvious and becomes 'mainstream'.
- **Transparency supports diversity** because, without transparency, individuals are herded together by uncertainty and fear. The rationale under uncertainty is not to stand out of the mass because the mass would barely err (cf. Lundblad's notion of a *noise society* [14]).

The objective of the remainder of this paper is to develop a formal model with which the conflict between these two hypothesis can be resolved. While fully acknowledging the potential problems of formal models, we will propose (and put up for discussion) a multi-period game with heterogeneous preferences and analyse under which conditions this prior heterogeneity is best preserved in rational individuals' actions.

2 Model

Imagine a world where each individual stores all information about social interactions, possibly combines his or her database with others (in a market for information, so prices for database peering may be negotiated), and uses this information as decision support in future transactions. For simplicity, we rule out any ambiguity and assume that all information is authentic and individuals are perfectly identifiable.

2.1 Assumptions

The following list of assumption defines our model. The rationales behind them are reported separately in Sect. 2.3 for the sake of clarity. A list of all symbols used in this paper can be found in the Appendix.

1. Individuals $I^{(1)},\dots,I^{(n)}$ are endowed with heterogeneous private preferences $p^{(i)}$ and initial wealth $v^{(i)}=1$.
2. The preference space is the circumference of a unit ring, with position drawn independently from a uniform distribution between 0 an 1, i.e. $p^{(i)}\in[0,1)$.
3. The system is updated in rounds. In each round k, all individuals emit a signal $s_k^{(i)}\in[0,1)$.
4. The cost of emitting a signal is a weighted sum of two components $c_{\text{emit}}=\alpha c_{\text{pret}}^{(i)}+\beta c_{\text{disc}}^{(i)}$.
   (a) The *pretence* component increases with the distance from the individual private preference $p^{(i)}$, hence $c_{\text{pret}}=D(s_k^{(i)},p^{(i)})$. We define function $D:[0,1)^2\to[0,1]$ as four times the square of the (shortest) distance between two points on the ring.

   $$D(x,y)=\begin{cases}4\cdot(x-y)^2 & \text{for } |x-y|\le\frac{1}{2}\\ 4\cdot(1-|x-y|)^2 & \text{otherwise}\end{cases}\tag{1}$$

   Note that D is symmetric and invariant to translation of its arguments on the unit ring: $D(x,y)=D(y,x)$ and $D(x,y)\equiv D(x+k \bmod 1, y+k \bmod 1)$.

   (b) The *discontinuity* component is proportional to the distance between the emitted signal in the current round $s_k^{(i)}$ and in the past round $s_{k-1}^{(i)}$, hence $c_{\text{disc}}=D(s_k^{(i)},s_{k-1}^{(i)})$. $c_{\text{disc}}^{(i)}:=0$ in the first round of each individual.

   Parameters α and β control the discomfort of dynamic adjustment in relation to the discomfort of pretending different preferences.
5. There is a global entity who punishes individuals depending on their emitted signals. The *penalty* is calculated as inverse distance between the signal s and a focal point d: $c_{\text{pen}}^{(i)}=\left(1-D(s_k^{(i)},d)^{\frac{1}{2}}\right)^2$.
6. The existence of TET is modelled as knowledge about d. We will compare a scenario without TET, where individuals do not know d, with one in which all individuals know the exact position of d (through TET).
7. Total cost

   $$c_{\text{tot}}^{(i)}=c_{\text{emit}}^{(i)}+\gamma c_{\text{pen}}^{(i)}+\nu=\alpha c_{\text{pret}}^{(i)}+\beta c_{\text{disc}}^{(i)}+\gamma c_{\text{pen}}^{(i)}+\nu\tag{2}$$

   are deducted from wealth $v^{(i)}$ at the end of each round.
8. Individuals default if their wealth $v^{(i)}$ turns negative. There is no possibility to transfer wealth between individuals, so no borrowing is allowed. Defaulted individuals are re-initialised in the next round ($v^{(i)}=1$, new realisation of $p^{(i)}$), thereby losing their history of observations.
9. Individuals know the global parameters α,β, γ and ν as well as the set of emitted signals from the last round. Apart from that, there is no communication between individuals (in particular no sharing of knowledge about the possible position of focal point d).
10. Individuals act fully rational and maximise their own expected time to default. When indifferent between two alternatives which would lead to the same number of rounds before default, they prefer the option where $|v^{(i)}|$ is smaller after default.

2.2 Problem statement

We use this model to study the relation between transparency and conformity with Monte-Carlo simulations. After initialisation, the model is updated over N rounds. At the end of each round, we compute two dependent variables:

1. A measure of *conformity* between individuals ψ_k, which is defined as the square sum of the (shortest) absolute distance between neighbouring signals $s_k^{(i)}$ on the preference ring, linearly scaled to the range from $\psi_{\min} = 0$ (perfect distribution; all signals are equidistant) to $\psi_{\max} = 100$ (full conformity; all signals equal). This metric, aggregated over all rounds, serves as indicator variable to answer the research question.
2. The *mean time to default* of all individuals who have defaulted in this round. This metric can be interpreted as a co-variate for a concept like (negative) 'social cost of information asymmetries'.

Both measures are calculated *per round* and than aggregated *over time*. This means that conformity should not be interpreted in an inter-temporal fashion, like concepts such as stability over time. Note that the valuation of diversity (i.e., inverse conformity) as a desirable property, as outlined in Sect. 1.2, is exogenous to this model and not accounted (e.g., as negative social cost) in our metric for the mean time to default. We do not make an attempt to combine both metrics to a single scalar utility metric.

2.3 Rationales for the assumptions

In the following we list the rationales that have lead to our model formulation. The ones printed in bold are important for understanding our design decisions.

- As to assumption 2: We choose the circumference of a unit circle to avoid discontinuities at the margins of the preference space. This also ensures that a pair of locations is equally distant from d. The distribution between these points, based on individual preferences, can be interpreted as diversity.
- **As to 3:** Signals correspond to information disclosed in social interaction. Individuals have the possibility to hide their true preferences if they deem this advantageous in the long run. However this comes at a cost. For example, if someone prefers not to disclose his home address to an online retailer, he or she has to bear the transaction costs of going to a bricks-and-mortar store. Also refraining from engaging in a transaction for privacy concerns can be seen as incurred opportunity cost.
- As to 4a: A quadratic distance function is a technical assumption to ensure that unique minima exist (apart from some pathologic cases where two options are possible due to symmetry).
- As to 4b: The discontinuity component constrains dynamic adjustment and thus learning. If adjustment is too cheap, then some individuals will infer the centre of the penalty distribution from observations so that they gain 'transparency by experience' even in the condition without transparency.

Contrary, if adjustment is expensive, then the expectations formed in the very first round of each individual are much more important for its survival. Aside from technical considerations, discontinuity costs can be interpreted as the social cost of changing one's image, or sunk costs associated with previous actions. The fact that discontinuity costs are quadratic in the distance between two signals implies that individuals prefer making small steps over a couple of rounds rather than a single big leap.

- **As to 5**: The punishing entity models the disadvantage individuals might incur from disclosing particular personal information. Although in reality, privacy risks are caused by other people, we have chosen this asymmetric setting (in fact, a player-versus-nature game) to keep the number of strategic interdependencies low. We do not believe that this is a major factor constraining the model's external validity.
- As to 7: Cost $\nu > 0$ is a small technical offset charged in each round independent of the individual's preference and behaviour to ensure that all individuals have finite time to default. (Otherwise the model could converge in a deterministic state.)
- As to 10: Assuming unbounded rationality is often criticised (rightly so). In assuming rational behaviour, our model abstracts from what we call *awareness* aspects, which deal with the problem that people do not understand or cannot interpret the information they have – in theory – at their disposal. We acknowledge that awareness is at least as important in practice as transparency, but both concepts must be differentiated and studied separately before drawing conclusions about their joint effect.

2.4 Analytic approach

We will first discuss the optimal strategy for individuals in the simpler case of full information before we advance to cases where d is unknown.

Strategy of individuals in regime with TET Individuals $I^{(i)}$ enter the game with knowledge of d and adjust $s^{(i)} = s_k^{(i)}\ \forall k$ with respect to $p^{(i)}$ to maximise their expected lifetime, that is minimise $c_{\text{tot}}^{(i)} - \nu$.

$$c_{\text{tot}}^{(i)} - \nu = \alpha\, c_{\text{pret}}^{(i)} + \beta\, c_{\text{disc}}^{(i)} + \gamma\, c_{\text{pen}}^{(i)} \tag{3}$$

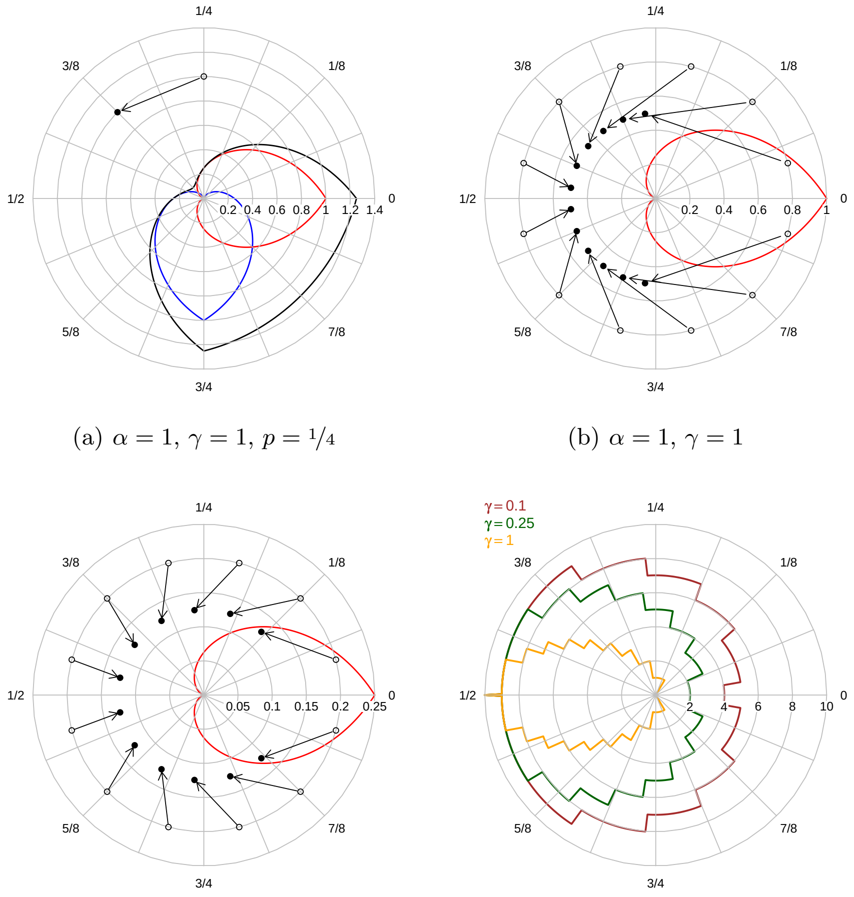
$$= \alpha\, D(s^{(i)}, p^{(i)}) + \beta \cdot 0 + \gamma \left(1 - D(s^{(i)}, d)^{\frac{1}{2}}\right)^2 \tag{4}$$

Using the fact $|s^{(i)} - d| \leq \frac{1}{2}$ (from symmetry) and regarding only cases where $d \leq \frac{1}{2}$, $d \leq p^{(i)} \leq 1$ and thus $d < s^{(i)} \leq 1$,

$$c_{\text{tot}}^{(i)} - \nu = 4\alpha\,(s^{(i)} - p^{(i)})^2 + \gamma \left(1 - 2(s^{(i)} - d)\right)^2 \tag{5}$$

The first-order condition of the minimisation problem (for $\alpha + \gamma > 0$) is

$$s^{(i)} = \frac{\alpha\, p^{(i)} + \gamma\,(d + \frac{1}{2})}{\alpha + \gamma}\ . \tag{6}$$



(a) $\alpha = 1$, $\gamma = 1$, $p = {}^1\!/_4$

(b) $\alpha = 1$, $\gamma = 1$

(c) $\alpha = 1$, $\gamma = {}^1\!/_4$

(d) $\alpha = 1$, $\nu = {}^1\!/_{10}$

Fig. 1. Preference and emitted signals under transparency. Radial plots of preference space. Symbols $\circ$ for preferences p, $\bullet$ for emitted signals s; connected lines are functions of s: red for penalty, blue for pretence cost, and black for total cost. (a): cost components and adjustment for example individual; (b) + (c): adjustment of heterogeneous individuals for varying γ (different radius for presentation clarity only); (d): time to default (in rounds) for young individuals as a function of emitted signal $s^{(i)}$ for varying γ. Focal point $d = 0$ in all plots.

Using translation invariance of D, we obtain the formula for $d > p$:

$$s^{(i)} \equiv \frac{\alpha\,(1-d+p^{(i)})+\frac{\gamma}{2}}{\alpha+\gamma} + d \bmod 1 \tag{7}$$

Equations (6) and (7) define the strategy for all individuals in the scenario where TET is available and d is public (by definition). As the cost function is fully deterministic and does not depend on other individuals' behaviour, there is no need to adjust the position in rounds $k > 1$. As a result, the weight for discontinuity costs β does not appear in the optimal strategy in this case.

We will further derive a number of metrics as a function of the absolute distance $|x-d|, 0 \leq x \leq 1/2$, which are needed below for the strategy in a regime without TET. The probability distribution function for 'young' individuals (age $k=1$) directly follows from the uniform distribution assumption for realisations of $p^{(i)}$ in $[0,1)$ and Eq. (6) solved for $p^{(i)}$.

$$f_{s_1}(x) = \text{Prob}(|s_1^{(i)} - d| < x) = \begin{cases} \frac{1}{\alpha}\left(x(\alpha+\gamma)-\frac{\gamma}{2}\right) & \text{for } \frac{\gamma}{2(\alpha+\gamma)} < x \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

As can be seen, $f_{s_1}(s^{(i)}) \equiv f_{s_1}(x + d \bmod 1)$ is a uniform distribution in the interval $\left[d + \frac{\gamma}{2(\alpha+\gamma)} \bmod 1, d - \frac{\gamma}{2(\alpha+\gamma)} \bmod 1\right]$ with density $\frac{\alpha+\gamma}{\alpha}$.

The expected time to default (measured in rounds) of young individuals with observed signals $s^{(i)}$ as a function of $x = |s^{(i)} - d|, x < \frac{1}{2}$ can be obtained directly from the cost function (assumption 7):

$$K(x) = \left\lfloor (c_{\text{tot}}^{(i)})^{-1} \right\rfloor = \left\lfloor \left[\alpha\, D(x, |p-d|) + \gamma\left(1 - D(x,0)^{\frac{1}{2}}\right)^2 + \nu\right]^{-1} \right\rfloor \tag{9}$$

$$= \left\lfloor \left[4\alpha\left(x - \underbrace{\frac{1}{\alpha}\left(x(\alpha+\gamma)-\frac{\gamma}{2}\right)}_{\text{Eq. 6 solved for } p^{(i)}}\right)^2 + \gamma\,(1-2x)^2 + \nu\right]^{-1} \right\rfloor \tag{10}$$

$$= \left\lfloor \frac{\alpha}{\gamma(\alpha+\gamma)(1-2x)^2 + \alpha\,\nu} \right\rfloor \tag{11}$$

Here we see that offset $\nu > 0$ is essential to avoid a zero denominator. Fig. 1 (d) depicts the time to default as a function of $s^{(i)}$. In the repeated game, the distribution of all observable signals f_{s_k} (as opposed to f_{s_1} for young individuals only) is proportional to the time to default.

Strategy of individuals in regime without TET We use a heuristic strategy to model the behaviour of individuals if d is unknown.[2]

[2] Although we have no proof that our strategy is optimal in the sense that it makes best use of all available information to narrow down the position of d as tight as possible, we believe that our algorithm is a quite good approximation. This conjecture is supported by experiments with small deviations in our simulation environment.

Step 1 – Choice of $s_1^{(i)}$: After initialisation, an individual $I^{(i)}$ knows the rules of the game, the global parameters $(\alpha, \beta, \gamma, \nu)$, its own preference $p^{(i)}$ and $m < n$ signals $s_0^{(j)}$ of individuals in the previous round. Neither $p^{(j)}, v^{(j)}, (j \neq i)$, nor the age of other individuals are observable.

The best guess of d is a solution to the maximum-likelihood (ML) problem

$$\begin{aligned} \hat{d}_1^{(i)} &= \arg\max_x \ \text{Prob}(s^{(1)}, \dots, s^{(j)}, \dots, s^{(m)} | d = x, \forall j \neq i) \\ &= \arg\max_x \prod_j \text{Prob}(s^{(j)} | d = x) = \arg\min_x \sum_j - \log \text{Prob}(s^{(j)} | d = x) \end{aligned} \tag{12}$$

The conditional probability can be obtained from Eq. (11), where we omit the truncation to integers to smooth the gradient for numerical solvers, and scale to

$$\int_{\frac{\gamma}{2(\alpha+\gamma)}}^{1-\frac{\gamma}{2(\alpha+\gamma)}} f_{s_k}(x)\, dx = 1 \ . \tag{13}$$

Then, $s_1^{(i)}$ is calculated from $\hat{d}_1^{(i)}$ using Eqs. (6) and (7), as in the case of transparency. Performance indicators for the ML estimate of d dependent on γ and the number of individuals n are reported in Tab. 2 in the appendix.

Step 2 – Two candidates for $\hat{d}_2^{(i)}$: In the second round, individual $I^{(i)}$ has experienced cost $c_{\text{tot},1}^{(i)}$ and thus can find out $c_{\text{pen},1}^{(i)}$ using Eq. (2). Since $c_{\text{pen},1}^{(i)}$ reveals distance $|s_1^{(i)} - d|$, this narrows down the possible location of d to two candidates, $\hat{d}_{2+}^{(i)}$ and $\hat{d}_{2-}^{(i)}$. There are at least two options to decide between the candidates.

1. The *static* solution is to compare the likelihood for $\hat{d}_{2+}^{(i)}$ and $\hat{d}_{2-}^{(i)}$, possibly with observations from both rounds $s_0^{(j)}$ and $s_1^{(j)}$ to lower the estimation standard error (although not a lot, as $s_0^{(j)}$ and $s_1^{(j)}$ are not independent).
2. There is also a *dynamic* solution based on the rationale that no individual would ever reduce its distance to d. Therefore a comparison of signals $s_0^{(j)}$ and $s_1^{(j)}$ contains information about the dynamic adjustment of other individuals and therefore conveys information about the most likely location of d.

In practice, both solutions come to the same conclusions in the large majority of cases. We have not investigated ways to combine the information optimally or resolve conflicting results. Our experiments are based on a static update of $\hat{d}_2^{(i)}$. Signal $s_2^{(i)}$ is choses using the step size rule described below (Eq. 17) with a target position calculated from the refined estimate $\hat{d}_2^{(i)}$.

Steps 3 and later – Optimal adjustment to d: The focal point d can be obtained exactly from $c_{\text{pen},1}^{(i)}$, $c_{\text{pen},2}^{(i)}$, $s_1^{(i)}$ and $s_2^{(i)}$. Finding the optimal step sizes to approach the ideal position $s_*^{(i)}$ conditional to d is a discrete dynamic optimisation problem. However, we argue that for $0 < \beta \leq 5$, the problem is posed in such a way that a reasonably good solution can be found sequentially by minimising the cost *in the current round.*[3] So, again, we minimise $c^{(i)}_{\mathrm{tot},k} - \nu$.

$$c^{(i)}_{\mathrm{tot},k} - \nu = \alpha\, c^{(i)}_{\mathrm{pret},k} + \beta\, c^{(i)}_{\mathrm{disc},k} + \gamma\, c^{(i)}_{\mathrm{pen},k} \tag{14}$$

$$= \alpha\, D(s^{(i)}_k, p^{(i)}) + \beta\, D(s^{(i)}_k, s^{(i)}_{k-1}) + \gamma \left(1 - D(s^{(i)}_k, d)^{\frac{1}{2}}\right)^2 \tag{15}$$

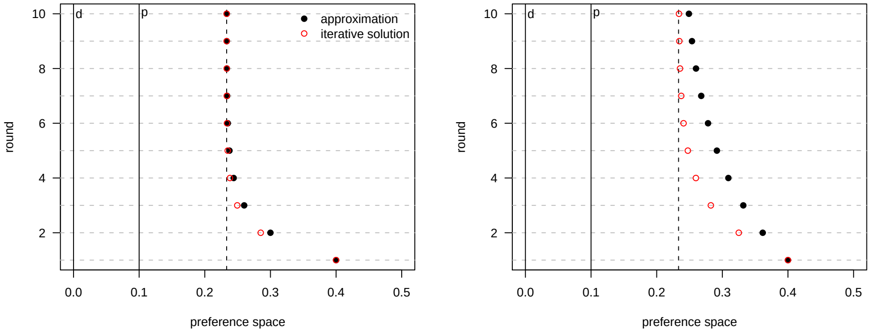
Restricting the analysis to cases where $d \leq \frac{1}{2}$, $d \leq p^{(i)} \leq 1$ and thus $d < s^{(i)}_k, s^{(i)}_{k-1} \leq 1$.

$$c^{(i)}_{\mathrm{tot},k} - \nu = 4\alpha\,(s^{(i)}_k - p^{(i)})^2 + 4\beta\,(s^{(i)}_k - s^{(i)}_{k-1})^2 + \gamma \left(1 - 2(s^{(i)}_k - d)\right)^2 \tag{16}$$

This leads to the first-order condition (and symmetric equivalents)

$$s^{(i)}_k = \frac{\alpha\, p^{(i)} + \beta\, s^{(i)}_{k-1} + \gamma\left(d + \frac{1}{2}\right)}{\alpha + \beta + \gamma}\ . \tag{17}$$

We have also implemented a numeric iterative solver for the dynamic minimisation problem and found that it leads to substantially better solutions only when β is large (see Fig. 2) and initial estimates $\hat{d}^{(i)}_1$ bad. Both occurs rarely in our experiments.



(a) $\beta = 1$ (b) $\beta = 5$

Fig. 2. Difference in adjustment step sizes between myopic and inter-temporal optimal solution. $d = 0, p = 1/10, \alpha = 1, \gamma = 1/2$. Signal $s^{(i)}_{k-1} = 0.4$ is very unrealistic and only chosen to emphasise the difference. The total cost disadvantage of the approximation (until convergence) is 0.7 % (left) and 4.7 % (right).

[3] This means that individuals are myopic or uncertain about the default threshold.

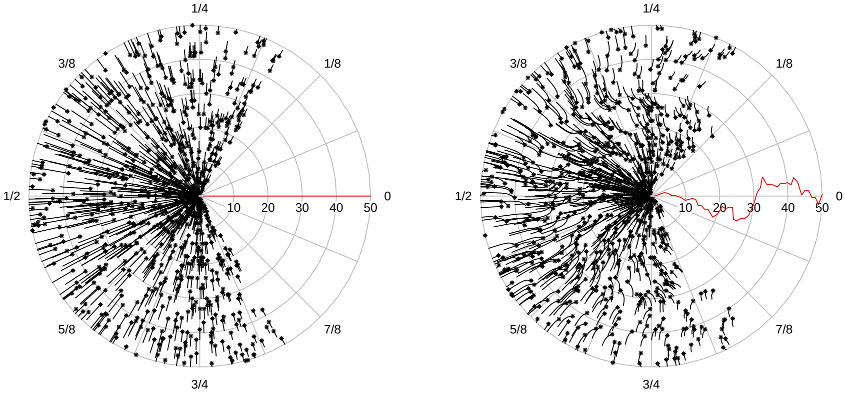
3 Results

It is obvious that the diversity measure depends on parameters $(\alpha, \beta, \gamma, \nu)$ as discontinuity costs clearly determine the individuals' ability to emit favourable signals. Therefore, we will compare the diversity of systems conditional to these parameters.

To structure the discussion of results, we fix parameters $\alpha = 1$, $\nu = 1/10$ and $n = 100$ for what we call *baseline results*. We compare the case of transparency (TET available) with no transparency for different severity of disadvantage due to unfavourable personal attributes: small ($\gamma = 1/10$), medium ($\gamma = 1/2$), and substantial ($\gamma = 1$) 'privacy infringement'. In the case of diversity, we further differentiate between low ($\beta = 1/10$), medium ($\beta = 1$) and high ($\beta = 4$) discontinuity costs.

Figure 3 shows two simulation snapshots for selected parameters with and without transparency (see figure caption for more details), and aggregate measures of conformity and time to default are reported in Tab. 1 for all relevant parameter combinations.

For the *baseline results*, it turns out that conformity ψ is always maximal in the case of transparency, although the relative difference to the simulations without transparency is rather small. We could confirm this tendency in many other parameter settings not reported here. We interpret this as a tentative support of



(a) transparency

(b) no transparency, $\beta = 1$

Fig. 3. Visualisation of a simulation snapshot of 50 subsequent rounds. The preference space is mapped to the polar axis and time increases with distance from the origin. Emitted signals from unique individuals are black connected lines. Each $s_1^{(i)}$ is annotated with symbol $*$. The red line shows the evolution of $\hat{d}_1$ over time. Parameters are $n = 100, d = 0, \alpha = 1, \gamma = 1/2, \nu = 1/10$.

Table 1. Simulation results: Impact of transparency-enhancing technologies

	conformity ψ				mean time to default			
	transparency	no transparency			transparency	no transparency		
		$\beta = 1/10$	$\beta = 1$	$\beta = 4$		$\beta = 1/10$	$\beta = 1$	$\beta = 4$
Baseline results								
$n = 100$, $\nu = 1/10$								
$\gamma = 1$	27.5	27.3	26.3	26.2	5.6	5.6	5.6	5.6
$\gamma = 1/2$	12.8	11.6	12.2	12.4	6.3	6.3	6.3	6.3
$\gamma = 1/10$	2.0	1.5	2.0	1.9	8.4	8.4	8.4	8.4
Early default								
$n = 100$, $\nu = 1/4$								
$\gamma = 1$	26.1	24.5	24.5	23.7	3.1	3.1	3.1	3.1
$\gamma = 1/2$	12.4	11.7	11.5	10.0	3.4	3.4	3.4	3.4
$\gamma = 1/10$	1.9	1.4	1.2	1.4	4.0	3.9	3.9	3.9
Limited population sample								
$n = 10$, $\nu = 1/10$								
$\gamma = 1$	40.1	39.3	37.7	38.9	5.6	5.3	5.4	5.5
$\gamma = 1/2$	32.6	20.7	20.5	26.9	6.2	6.0	6.1	6.3
$\gamma = 1/10$	10.6	10.1	11.7	8.4	8.5	8.2	8.3	8.3

Aggregate metrics computed from 1000 iterations ($\alpha = 1$)

our first hypothesis (transparency supports conformity), but the probably more interesting result is that the influence of transparency on diversity is so small. The much higher differences in ψ and mean time to default *between* different values for γ are not surprising, as γ directly influences the dispersion of the ideal distribution of individuals in the signal space (see Eq. 8). The mean time to default is approximately independent from the presence of transparency (in fact, at a higher precision, individuals in games without transparency default slightly earlier on average). This observation as well as the constant time to default for all values of β indicates that $n = 100$ individuals provide enough information for sufficiently precise estimates of $\hat{d}_1$. In other words, the information disadvantage of fully rational individuals without TET is rather small in our model compared to full transparency. Arguably, this can be seen as unrealistic, so we check the robustness of our results with two different parameter settings that both aim at limiting the 'information leakage' from older individuals to young ones.

The *early default* result set accomplishes this goal by artificially high constant costs ν. As a result, mean time to default drops to roughly one half of the baseline results. This ensures that the fraction of experienced (i.e., 'old' an thus better adjusted) individuals in the population decreases (see Fig. 4 (a) in the appendix). Nevertheless, this does not alter our conclusion on conformity; quite the contrary: the relative conformity gain in the case of transparency even widens.

The findings on conformity remain broadly stable in the third result set tagged *limited population sample*.[4] The idea here is to increase the uncertainty of the estimate $\hat{d}_1$ by reducing the number of individuals in the game to $n = 10$. This can be interpreted as a kind of awareness constraint in reality, i.e. individuals typically have no means to observe the population as a whole but rather some randomly drawn peers. The higher estimation errors that cause discontinuity costs during adjustment can be observed in Fig. 4 (b) (in the appendix) and also cause measurable differences in the time to default between the cases with and without transparency. This is coherent with the interpretation that information asymmetries cause higher social costs if individuals have less observations at their disposal to approximate hidden parameters.

All in all, keeping in mind all the limitations and caveats that go along with the methodological approach, we conclude that our model suggests relatively little impact of transparency on diversity, although with a slight tendency towards a positive correlation between transparency and conformity. We did not find supporting evidence for the opposite hypothesis.

4 Discussion

We see this work as an attempt to conceive a formal model of individual behaviour in different regimes of public knowledge about the consequences of personal data exposure. It is far too early to draw relevant conclusions for the real world from such a small model, or to derive policy recommendations. Here it is important to recall that the model compares two 'second best' options, and favourable policies might include elements not captured in the models, say, a combination of transparency, restrictions on personal data processing and a ban of obvious discrimination by personal attributes (i.e., decreasing γ rather than communicating d). This is why we rather see our proposal as a framework to support structured reasoning about social aspects of privacy and transparency, as well as a subject to critique and improvement. Our current list of ideas for model extensions which may be considered, one by one, in further refinements is given below. Where appropriate, we also comment qualitatively on the technical consequences for the model and possible interpretations.

- **Perception bias** In the current model, the maximum-likelihood estimate of d is efficient because the individual has access to a representative sample of the signals emitted in the population. This is unrealistic, as people observe their peers over connections in social networks, where nodes in close proximity tend to share similar preferences. The model could be augmented by an observability rule (technically, a filter on $s^{(j)}, i \neq j$) that reflects these restrictions.
- **Multi-modal preference distribution** In line with the previous point, preferences are most likely not distributed evenly between individuals, but rather in clusters. This is partly due to socialisation between peers, but

[4] with one single exception for $\beta = 1$ and $\gamma = {}^{1}/_{10}$

since social adaptation is not captured in our model, an exogenous multimodal (mixture) distribution for preferences could help to emulate this phenomenon. However, care must be taken to keep the number of parameters tractable, and whether individuals know them or not.

- **Higher-dimensional preference space** The low dimension of the preference space restricts individuals in the choice of trade-offs between their private preferences and their public 'image' (communicated through signals). Possible candidates for higher-dimensional preference spaces are surfaces of k-toruses, hyperspheres (both share the useful property that no discontinuities exists at margins) or, for a discrete case, a binary vector. Higher-dimensional preference spaces have the advantage that the penalty function can be a distance measure in a lower-dimensional projection. Knowledge about which dimensions are relevant (i.e., the coefficients of the projection matrix) could be a distinguishing feature between transparency and obscurity. This not only allows to adjust the degree of transparency more gradually, but might also be deemed as closer to reality where information asymmetries tend to exist on the *selection of attributes* used as discriminating features (e.g., your high-school degree for credit scoring) while the *direction* of influence is more obvious (e.g., better grades, on average, imply better jobs and thus lower credit risk). Higher-dimensional preference spaces also enable the reflection of dependencies between attributes.
- **Stochastic penalty** One problem of our model that might drive the results is the fact that individuals learn the position of the focal point so quickly from incurred cost (for large n, more than 90 % of individuals know d in second round, and with certainty in the third round). Although we try to compensate for this by keeping time to default short, and thus the proportion of uninformed individuals high, it would be desirable to find ways to cut the immediate feedback. One option is to make the penalty discrete and stochastic, so that individuals optimise over expected costs. This would clearly add noise to the observations and impede learning. However, stochastic penalty also complicates the model as lot, in particular since individuals would have to make trade-offs between expected value and volatility. So additional assumptions on risk aversion are needed.
- **Penalty dependent on other individuals' actions** Yet another direction are penalty functions that depend on the individuals' behaviour *relative to others* (e.g., the cost is born by the q individuals closest to d). Such a penalty function could mirror the dynamics of social norms, which have empirically been found to affect individuals' cost to emit certain signals (perceived abnormality implies higher cost [15]). This seemingly simple change has tremendous implications, as the setting would become a non-zero-sum game *between* individuals. So optimal strategies will need observations of others not anymore just to compensate information asymmetries, but also to anticipate the (re-)actions of others. So making the penalty function dependent on others seems difficult and might not be a good idea unless the focus of study is strategic competition between individuals.

- **Endogenous penalty function** Related to the previous point, one could also consider to make the penalty function dependent on the (aggregate) signals. This would reflect the property that the 'punishing entity' is part of the society and formed by it in more or less institutionalised ways, such as democratic decisions, populism in policy-making, or public uproar and revolution. Strictly speaking, endogenous penalty functions imply that the model turns into a game between individuals (see above). However, if n is large, one can make the common assumption that individuals are 'price takers' to justify that strategic interactions are disregarded.
- **Behavioural features** Finally, the rationality assumption could be weakened by allowing for well-understood behavioural phenomena that are deemed relevant for perceptions (and following action) in the area of privacy and transparency, for instance through hyperbolic discounting of uncertain costs in the future [16].

We would like to stress that this list of options is not very specific to the research question studied in this paper, but applies to more general aspects of modelling the distribution of personal information in a society. An overview of literature that addresses topics at the intersection between privacy and technology with a similar methodology, though in a more or less formal manner, is given in the next section.

5 Related work

Social implications of permanent data traces have been studied by Friedman and Resnick ('social cost of cheap pseudonyms') [17], Blanchette and Johnson ('forgetfulness') [18] and lately also by Mayer-Schönberger [19]. Odlyzko [20] has added that costs of a lack of privacy can also materialise in supplier rents through better possibilities for price discrimination. A broader survey on the economics of privacy has been compiled by Hui and Png [21].

Transparency as a remedy to personal information abuses as received little attention so far. TETs in conjunction with PETs can seen as enabling tools for Jiang et al.'s [22] *principle of minimum asymmetry.* This principle has been developed in the broader context of privacy issues in ambient intelligence. It is based on the assumption that information asymmetries between two parties, data owner and data controller, negatively impacts the data owner in making an informed decisions about disclosure of personal information to the data controller. This is so because the data owner is uncertain about negative externalities arising from the re-use of his or her personal by third parties that collude with the data controller. These externalities correspond to our penalty function, and the logic that technology cures negative externalities indirectly (via reducing information asymmetries) is compatible with our model. The solutions proposed in this framework differ from our model in several ways. First, user-controllable data avoidance (i.e. PETs) are considered by Jiang et al. as complementary technologies, but do not appear in our model (data avoidance is fixed to constantly hide the private preferences). Instead, we allow the individuals to alter their signal (personal attributes), though it comes at a cost. Finally, Jiang et al.'s framework includes the concept of prevention by deterrence: technology supports mechanisms to detect data abuse and a legal framework ensures that malicious data controllers are held accountable. This channel has no corresponding element in our model. Also Brin's (pointed and admittedly unrealistic) concept of a transparent society [23] can also be seen as a spiritual forerunner of our work, however without leaving individuals the choice to emit a different signal than their endowed preference (that is not private any longer).

One of the key ingredients of modelling privacy-related behaviour on the individual level is the assumption of heterogeneous attributes between individuals (our model does this by means of preferences). While this design decision is quite obvious – otherwise, if all individuals were identical, hiding attributes shared with all others is not very meaningful – researchers disagree in whether the attitude towards privacy should itself be modelled as heterogeneous (e.g. in [24, 25]) or not (for example [26]). Clearly, empirical evidence suggests the existence of different stereotypes, such as privacy fundamentalists as well as pragmatists [27]. However, adhering to the *lex parsimoniae* (parsimony principle), one may consider to omit this detail. Dodds [28] approaches this important question with evolutionary theory and proposes a model in which heterogeneous privacy concerns are more stable than homogeneity, although the exact transition paths depend on a number of (arbitrarily chosen) parameters. Note that privacy concerns are heterogeneous in our model as well. They follow implicitly from rational behaviour given heterogeneous preferences.

6 Summary and outlook

Our research was motivated by the debate about appropriate tools and technologies to assist people in dealing with their personal information in a world where storage and processing of data becomes ever cheaper. We have argued that the data avoidance approach pursued by advocates of so-called privacy-enhancing technologies (PET) is impractical and unrealistic in many situations, so that transparency-enhancing technologies (TET) are seen as a promising alternative. This led us to the research question, how transparency on the consequences of disclosure of particular personal attributes affects macro-social properties, such as diversity and conformity. We have proposed a micro-economic model of rational agents adjusting their data disclosure under various constraints, and presented solutions for the optimal individual strategy in either case. Simulation results tentatively suggest that transparency in fact fosters conformity, although the effects we found are rather weak. Beyond this particular result, we see the main contribution of this paper in the model proposal and the reflections on possible extensions, which may serve as a starting point for more complete (or more parsimonious) models, which one day may be augmented by a measurement part to be fit to empirical data.

Acknowledgements

We thank Thomas Gloe for stating the proposition “If TETs ever provide full transparency, then everybody will behave the same” in an informal conversation, which led us to the research question. The final model is a result of fruitful discussion with Mike Bergmann, Stefan Berthold, Matthias Kirchner, Stefan Köpsell, and Immanuel Scholz. As usual, the responsibility for errors and omissions remains our own.

This work was supported in part by the EU Network of Excellence “Future of Identity in the Information Society” (FIDIS, `http://www.fidis.net`). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies.

References

1. Acquisti, A., Varian, H.R.: Conditioning prices on purchase history. *Marketing Science* **24** (2005) 1–15 `http://www.heinz.cmu.edu/~acquisti/papers/privacy.pdf`.
2. Calzolari, G., Pavan, A.: On the optimality of privacy in sequential contracting. *Journal of Economic Theory* **30** (2005) 168–204
3. Kim, E., Lee, B., Zhu, K.: CRM and the incentive to share customer information. Workshop on Information Systems and Economics (WISE) (2005)
4. Baran, P.: Communications, computers and people. Technical report, RAND Corporation, Santa Monica, CA (1965)
5. Cranor, L.F.: P3P : Making privacy policies more useful. *IEEE Security & Privacy* **1** (2003) 50–55
6. Ferraiolo, D., Kuhn, D.: Role based access control. In: *Proc. 15th National Computer Security Conference.* (1992) `http://csrc.nist.gov/rbac/ferraiolo-kuhn-92.pdf`.
7. Goldberg, I., Wagner, D., Brewer, E.: Privacy-enhancing technologies for the internet. In: *Proc. of 42nd IEEE Spring COMPCON.* (1997) 103–109
8. Camp, L.J., Osorio, C.: Privacy enhancing technologies for internet commerce. In: *Trust in the Network Economy.* Springer-Verlag, Berlin (2003)
9. Adams, C.: A classification for privacy techniques. *University of Ottawa Law & Technology Journal* **3** (2006) 35–52
10. Borcea-Pfitzmann, K., Hansen, M., Liesebach, K., Pfitzmann, A., Steinbrecher, S.: Managing one’s identities in organisational and social settings. *Datenschutz und Datensicherheit [Data Protection and Data Security]* **31** (2007) 671–675 `http://dud.inf.tu-dresden.de/literatur/DuD_2007-09.pdf`.
11. Brandt, F.: Fully private auctions in a constant number of rounds. In Wright, R.N., ed.: *Proc. of Financial Cryptography.* Volume LNCS 2742. (2003) 223–238 Revised version of February 2004 available at `http://www.tcs.ifi.lmu.de/~brandtf/papers/fc2003.pdf`.
12. Hildebrandt, M.: Profiling into the future: An assessment of profiling technologies in the context of Ambient Intelligence. FIDIS In-house Journal (2007) Online available at `http://journal.fidis.net/fileadmin/journal/issues/1-2007/Profiling_into_the_future.pdf`.
13. Bellotti, V., Sellen, A.: Design for privacy in ubiquitous computing environments. In: *ECSCW’93: Proc. of European Conference on Computer-Supported Cooperative Work*, Norwell, MA, Kluwer Academic Publishers (1993) 77–92

14. Lundblad, N.: Privacy in a noise society. WHOLES Workshop: A Multiple View of Individual Privacy in a Networked World (2004) http://www.sics.se/privacy/wholes2004/papers/lundblad.pdf.
15. Huberman, B.A., Adar, E., Fine, L.R.: Valuating privacy. *IEEE Security & Privacy* **3** (2005) 22–25
16. Acquisti, A., Grossklags, J.: Losses, gains, and hyperbolic discounting: An experimental approach to personal information security attitudes and behavior. In: *The Economics of Information Security.* Kluwer (2004) http://www.heinz.cmu.edu/~acquisti/papers/acquisti_grossklags_eis_refs.pdf.
17. Friedman, E., Resnick, P.: The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy* **10** (2001) 173–199
18. Blanchette, J.F., Johnson, D.G.: Data retention and the panoptic society: The social benefits of forgetfulness. *Information Society* **18** (2002) 33–45
19. Mayer-Schönberger, V.: Useful void: The art of forgetting in the age of ubiquitous computing. KSG Faculty Research Working Paper RWP07-022 (2007) http://ksgnotes1.harvard.edu/Research/wpaper.nsf/rwp/RWP07-022.
20. Odlyzko, A.: Privacy, economics, and price discrimination on the internet. In Sadeh, N., ed.: *ICEC2003: Fifth International Conference on Electronic Commerce.* (2003) 355–366 http://www.dtc.umn.edu/~odlyzko/doc/privacy.economics.pdf.
21. Hui, K.L., Png, I.P.: The economics of privacy. In Hendershott, T.J., ed.: *Handbooks in Information System and Economics.* Volume 1. Elsevier (2006)
22. Jiang, X., Hong, J.I., Landay, J.A.: Approximate information flows: Socially-based modeling of privacy in ubiquitous computing. In: *UbiComp '02: Proc. of Ubiquitous Computing.* Volume LNCS 2498., Berlin, Springer-Verlag (2002) 176–193
23. Brin, D.: *The Transparent Society.* Perseus Books, Reading, MA (1998)
24. Böhme, R., Koble, S.: On the viability of privacy-enhancing technologies in a self-regulated business-to-consumer market: Will privacy remain a luxury good? *Workshop on the Economics of Information Security (WEIS)*, Carnegie Mellon University, Pittsburgh, PA (2007) http://www.inf.tu-dresden.de/~rb21/publications/BK2007_PET_Viability_WEIS.pdf.
25. Chellappa, R., Shivendu, S.: An economic model of privacy: A property rights approach to regulatory choices for online personalization. *Journal of Management Information Systems* **24** (2007) 193–225
26. Bouckaert, J., Degryse, H.: Opt in versus opt out: A free-entry analysis of privacy policies. *Workshop on the Economics of Information Security (WEIS)*, Robinson College, University of Cambridge, UK (2006) http://weis2006.econinfosec.org/docs/34.pdf.
27. Kumaraguru, P., Cranor, L.F.: Privacy indexes: A survey of Westin's studies. Tech. Report CMU-ISRI-5-138, Carnegie Mellon University (2005)
28. Dodds, S.: Hiding, seeking, and the evolution of privacy behaviour. *Workshop on the Economics of Information Security (WEIS)*, Carnegie Mellon University, Pittsburgh, PA (2007) http://www.econ.yorku.ca/seminars/2006-2007/Stefan_Dodds.pdf.

Appendix

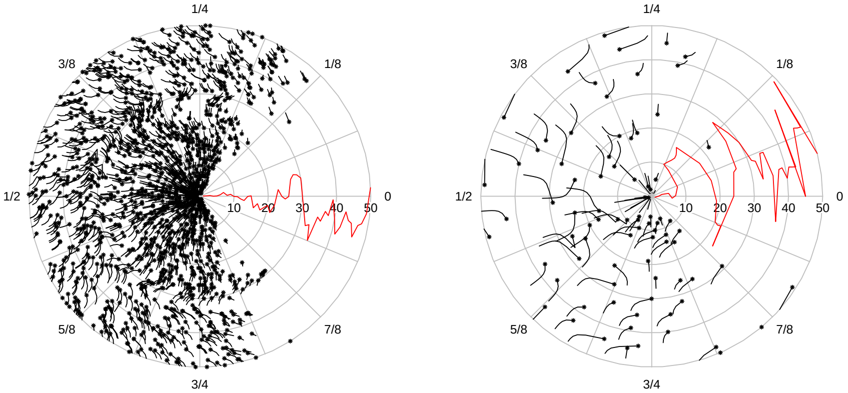
List of symbols

α	weight of pretence component in total cost
β	weight of discontinuity component in total cost
γ	weight of penalty component in total cost
ν	cost offset per round (technical constant to prevent convergence)
ψ	measure of conformity (variable of interest)
c_{disc}	discontinuity cost ($\propto$ distance to previous signals)
c_{pen}	penalty cost ($\propto$ neg. distance between signal s and focal point d)
c_{pret}	pretence cost ($\propto$ distance between preference p and signal s)
$c_{\text{tot}}^{(i)}$	total cost of i-th individual
d	location of vocal point (max. penalty); transparency $\Rightarrow d$ is known
$\hat{d}_k^{(i)}$	estimated for d formed by the i-th individual in round k
D	distance function in preference/signal space
f_{s_1}	probability distribution of 'young' individuals' signals
f_{s_k}	probability distribution of all individuals' signals
i	index for individual
$I^{(i)}$	individual (agent in the model) with index i
j	alternative index for individual
k	round index (as suffix)
K	expected time to default (function over signal/preferece space)
m	number of observable signals from previous round
n	number of individuals in the model
$p^{(i)}$	private preference of i-th individual ($0 \leq p^{(i)} < 1$)
q	quantile among individuals ($0 \leq q \leq n$)
$s^{(i)}$	signal emitted by i-th individual ($0 \leq s^{(i)} < 1$)
$v^{(i)}$	wealth of i-th individual ($v^{(i)} \leq 1$)

Table 2. Performance of the ML estimator for $\hat{d}$ (in % pts.)

	mean absolute error (MAE)				
n	$\gamma = 1/20$	$\gamma = 1/10$	$\gamma = 1/4$	$\gamma = 1/2$	$\gamma = 1$
5	14.38	10.16	6.91	4.33	2.75
10	10.77	7.41	4.85	3.15	1.88
25	8.07	5.88	3.23	2.47	1.55
50	7.30	4.31	2.59	1.61	1.07
100	4.47	3.20	1.91	1.52	0.80
200	3.81	2.46	1.43	0.91	0.61

Metrics computed from 100 runs after 50 iterations ($\alpha = 1, \nu = .1$)



(a) no transparency, $n = 100, \nu = 1/4$

(b) no transparency, $n = 10, \nu = 1/10$

Fig. 4. Supplemental simulation snapshots. $d = 0, \alpha = 1, \beta = 1, \gamma = 1/2$. See caption of Fig. 3 for more details.