

Patel, Amrish; Cartwright, Edward

**Working Paper**

## Social norms and naïve beliefs

Department of Economics Discussion Paper, No. 09,06

**Provided in Cooperation with:**

University of Kent, School of Economics

*Suggested Citation:* Patel, Amrish; Cartwright, Edward (2009) : Social norms and naïve beliefs, Department of Economics Discussion Paper, No. 09,06, University of Kent, Department of Economics, Canterbury

This Version is available at:

<https://hdl.handle.net/10419/50592>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

University of Kent

Department of Economics Discussion Papers

**Social Norms and Naïve Beliefs.**

Amrish Patel and Edward Cartwright

March 2009

KDPE 0906



# Social Norms and Naïve Beliefs

Amrish Patel  
Department of Economics,  
Keynes College,  
University of Kent,  
Canterbury,  
Kent. CT2 7NP. UK.  
ap291@kent.ac.uk

Edward Cartwright  
Department of Economics,  
Keynes College,  
University of Kent,  
Canterbury,  
Kent. CT2 7NP. UK.  
E.J.Cartwright@kent.ac.uk

March 10, 2009

## Abstract

In this paper we analyse the effect that naïve agents (those who take behavior at “face value”) have on the nature of social norms. After reviewing the use of signalling models to model conformity, we argue in favour of modelling naïve inferences in tandem with standard Bayes rational inferences. Naïve agents weaken the existence of social norms and reduce the range of actions that can become social norms.

**Keywords:** Signalling, Conformity, Social Norms, Naïve Beliefs.

**JEL codes:** D82, D83, Z13.

# 1 Introduction

Conformity, the act of changing one's behavior to match that of others, is a commonly observed and important component of behavior. It is hardly surprising, therefore, that economists have joined social scientists in trying to understand why people do conform (e.g. Elster, 1989; Coleman, 1990; Hechter and Opp, 2001; Cialdini and Goldstein, 2004; Young, 2008). Explanations for conformity discussed in the literature are many but in this paper we shall focus on normative conformity.<sup>1</sup> In short, normative conformity occurs when a person conforms to some norm of behavior because deviating from the norm would result in some loss of utility through guilt, loss of reputation etc. (Deutsch and Gerard, 1955; Sugden, 1986; Coleman, 1990; Cialdini and Trost, 1998).<sup>2</sup> Examples of how normative conformity can have important economic consequences include the possibility that concerns about reputation may prevent an employer breaking the custom to give a high-wage (Akerlof, 1980), or fear of social stigma may stop a person free-riding on a social insurance system (Lindbeck, 1997; Lindbeck et al., 2003).<sup>3</sup>

In order to understand normative conformity two questions seem crucial: First, why do people care about reputation or social approval, and, second, why should deviation from a norm lead to a loss of reputation or social approval. The seminal contribution of Bernheim (1994) goes a long way to answering the second question by persuasively demonstrating that normative conformity can be appropriately modelled as a *signalling game*. This approach is appealing because it can *explain* why deviating from a

---

<sup>1</sup>*Informational conformity* occurs if a person imitates others because he believes that their actions signal payoff relevant information (Bikhchandani, Hirshleifer and Welch, 1998; Chamley, 2003). *Convention* occurs when people coordinate on the same action to exploit mutual, positive externalities (Young, 1996, 2001). *Bounded rationality* occurs if a person uses the simple heuristic of imitating others (Hayakawa, 2000).

<sup>2</sup>Evidence for normative influence dates back to Asch's (1955) famous line length experiment but economists have also recognised the important effects of status, esteem and reputation on individual behaviour (Veblen, 1899; Brennan and Pettit, 2004).

<sup>3</sup>See also: Nyborg and Rege (2003) who model how smokers trade-off social approval from non-smokers against the personal costs of refraining to decide on optimal smoking; Rege (2004) who shows that a concern for social approval can explain the lack of free-riding observed in public goods games.

norm (even if only a little bit) could lead to a large drop in payoff. Crucial, however, in any signalling game are the *beliefs* and *inferences* of players. In modelling conformity this issue becomes particularly apparent because it is precisely the beliefs and inferences of players that will determine how non-conformity is viewed by others.

Current modelling of conformity as a signalling game (including Bernheim, 1994) assumes that inferences are formed using Bayes rational updating. Evidence from psychology and experimental economics would suggest, however, that people do not always form rational inferences but are often more ‘naïve’. We suggest that this makes it crucial *to understand the consequences of ‘more naïve’ inferences in signalling models of conformity*. The motivation for this paper is to discuss and highlight this issue while also suggesting that it is an area where economics and psychology can constructively overlap. In order to illustrate the consequences that more naïve inferences can have we shall consider a special case of the model presented in Bernheim (1994) and demonstrate that the presence of naïve inferences: (1) reduces the chances that conformity emerges, and, (2) reduces the set of actions that could potentially be norms. Once the model is understood, neither of these results will probably be particularly surprising, but they do appear meaningful and serve to demonstrate the importance of better understanding the connection between inferences and conformity.

We proceed as follows: In Section 2 we shall explain in more detail how a signalling game can be used to model conformity. In doing so we shall introduce a simple example to be used in the remainder of the paper. In Section 3 we highlight the importance of inferences, and argue that naïve inferences need to be considered. In Section 4 we demonstrate how naïve inferences can impact substantially upon the conformity one observes. In Section 5 we conclude.

## **2 Signalling models of social norms**

Our aim in this section is to briefly explain and motivate how a signalling game can be used to model conformity. A signalling game is composed of

two types of player, *sender* and *receiver*. Senders have some private, non-verifiable and payoff-relevant information, often called the sender's *type*. Examples include generosity (Benabou and Tirole, 2006; Cartwright and Patel, 2008), fairness (Andreoni and Bernheim, 2009; Grossman, 2009), dedication, honesty, innate productivity (Dufwenberg and Lundholm, 2001), discount rate (Posner, 2000), or relative preference for their first child (Bernheim and Severinov, 2003). While sender's type is not observable, the sender does undertake some *action* which is *observed* by the receiver. For example, the sender may donate \$1000 to charity, spend 8 hours at work or keep his promise. Having observed the action, receivers can try to *infer* which type of sender would have undertaken the action. In doing so, they form a *belief* about the sender's type and given these beliefs may want to 'reward' or 'punish' the sender in some way. For example, if the sender donates \$1000 to charity the receiver will try to infer the generosity of the sender and perhaps give more esteem to the sender if he infers him to be generous.<sup>4</sup>

In order to provide a working example we shall consider a special case of a model due to Bernheim (1994) with the interpretation of type 'as willingness to work hard'. A *worker* chooses how many hours  $h$  to work (each day) from the *action set*  $H \in [0, 20]$ . *Hours worked is publicly observable*. The worker has a *type*  $t$  randomly drawn from the *type set*  $T = [0, 20]$  according to the uniform distribution. The worker knows his type but *type is not publicly observable*. A worker's payoff is the sum of intrinsic and esteem utility. The *intrinsic utility* a worker receives if he is type  $t$  and works  $h$  hours is

$$g(h - t) = -(h - t)^2.$$

Type  $t$  is interpreted as the worker's *intrinsic bliss point (IBP)*. Clearly, a type  $t$  worker maximizes his intrinsic utility by working  $h = t$  hours. The *esteem utility* a worker receives is based on 'inferred type'. We shall say more later about what is meant by 'inferred type'. At this point it is sufficient to

---

<sup>4</sup>To make this setting interesting there needs to be genuine ex-ante uncertainty over the senders type and the receiver must care about the type of the sender.

say that a worker inferred to be of type  $b$  will receive esteem

$$e(b) = -(10 - b)^2.$$

A worker of the *ideal type* thus intrinsically prefers to work 10 hours. A worker believed to be of a lower type may receive less esteem because he is, say, considered lazy, while a worker believed to be of a higher type may receive less esteem because he is seen as someone who neglects other commitments like his family. The total payoff of a worker is a weighted sum of intrinsic and esteem utility.

## 2.1 Signalling equilibria

A *signalling equilibrium* consists of actions for senders and beliefs for receivers such that actions are optimal given beliefs, beliefs are optimal given actions and beliefs are updated using Bayes' rule (Fudenberg and Tirole, 1991). The key thing for us to note at this stage is how signalling equilibria can be divided into two types, *separating* and *pooling*. In a *separating equilibrium*, each type of sender chooses a different action. This means that the senders type can be inferred from his action and so it is difficult to think of there being conformity. By contrast, in a *pooling equilibrium* at least *two different types of sender undertake the same action*. This means that type cannot be inferred by action and suggests that there is some conformity. To illustrate it is worth returning to the model and defining signalling equilibrium.

The two ingredients of a signalling equilibrium are an action function and an inference function. An *action function*  $\mu$  maps the set of types to the set of actions. In interpretation  $\mu(t)$  is the number of hours a worker of type  $t$  will choose to work. An *inference function*  $\phi(b, h)$  assigns a probability distribution over types for any choice  $h$ . Informally, we can think of  $\phi(b, h)$  as giving the probability that a worker who works  $h$  hours is inferred to be

type  $b$ .<sup>5</sup> The total payoff of a worker of type  $t$  who chooses  $h$  can be written

$$U(t, h, \phi) = g(h - t) + E \int_T e(b) \phi(b, h) db \quad (1)$$

where  $E > 0$  is the weight the worker puts on esteem. Action function  $\mu$  and inference function  $\phi$  are a *signalling equilibrium* if (i) actions are optimal given inferences, that is,  $U(t, \mu(t), \phi) \geq U(t, h, \phi)$  for all  $h \in H$  and  $t \in T$ , and (ii) inferences  $\phi$  are Bayes rational given action function  $\mu$ . This second criteria is harder to pin down to a simple definition. It does imply that positive probability should be put on a worker of type  $b$  choosing  $h$  if a worker of type  $b$  will actually choose  $h$  (that is,  $\phi(b, h) > 0$  if  $\mu(b) = h$ ) and this probability should be one if only a worker of type  $b$  would choose  $h$  (that is,  $\phi(b, h) = 1$  if  $\mu(b) = h$  and  $\mu(t) \neq h$  for all  $b \neq h$ ).

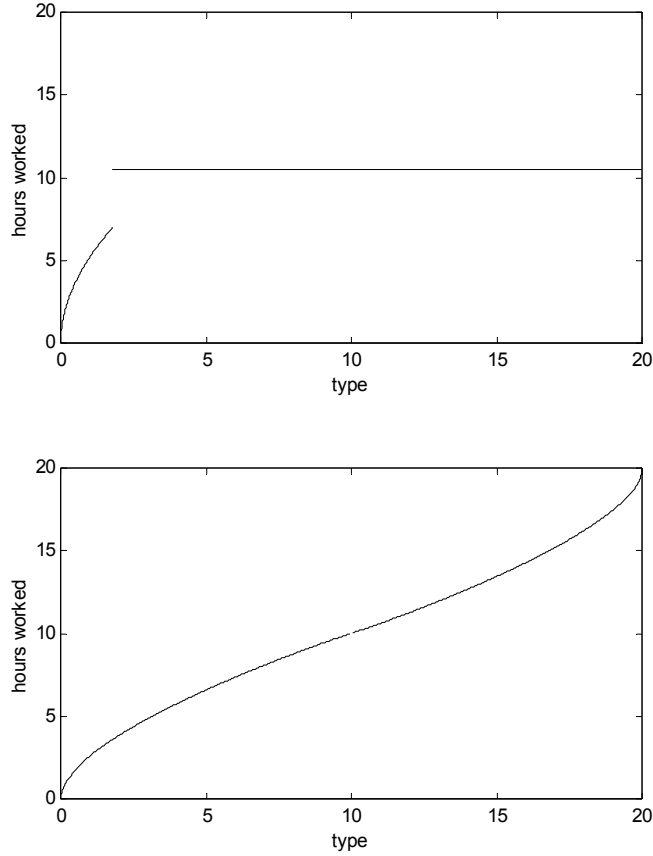
Figure 1 illustrates the nature of signalling equilibria by plotting two different equilibria. The top panel of Figure 1 illustrates a pooling equilibrium and the lower panel a separating equilibrium. In the pooling equilibrium there is a norm of 10.5 hours with any worker of type 2 or above conforming to the norm. Workers of type 2 and below do not conform. In the separating equilibrium hours worked is determined by type and the higher the type the more hours worked. Note, however, that a worker is still influenced by esteem by working more hours than he would intrinsically prefer if of type less than 10 and fewer hours if of type more than 10.

---

<sup>5</sup>It must be the case that,

$$\int_T \phi_l(b, h) db = 1 \quad \text{for all } x \in X.$$





**Figure 1:** The top panel illustrates a signalling equilibrium when  $E = 1.25$  and the norm is  $h = 10.5$ . The bottom panel illustrates a signalling equilibrium when  $E = 0.25$ .

A signalling equilibrium requires Bayes rational inferences about  $\mu$ . This means that in the equilibrium illustrated in the bottom panel of Figure 1 an observer should correctly infer the type of any worker. For example, a type 5 worker will work 6.5 hours but anyone who works 6.5 hours is correctly inferred to be of type 5. By contrast, in the equilibrium illustrated in the top panel an observer is not able to infer the type of anyone who conforms

to the norm of 10.5 hours. Instead he should infer anyone who works 10.5 hours to be of some type between 2 and 20.

More generally, given any pooling signalling equilibrium there exists a unique action  $h_p$  which we shall call the *norm* and many types of worker who choose  $h_p$ . In Figure 1 the norm illustrated is 10.5.<sup>6</sup> Typically there can be many different pooling equilibria each with a different norm. In particular, as in Figure 1, the norm need not correspond to the most desirable type. Henceforth we refer to a pooling signalling equilibrium as a *conformist signalling equilibrium*.

## 2.2 Why model conformity as a signalling game?

Before moving on we briefly ask ‘what insights can a signalling based model bring?’ In many models of normative conformity it is simply assumed that deviating from the norm leads to a loss of reputation (Akerlof 1980). Specifically, the ‘reputation function’ is assumed discontinuous around the norm. While this may yield insight it does effectively engineer conformity rather than explain it. A signalling model does not suffer from this problem because the esteem function can be continuous and yet there still be a discontinuous loss in utility from non-conformity. This discontinuity is produced endogenously as a result of equilibrium behavior rather than the model’s construction (Bernheim, 1994). There are non-signalling models of social conformity that do assume a continuous reputation function (e.g. Azar, 2004). These models suffer from the drawback that they only explain conformity to norms consistent with ‘preferred type’. They are not able to explain the existence of a norm ‘that most people do not like, or would rather change’. Clearly, however, such norms exist and are important. In order to explain such norms a discontinuous loss in utility from non-conformity is required.

---

<sup>6</sup>Unlike most signalling models (e.g. Spence, 1974) our esteem function  $h(b)$ , implies that the perception optimum is not at the boundary of the action set. This implies that we expect both high and low type agents to shade their choices toward the centre, creating central pools. We do not provide complete proofs of the characteristics of pooling equilibria here, Bernheim (1994) formally proves that there is a single central pool in the neighbourhood of unity. A number of other models also contain central pools, for example, Banks (1990) and Bernheim and Serverinov (2003).

The principle advantage, therefore, of a signalling model of conformity is that it can endogenously produce the discontinuous loss in utility from non-conformity that seems necessary to explain much normative conformity. A signalling framework has been found to be informative in understanding a number of specific social norms. For example, Dufwenberg and Lundholm (2001) analyze the effect of unemployment insurance on job-search. They argue that social respect is given to those who exert high search effort relative to their innate talent to find work. Innate talent is private information, motivating the signalling model. They find that more generous unemployment insurance leads to increased effort by untalented individuals in order to signal they are more talented than they actually are, thereby securing social respect. Modelling inheritance, Bernheim and Severinov (2003) argue that bequests are used to signal parents' relative altruistic preferences towards their children. They show that a relatively impartial parent divides the bequest equally among children because not doing so would lead to children feeling they were significantly less loved than they actually were, the social norm of equal division. Finally, more discursive applications of signalling theories explaining social norms are found in Posner (1998, 2000). Attitudes towards symbols, conspicuous consumption, marriage, voting and discrimination can all, at least in part, be explained within a signalling framework.

On the negative side a signalling model does leave important questions unanswered such as 'why reputation matters?' and 'how we come to arrive at a particular norm?' Duxbury (2001) makes a more fundamental point by arguing that actions are not always observable yet there are still social norms for these actions, voting for example. Clearly a valid point. One can, however, point towards the 'big brother' effect, in which a person can think that somebody 'might be watching him' as observed in tax evasion (e.g. Alm et al., 1995, 1999), as well as a desire to signal desirable attributes to oneself (Bodner and Prelec, 2003) to argue that a signalling model can still be relevant. Taking a different track, McAdams (2001) argues that signalling is inefficient and costly and so people will revert to more direct means to build a reputation. Again, there is clearly some merit in this view but it also seems to confuse the nature of a signalling model. In particular, the

argument that people with ‘desirable types’ can somehow signal their type is merely an argument that a separating equilibria exists. If this is the case we should not expect to observe conformity.

A further criticism that can be made of signalling models is that prediction is difficult. Indeterminacy is a problem that plagues signalling models in general, those for social norms are no different. For example, while a pooling equilibrium is usually characterized by a unique norm, there are still a continuum of potential norms making it difficult to identify the precise location of the norm. In addition, the range of indeterminacy (the set of potential norms) varies with the assumptions placed on beliefs. For example, Bernheim (1994) and Dufwenberg and Lundholm (2001) impose different restrictions on inferences, each will produce a different set of potential norms. This, however, may not be a criticism of signalling models per se but just a reflection of the impossibility of predicting the precise location of norms. We frequently observe social norms of behavior that appear entirely arbitrary and that could equally have been some other behavior, the clapping of hands to show appreciation, for example. Indeed, Posner (2000) argues that a costly arbitrary action that would never be undertaken for any reason other than signalling is the most effective signal. We will return to the issue of indeterminacy in Section 4.

### **3 The importance of inferences**

Inferences are fundamental to a signalling model of conformity. A person will conform if he thinks that by not conforming he will send a signal to others that leads to a significant loss in payoff. This means that the person must infer *how others will interpret his actions if he conforms* and infer *how others will interpret his actions if he does not conform*. We shall assume that a person does correctly infer how others will interpret his actions. This then leaves the question of how others will interpret his actions. It should be clear that this is fundamental to understanding whether conformity does or does not arise. Conformity only arises if non-conformity is expected to be inferred by others in such a way as to lead to a relatively large drop in

esteem.

That signalling models of conformity depend so crucially on inferences is a cause for concern for two reasons. First, Bayesian updating can leave open the question of what an worker should infer about someone who does *not* conform to a norm. If, for example, no-one should work 9 hours then Bayes updating is no help in interpreting the type of a worker who does work 9 hours. Clearly, however, this is crucial in knowing whether or not a worker would want to deviate from the norm and work 9 hours. Various criteria on out of equilibrium beliefs can and have been used, such as the D1 Criterion and Intuitive Criterion, to impose ‘reasonable’ assumptions on out of equilibrium beliefs (Cho and Kreps, 1987; Banks and Sobel 1987). Different criteria can produce, however, different equilibria and this is clearly a worry (Cartwright, 2009).

A second cause for concern, and the one we shall focus on here, is that signalling equilibrium implicitly assumes that people use Bayes updating in inferring the types of those who *do* conform, but is this assumption reasonable? We shall argue that often it is not and that the evidence points, instead, towards people making ‘naïve’ inferences about others. Before going further we should argue this point.

### **3.1 Naïve or Bayesian inferences?**

The assumption of Bayes updating implies that observers make full and correct use of information to interpret the actions of a worker. If observers do gain a higher payoff for making correct inferences then anything other than Bayes updating can be improved upon and so this assumption is justifiable. There are, however, good reasons to think that an observer may not use Bayes updating. Furthermore, these reasons point towards observes making ‘naïve’ inferences, in the sense that a worker who works  $h$  hours is simply seen as the type of worker who likes to work  $h$  hours.

A most basic reason why inferences may deviate from rationality is simply the complexity involved in calculating rational inferences. In order to Bayes update the observer needs lots of information about the environment

and choice facing the sender. Clearly this information may not be immediately available to the observer or the choice setting may be a complex one that the observer finds hard to relate to. Signalling experiments show that convergence to equilibrium can be very slow as subjects take time to learn even a relatively simple signalling game (Anderson and Camerer, 2000). Information and familiarity with a problem are important, meaningful context is found to be a good substitute for the early stages of learning (Cooper and Kagel, 2003).

Even if the observer does have relevant and meaningful information he may not update appropriately. A belief in the *law of small numbers*, for example, suggests that an observer may exaggerate the extent to which a small sample represents the population (Tversky and Kahneman, 1971; Rabin, 2002). Similarly, *confirmation bias* would suggest an observer seeks and interprets evidence to be consistent with initial priors, ignoring contradictory information (Rabin and Schrag, 1999; Charness and Levin, 2005). On a more basic level experimental evidence suggests an overconfidence in private information (Huck and Oechssler, 2000; and Nöth and Weber 2003) and an inability to think through the motivations of others in strategic contexts (Thaler, 1988; Eyster and Rabin, 2005).

One reason why inferences may be naïve is the so called *fundamental attribution error* (Jones and Harris, 1967; Ross, 1977). A person makes an attribution error if he attributes the behavior of others to an internal cause, such as preferences, while attributing own behavior to external causes, such as social pressure. This results in someone underestimating how much the decisions of others is effected by social influence (Miller and Prentice, 1994). The literature has now demonstrated widely the existence of a fundamental attribution error even in contexts where observers are well aware that social influence exists (Shweder and Bourne, 1982; Marriott, 1990; Fiske et al., 1998). The main thing for us to note, however, is that an observer who makes an attribution error is likely to make naïve inferences and interpret actions as merely reflecting type.

One possible consequence of an attribution error is *pluralistic ignorance*. This is a psychological state in which a person believes the attitudes and

preferences of others are different to his own, even if his behavior is the same as theirs (Miller and McFarland 1991). A classic example, would be the student, struggling to follow a lecture, who interprets the silence that follows the lecturer asking ‘do you have any questions’ as a signal that he is the only one who is struggling (Miller and McFarland 1987). Pluralistic ignorance can be used to explain why conformity does exist. For example, if all workers are working 10 hours then someone with pluralistic ignorance would think that all workers are of a type who want to work 10 hours, and this may put pressure on him to work 10 hours. Pluralistic ignorance also, however, naturally implies that anyone who deviates from a norm will be inferred as having behaved according to type (Bicchieri, 2006). Again, this points towards naïve inferences.

A further reason for naïve inferences is provided by *trust*. Trust has long been recognized as a cornerstone of all societies and relationships within them (Coleman, 1990; Putnam, 1993; Knack and Keefer, 1997). The moment we begin to question the motivation and actions of someone we trust then that trust can be lost with costly consequences.<sup>7</sup> Trusting individuals are less likely to fully explore the motivations that lay behind others’ actions. This can lead to mistaken inferences about the actions of others but this cost maybe worth paying if it sustains a trusting relationship. If someone’s actions are trusted then the outcome is naïve beliefs. What’s more it provides a reason why naïve beliefs may exist in settings where the observer and sender know each other well.

Having given some reasons why irrationality and naïvety in inferences may be expected, it is important to recognize that not all evidence points that way. Indeed, experimental economic evidence that directly addresses signalling games, suggests there is convergence to equilibrium behavior with sufficient time to learn (see Van Winden (1998) and chapter 8 of Camerer

---

<sup>7</sup>For instance, contrary to agency theory, recent research has shown that in many contexts, monitoring and explicit incentives increase shirking as they signal a lack or breach of trust. Benabou and Tirole (2003) model this result by highlighting the difference between intrinsic and extrinsic motivations; Frey (1993) and Barkema (1995) both show that CEO performance decreases with monitoring; and experimental work by Fehr and Gächter (2002) corroborates the idea.

(2003) for reviews of the literature). Of most relevance to us are studies that shed light on whether we observe pooling when there is a pooling equilibrium and separating when there is a separating equilibrium.<sup>8</sup> The evidence is that we do get behavior consistent with equilibrium if subjects have had time to learn (Cadsby et. al., 1990; Cooper et. al., 1997). Indeed it seems that senders learned to pool or separate as appropriate, and observers learned to distinguish when observers were pooling or separating. Furthermore once subjects have learnt how to play a particular signalling game they can transfer this knowledge to other games (Cooper and Kagel, 2008). The context considered in the literature is very different to the one in this paper but the suggestion is that rational inferences may not be too unrealistic.

The previous paragraph raises another important issue. We have primarily concentrated on the question of whether observers make naïve or Bayes rational inferences. The question that actually matters, however, at least in interpreting the actions of senders, is how senders *expect observers to interpret their actions*. If observers do have naïve inferences, do senders know this? Data from signalling experiments is consistent with the existence of a significant number of "sophisticated" individuals (Cooper and Kagel, 2008). Sophisticated players understand that other players behaviour may be motivated by boundedly rational beliefs, they thus best-respond to this behaviour rather than the behaviour that follows from Bayes rational beliefs.

In summary, there is strong evidence to suggest that observers make errors in inferring the motives of others and are likely to make naïve inferences. There is also, however, evidence that observers can learn over time to correctly interpret actions. This suggests, as a simplification, a *distinction between a set of observers who will make Bayesian inferences and a set who will make naïve inferences*. The former will include those who have more to gain from making an informed decision and have more information about

---

<sup>8</sup>Another strand of the literature, of some relevance to us, looks at the validity of equilibrium refinements based on 'reasonable' assumptions about out of equilibrium beliefs. There is good support for the most basic 'intuitive criterion' but less support for more refined criterion (Brandts and Holt, 1992; Banks, Camerer and Porter, 1994). Historical precedence/accident seems more important in selecting equilibria than the more refined criterion.



the decision being made. The latter will include those with less to gain and who have insufficient information to make an informed decision. This is the approach we follow below.

### 3.2 Modelling naïve inferences

Naïve inferences are unlikely to be ‘rational’ and so allowing for naïvety requires a departure from signalling equilibrium. It seems, however, advantageous to have a model with naïve inferences that still fits the ‘standard equilibrium’ framework of economics. We shall do this by introducing a ‘naïve equilibrium’ where the sender and ‘informed observers’ are rational while ‘uninformed observers’ may be naïve.

The notion that stable equilibria exist with individuals that suffer from systematic belief biases is not new in the literature. For example: Self-Confirming Equilibrium (Fudenberg and Levine, 1993), where beliefs are only correct on the path of play; Cursed Equilibrium (Eyster and Rabin, 2005), where agents underestimate the correlation between actions and private information; Analogy-Based Expectation Equilibrium (Jehiel, 2005), where agents form beliefs for analogy classes of situations rather than each specific situation; and Behavioral Equilibrium (Esponda, 2008), where agents fail to account for the informational content of other agents’ actions. While these equilibria are informative in a number of different settings, they neither directly address signalling games nor consider the "taking actions at face value" interpretation of naïvety motivated here.

To our knowledge, there has been limited research on naïvety in signalling games. Building on their model of Cursed Equilibrium, Eyster and Rabin (2007) introduce the idea that uncursed agents may exaggerate the extent to which others are cursed, that is, they may be "inferentially naïve". Their notion of Credulous Play is similar to our naïve equilibrium. Receivers in Credulous Play do not think that senders choose their actions so as to manipulate beliefs, but in fact, senders are rational and best-respond to naïve inferences. Applying their ideas to a variant of the standard job-market signalling game, they show that the greater the degree of inferential naïvety

the more over-paid workers are. While their model is able to identify some of the effects of naïve beliefs, they are silent on social norms. The infinite action and type spaces mean that although pooling equilibria do exist analysis is difficult as there are an infinite number of them. With closed action and type spaces, our model is more informative on how naïve beliefs affect conformity.

Naïvety could also be important in self-signalling contexts. Bodner and Prelec (2003) consider an agent with some predisposition for committing a vice who infers this predisposition from his actions. The agent's choice determines his outcome utility and diagnostic utility (equivalent to intrinsic and esteem utility). They contrast the case of "face value" interpretations, where the agent ignores the diagnostic motive for actions and assumes all actions are driven by outcome utility, with that of "true" interpretations, where the agent is Bayes' rational. Although the application and details of the model are very different, the similarity between their approach and ours is clear. One significant difference is that our modelling of naïvety is more general in that we always permit the co-existence of naïve and Bayes' rational inferences. In the context of self-signalling it may be difficult to imagine simultaneously having multiple beliefs. For social-signalling, however, it seems important to account for the fact that different observers may have different beliefs.

Bodner and Prelec (2003) shares an important feature with Cartwright and Patel (2008) where we consider naïve inferences in a signalling model of public good provision. Both these papers assume the ideal type is at the boundary of the action set: the less predisposed to a vice you are, the higher your self-esteem; the more generous you are, the higher your social esteem. In the current paper, however, the ideal is some intermediate type. This distinction proves important in terms of the nature of conformity. More specifically, in the framework of Cartwright and Patel or Bodner and Prelec there can be only one action that is the norm, while in the framework of this paper there can be many actions that are potential norms. We shall see this in the following section.

## 4 What difference do naïve inferences make?

To illustrate what consequences naïve inferences can make we shall work through the simple model of conformity introduced in Section 2 and contrast a setting where all observers make Bayes rational inferences with one where some make naïve inferences. More specifically, we split the set of observers into two sets, *informed observers* and *uninformed observers*. As the name may suggest, informed observers will be assumed to make rational inferences. We shall then contrast the case where uninformed observers make naïve inferences to that where they make rational inferences.

Letting  $\phi_I$  and  $\phi_U$  denote the inference functions of informed and uninformed observers, the payoff of a worker of type  $t$  who chooses  $h$  can be written

$$U(t, h, \phi_I, \phi_U) = g(h - t) + \lambda \int_T e(b) \phi_I(b, h) db + \theta \int_T e(b) \phi_U(b, h) db, \quad (2)$$

where real numbers  $\lambda, \theta > 0$  are the weights on esteem from informed and uninformed observers, both relative to intrinsic utility. To compare this to equation (1) we can think of  $\lambda + \theta$  as equal to  $E$ .<sup>9</sup> If both informed and uninformed observers make rational inferences we will obtain a signalling equilibrium, as previously defined. In this context a *signalling equilibrium* is characterized by an action function,  $\mu$ , and inference functions,  $\phi_I$  and  $\phi_U$ , such that (i) actions are optimal given inferences, and (ii) inferences  $\phi_I$  and  $\phi_U$  are Bayes rational. A signalling equilibrium of the model is thus characterized by *all observers* having Bayes rational inferences meaning that  $\phi_I = \phi_U$ .

---

<sup>9</sup>More formally, equation (2) can be rewritten

$$U(t, h, \phi_I, \phi_U) = g(h - t) + (\lambda + \theta) \int_T e(b) \phi(b, x) db$$

where

$$\phi(b, x) = \frac{\lambda \phi_I(b, x) + \theta \phi_U(b, x)}{\lambda + \theta}.$$

Note that  $\int_T \phi(b, x) db = 1$  and so  $\phi$  is a valid inference function.

We say that the uninformed observers have *naïve inferences* if,

$$\phi_U(b, h) = \begin{cases} 1 & \text{if } b = h \\ 0 & \text{otherwise} \end{cases} .$$

In other words, if a worker works  $h$  hours, an uninformed observer will naïvely infer that the worker's type is equal to  $h$ . The fact that naïve observers make naïve inferences means we are unlikely to obtain a signalling equilibrium (because the naïve observers will have incorrect beliefs). We, therefore, consider a refined notion of equilibrium. A *naïve (signalling) equilibrium* is characterized by an action function  $\mu$ , and inference functions,  $\phi_I$  and  $\phi_U$ , such that (i) actions are optimal given inferences, (ii) inferences  $\phi_I$  are Bayes rational, and (iii) inferences  $\phi_U$  are naïve.<sup>10</sup> Naïve equilibria can also be distinguished between separating and pooling equilibria and again we shall call a pooling naïve equilibria a *conformist naïve equilibrium*.

#### 4.1 The Effect of Naïve Inferences

Within the model we are considering there will always exist a signalling equilibrium and naïve signalling equilibrium. We have noted, however, the important distinction between a *separating* equilibrium, where there is no conformity, and a *pooling*, or *conformist*, equilibrium, where there is conformity. The issues of interest to us, therefore, is under what conditions there exists a conformist equilibrium, that is, an equilibrium characterized by a norm to which multiple types of workers will conform. Given our focus on comparing rational versus naïve inferences, this gives rise to the question: *‘Does there exist a conformist naïve equilibrium whenever there exists a conformist signalling equilibrium, and vice versa?’* The answer is simple:

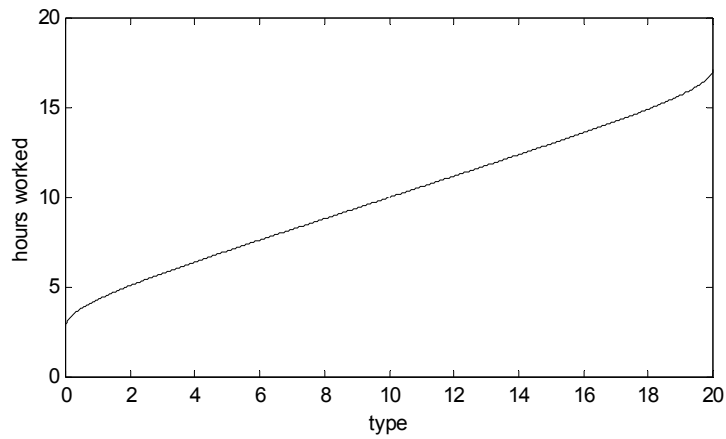
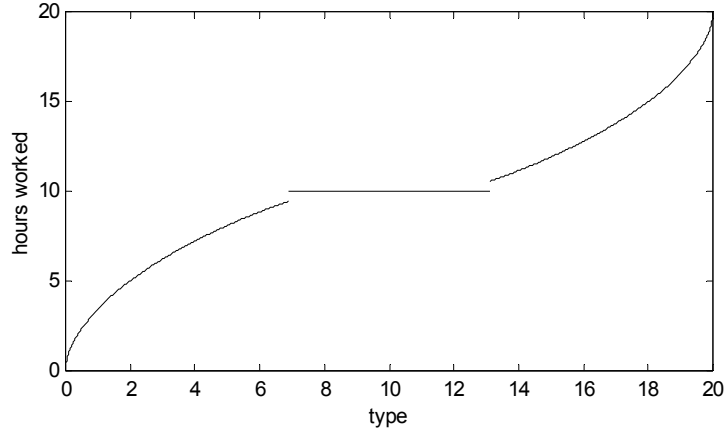
**Proposition:** If there exists a conformist naïve equilibrium with norm  $h_p$  then there exists a conformist signalling equilibrium with norm  $h_p$ . The converse need not be true.

---

<sup>10</sup>Note that while the uninformed are considered naïve the informed are considered rational. This means that the informed understands that the uninformed are naïve and that the worker knows this.

In short, naïve inferences make it less likely for there to be conformity. To better understand what this means, and why it is the case, we decompose the problem into two parts. First, given a set of parameters we can ask whether there exists a conformist signalling equilibrium and/or conformist naïve equilibrium (for any norm). This basically amounts to questioning whether there exists a conformist equilibrium with a norm to work 10 hours (the hours worked by the ideal type). Second, fixing a specific action (different to 10 hours) as the possible the norm, we can ask whether there exists a conformist signalling equilibrium and/or conformist naïve equilibrium with this action as the norm. This amounts to asking what actions can be norms in equilibrium.

The question of when there exists a conformist equilibrium, with norm to work 10 hours, is relatively simple to answer (see the Appendix for the details). There exists a conformist signalling equilibrium if and only if  $\lambda + \theta > 0.25$  and there exists a conformist naïve equilibrium if and only if  $\lambda(1 + \theta) > 0.25$ . We see, therefore, that whenever there exists a conformist naïve equilibrium there must exist a conformist signalling equilibrium. Indeed, there can only exist a conformist naïve equilibrium if  $\lambda$  is sufficiently high meaning that the worker must value the esteem of informed workers. To illustrate, and explain why, Figure 2 plots the conformist signalling equilibrium and naïve separating equilibrium when  $\lambda = 0.1$  and  $\theta = 0.4$ .



**Figure 2:** Setting  $\lambda = 0.1$  and  $\theta = 0.4$ , the top panel illustrates a conformist signalling equilibrium with norm  $h_p = 10$  while the bottom panel illustrates a naïve separating equilibrium.

In the top panel of Figure 2 we see that when uninformed observers are rational a worker of type 7 to 13 conforms to a norm of working 10 hours. They do so, because deviating from this norm would lead to a sufficiently large *discontinuous* drop in the esteem that they would receive. This discontinuity arises because any worker who works less (or more) than 10 hours

is inferred to *not* have a type between 7 and 13. Thus, anyone who works 10 hours gets significantly more esteem than anyone who does not work 10 hours (however close they may be to working 10 hours).<sup>11</sup>

Now, suppose that uninformed observers make naïve inferences. Given that a naïve observer equates action with type any deviations from working 10 hours result in a *continuous* change in esteem. A worker who works 9 hours 59 minutes, for example, would get approximately the same esteem as someone who conforms and works 10 hours. The presence of rational informed observers would still suggest a discontinuous drop in esteem for deviating from the norm but this drop would now be much less. This may result in insufficient incentive to conform, and this is what we observe in the bottom panel of Figure 2.

In short, the fact that uninformed observers have naïve inferences makes it ‘easier’ for a worker to deviate from a norm and not lose ‘too much’ esteem.<sup>12</sup> This acts to break down conformity and result in a separating equilibrium. It is clear that, provided there is concavity of the esteem and intrinsic utility functions this will always be the case. If the norm is 10 hours, for example, then by working 9 hours 59 minutes, rather than 10 hours, a worker who prefers to work less than 10 hours, will gain more in intrinsic utility than he will lose in esteem from any naïve observers. Naïve inferences thus undermine the reasons to conform.

Naïve inferences undermine conformity even more if the norm is something other than to work 10 hours. To provide an illustration of this consider Figure 3. For varying degrees of weight on the esteem of the uninformed observers, Figure 3 plots the maximum possible hours that can be a norm in equilibrium.<sup>13</sup> We see, for example, that when the uninformed have rational

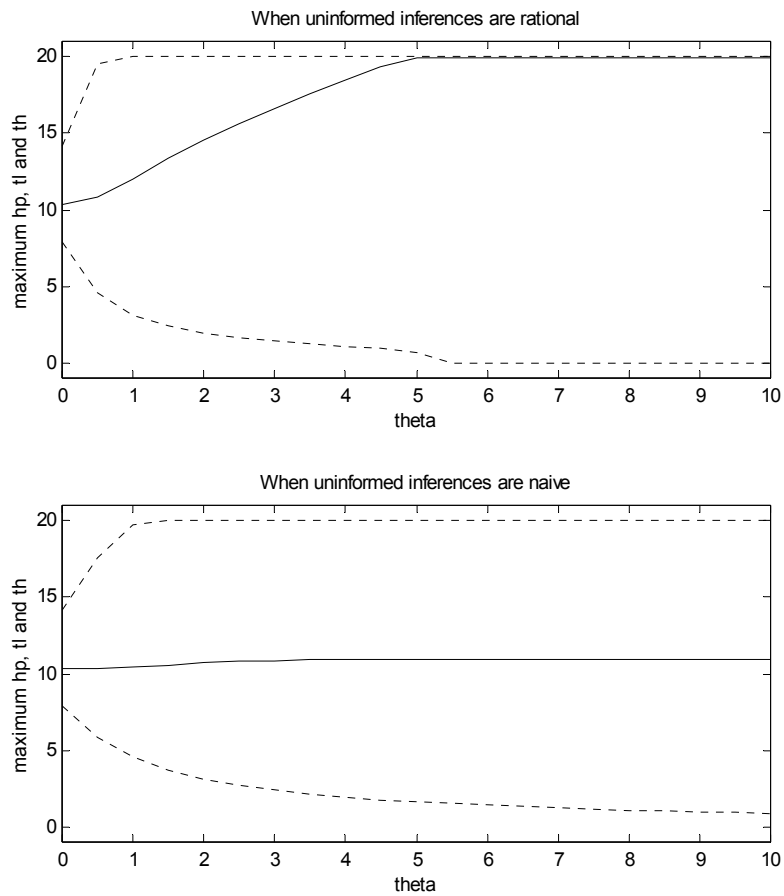
---

<sup>11</sup>In a setting of a continuous action space like this it may be questioned whether such a discontinuity in esteem seems realistic. Is it plausible, for instance, that someone who works 9 hours and 59 minutes gets less esteem than someone who works 10 hours? We think the answer can be yes when there is a norm to work 10 hours. More specifically, our anecdotal evidence of some workplaces is that ‘clock watching’ can happen with it noted that someone left *before*, say, 5 p.m., even if it was only 1 minute before 5 p.m.

<sup>12</sup>The fact that informed observers know this means that they will also not drop esteem so much for deviators from the norm.

<sup>13</sup>If  $h_p^{\max}$  is the maximum possible norm, then any action in  $[20 - h_p^{\max}, h_p^{\max}]$  can also

inferences and  $\theta = 1$  there exists a signalling equilibrium with a norm to work 12 hours, and an agent with any type between 4 and 20 will conform. When the uninformed observers have naïve inferences there cannot be a naïve equilibrium to work 11 hours or more. Clearly, whether uninformed observers are rational or naïve makes a big difference.



**Figure 3:** The maximum possible norm  $h_p$  (and corresponding values of  $t_l$  and  $t_h$ ) for different values of  $\theta$ , when  $\lambda = 0.5$ .

---

be a norm.



The reason why the distinction between naïve and rational inferences proves so significant in determining what actions can be norms is fairly simple. If uninformed observers are rational then the higher is the weight on esteem the more an agent will want to conform to the norm, whatever the norm may happen to be. This is because anyone deviating from the norm will be assumed to have an ‘extreme type’ and so the more the worker values the esteem of observers the more he is willing to sacrifice intrinsic utility to conform. A norm of working 20 hours can, therefore, be maintained if esteem is sufficiently highly weighted.<sup>14</sup> If uninformed observers are naïve then there is much less incentive to conform to a norm different than the ideal of 10 hours. This is because anyone who works 10 hours will receive maximal esteem from the uninformed and so ‘why conform to a norm of working other than 10 hours?’. To explain further, let us return to Figure 3 and imagine that  $\theta = 6$  and the norm is to work 20 hours. If the uninformed are rational then this norm is consistent with equilibrium because anyone who deviated and worked, for example, 10 hours would be inferred to have type  $t_l = 0$  and so receive minimum esteem. If the uninformed are naïve then the norm of 20 hours is not consistent with equilibrium because an agent can deviate to working 10 hours and receive maximal esteem from the uninformed because he is inferred to have the ideal type  $t = 10$ .

The above discussion makes clear, that the set of actions that can be norms when the uninformed are rational is increasing in  $\theta$ . By contrast, the set of actions that can be norms when the uninformed are naïve must be decreasing for sufficiently large  $\theta$ , and ultimately fall to 10. This means that for large  $\theta$  the differences between a setting with rational and naïve uninformed agents must be large. To better appreciate this, Figure 4 plots the maximum norm  $h_p$  for combinations of  $\theta$  and  $\lambda$  if the uninformed are naïve. We can see that for any level of  $\theta$  an increase in  $\lambda$  increases the

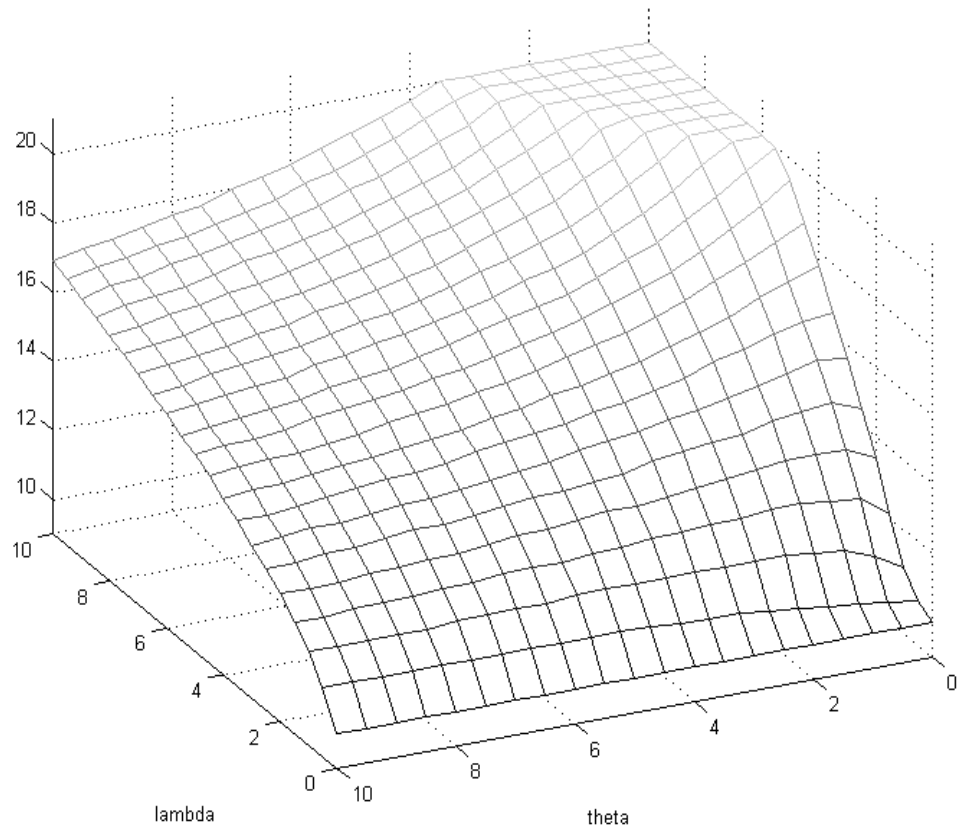
---

<sup>14</sup>To be more explicit, we need to find the level of  $\lambda + \theta$  such that a type 0 agent prefers to work 20 hours and receive relatively more esteem versus working 0 hours and receiving low esteem. This requires,

$$-2^2 - \frac{1}{3}(\lambda + \theta) > -(\lambda + \theta)$$

which gives  $\lambda + \theta > 6$ . Note that this is consistent with Figure 3 where  $t_l = 0$  for  $\lambda = 0.5$  and  $\theta > 5.5$ .

maximum norm, while for any level of  $\lambda$  an increase in  $\theta$  (generally speaking) decreases the maximum norm. What this means is that the set of actions that can potentially be norms is much smaller if the uninformed have naïve inferences. It is as if naïve inferences act as an equilibrium selection device.



**Figure 4:** The maximum possible norm  $h_p$  when the uninformed make naïve inferences, for different values of  $\lambda$  and  $\theta$ .

To summarize we see that naïve inferences have distinct consequences. First, naïve inferences make it less likely that there will be a conformist equilibrium with a norm to work the ‘ideal amount’ of 10 hours. This is

because, in a setting of naïve inferences, unlike that of rational inferences, a worker gains relatively little extra esteem from working 10 hours than, say, 9 hours 59 minutes. Second, naïve inferences make it harder to sustain norms that are not close to the ideal type of 10 hours. This is because, in a setting of naïve inferences, again unlike that of rational inferences, a worker who works 10 hours will necessarily receive relatively high esteem. All together, therefore, naïve inferences make conformity less likely and reduce the set of possible norms.

## 5 Conclusion

The basic points we wished to make in this paper can be summarized: (i) a signalling game is a good, useful way to model normative conformity, but (ii) signalling models can be sensitive to how individuals interpret and infer from the actions of others, which suggests (iii) we need to better understand how individuals do interpret and infer the actions of others. We would further add that (iv) this is an area where economists and psychologists can potentially learn a lot from each other. We hope to have argued point (i) and shown by example point (ii) so in this concluding section we focus more on point (iii) by providing some issues that that we think may be worth pursuing.

The first thing that we can pick up on is how our results fit with the notion of pluralistic ignorance. Specifically, we suggest that naïve inferences narrow the set of actions that can be norms. It is typical, however, to use pluralistic ignorance, which is also a form of naïve beliefs, to argue that many actions can be norms (Bicchieri 2006). At first, this seems contradictory, but it is not. The key issue is whether the ‘ideal type’ is common knowledge or not. Naïve inferences mean that only actions ‘close’ to the *perceived* ideal type can be norms. We have assumed that the ideal type is common knowledge and so perceived ideal type equals the actual ideal type. In the case of pluralistic ignorance, the ideal type is not common knowledge but the action that is the norm is the perceived ideal type. Hence there is no inconsistency between our results and pluralistic ignorance. Instead, it raises

the question of whether the ideal type is known, or at least learnt over time. This, in turn, leads to questions of how norms change over time.

A further issue is whether the relative value a person places on esteem will depend on the inferences of observers. In their analysis of audience effects, Brennan and Pettit (2004) rather cynically (although reasonably) argue that individuals of high ability will try to attract the attention of more informed audiences, whereas those of low ability will do the opposite. More generally, agents of the ‘ideal type’ may come to value more highly the esteem of those with rational inferences while agents of ‘less ideal types’ may value the esteem of those with naïve inferences. If correct this would mean, in terms of our model, that  $\lambda$  and  $\theta$  will also depend on the type of agent. This could have some interesting consequences for equilibrium conformity.

Of interest with respect to both of the issues just raised is the possibility of an “overconfidence” bias in which an agent systematically overestimates his ability relative to others (DellaVigna, 2007). This could result in an agent inferring that the ideal type is more ‘like him’ than it actually is and consequently lead to him seeking a more informed audience. More generally, this raises the question of whether senders (or workers in our model) can be expected to behave rationally or, at least, learn to behave rationally.

Finally, we have focussed on the dichotomy between ‘fully’ rational and ‘fully’ naïve inferences. More realistic models should draw on theories of learning and have a continuous degree of belief sophistication. This could allow observers to become more informed with experience. The effect this could have on social norms is not clear. If a social norm forms when a large proportion of the population is naïve, does this norm change as these agents become more rational? Our results suggest that fewer naïve agents would expand the set of feasible norms, but because of history-dependence the norm may not change. This raises interesting questions about how observers do learn and about how norms can change.

## 6 Deriving signalling equilibria

If uninformed observers are naïve and informed observers perceive the worker to be type  $b$  then the workers utility is

$$U(t, h, b) = -(h - t)^2 - \lambda(10 - b)^2 - \theta(10 - h)^2.$$

Noting that

$$\frac{U(t, h, b)}{100} = -\left(\frac{h}{10} - \frac{t}{10}\right)^2 - \lambda\left(1 - \frac{b}{10}\right)^2 - \theta\left(1 - \frac{h}{10}\right)^2$$

we can normalize so that  $h$  and  $t$  range between 0 and 2. The equilibrium action profile can then be found as described by Bernheim (1994) and Cartwright (2009). Specifically, indifference curves in the  $(h, b)$  plane for a worker of type  $t$  are given by,  $(h - t)^2 + \lambda(1 - b)^2 + \theta(1 - h)^2 = D$ , where  $D$  is an arbitrary constant. We can calculate the slope of an indifference curve of a type  $t$  worker through the point  $(b, x)$  as,

$$\frac{db}{dx} = -\frac{\partial U/\partial x}{\partial U/\partial b} = \frac{(1 + \theta)h - t - \theta}{\lambda(1 - b)}.$$

In equilibrium there must (i) be a tangency between inference function  $\phi_I$  and the indifference curve and (ii) the inferences of rational observers must be correct implying that  $\phi_I(h) = b = t$ . Thus,

$$\phi_I'(h) = \frac{(1 + \theta)h - \phi_I(h) - \theta}{\lambda(1 - \phi_I(h))}. \quad (3)$$

The differential equation (3) can be rewritten as system

$$\begin{bmatrix} dt/dv \\ dh/dv \end{bmatrix} = \begin{bmatrix} -1 & 1 + \theta \\ -\lambda & 0 \end{bmatrix} \begin{bmatrix} t - 1 \\ h - 1 \end{bmatrix},$$

where  $v$  is some index. Rearranging the bottom equation gives

$$t = 1 - \frac{h'}{\lambda} \quad (4)$$

which can be inserted into the top equation to give the second-order differential equation

$$x'' + x' + \lambda(1 + \theta)x = \lambda(1 + \theta). \quad (5)$$

The solution to this differential equation is easily found and so values of  $h$  and  $t$  can be traced out to show the action  $h$  of a type  $t$  worker. From this can be derived appropriate inferences  $\phi_I$  and action function  $\mu$ . Note, however, that we cannot know at this stage whether  $\phi_I$  and  $\mu$  are consistent with equilibrium as, in particular, we may obtain an action  $h > 1$  which is not possible.

The characteristic equation of (5) is  $r^2 + r + \lambda(1 + \theta)$ . This equation has two distinct real roots if  $\lambda(1 + \theta) < 0.25$ , repeated roots if  $\lambda(1 + \theta) = 0.25$  and two distinct complex roots if  $\lambda(1 + \theta) > 0.25$ . All are clearly possible and so we need to distinguish these three cases.

Case (1):  $\lambda(1 + \theta) = 0.25$ . The solution to equation (5) is

$$h = 1 + C_1 e^{-\frac{v}{2}} + C_2 v e^{-\frac{v}{2}} \quad (6)$$

for some constants  $C_1$  and  $C_2$ . To derive appropriate initial conditions consider a worker of type  $t = 0$ . If a type 0 worker is correctly perceived to be of type 0 then his payoff is

$$U(h, 0, 0) = -h^2 - \lambda - \theta(1 - h)^2.$$

Setting  $\frac{du}{dh} = 0$  suggests that  $t = 0$  as  $h = \frac{\theta}{1 + \theta}$ . Appropriate initial conditions are thus  $t = 0$  and  $h = \frac{\theta}{1 + \theta}$  as  $v = 0$ . Using equations (6) and (4) in turn gives

$$C_1 = \frac{-1}{1 + \theta}; \quad C_2 = E\theta - \frac{1}{2(1 + \theta)}. \quad (7)$$

Case (2): If  $\lambda(1 + \theta) < 0.25$ . The solution to equation (5) is

$$h = 1 + C_3 e^{r_1 v} + C_4 e^{r_2 v} \quad (8)$$

where

$$r_1 = -\frac{1 + (1 - 4\lambda(1 + \theta))^{0.5}}{2}; \quad r_2 = -\frac{1 - (1 - 4\lambda(1 + \theta))^{0.5}}{2}$$

and  $C_1$  and  $C_2$  are constants. Appropriate initial conditions remain  $t = 0$  and  $h = \frac{\theta}{1+\theta}$  as  $v = 0$ . So, from (8) we obtain  $C_4 = \frac{-1}{1+\theta} - C_3$  and using (4) we get

$$\lambda = r_1 C_3 + r_2 C_4 = (r_1 - r_2) C_3 - \frac{r_2}{1 + \theta}.$$

Thus,

$$C_3 = -\frac{r_2 + \lambda(1 + \theta)}{(r_2 - r_1)(1 + \theta)} \text{ and } C_4 = \frac{r_1 + \lambda(1 + \theta)}{(r_2 - r_1)(1 + \theta)} - C_3.$$

Case (3):  $\lambda(1 + \theta) > 0.25$ . The characteristic equation of (5) has two distinct complex roots and so the solution of (5) has form,

$$h = 1 + e^{-\frac{1}{2}v} \left( C_5 \cos \frac{mv}{2} + C_6 \sin \frac{mv}{2} \right) \quad (9)$$

where  $m = (4\lambda(1 + \theta) - 1)^{0.5}$  and  $C_5$  and  $C_6$  are constants. Appropriate initial conditions remain  $t = 0$  and  $h = \frac{\theta}{1+\theta}$  as  $v = 0$ . Using equations (9) and (4) in turn gives

$$C_5 = \frac{-1}{1 + \theta}; \quad C_6 = \frac{2\lambda + C_5}{m}.$$

This describes an action function but can give  $h > 1$  implying that a pooling equilibrium will be obtained. In this case the equilibrium will also be characterized by a norm  $h_p$  and types  $t_l$  and  $t_h$ . A worker of type  $t \in [t_l, t_h]$  will work  $h_p$  hours while other workers work according to (9). Types  $t_l$  and  $t_h$  workers will be indifferent between working  $h_p$  hours and working according to (9). There is no simple way (that we know of) to find the values of  $t_l$  and  $t_h$ , but it is simple to check whether any combination of  $t_l, t_h$  and  $h_p$  are consistent with equilibrium. Thus, given  $h_p$  one can search for a combination of  $t_l$  and  $t_h$  that would give an equilibrium. One can then search for possible  $h_p$  that are consistent with equilibrium. ■

## References

- [1] Akerlof, G. A., (1980). "A theory of social custom, of which unemployment may be one consequence," *Quarterly Journal of Economics*, 94: 749-775.
- [2] Alm, J., Sanchez, I. and De Juan, A., (1995). "Economic and noneconomic factors in tax compliance," *Kyklos*, 48(1): 3-18.
- [3] Alm, J., McClelland, G. H. and Schulze, W. D., (1999). "Changing the social norm of tax compliance by voting," *Kyklos*, 52(2): 141-171.
- [4] Anderson, C. M. and Camerer, C. F., (2000). "Experience-weighted attraction learning in sender-receiver signaling games," *Economic Theory*, 16: 689-718.
- [5] Andreoni, J. and Bernheim, D. B., (2009). "Social image and the 50:50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*.
- [6] Asch, S. E., (1955). "Opinions and social pressure," *Scientific American*, 193(5): 31-35.
- [7] Azar, O. H., (2004). "What sustains social norms and how they evolve? The case of tipping," *Journal of Economic Behavior and Organization*, 54: 49-64.
- [8] Banks, J. S., (1990). "A model of electoral competition with incomplete information," *Journal of Economic Theory*, 50: 309-325.
- [9] Banks, J. S., C. Camerer and D. Porter, (1994). "An experimental analysis of Nash refinements in signaling games," *Games and Economic Behavior* 6: 1-31.
- [10] Banks, J. S. and Sobel, J., (1987). "Equilibrium selection in signaling games," *Econometrica*, 55(3): 647-661.



- [11] Barkema, H. G., (1995). "Do top managers work harder when they are monitored," *Kyklos*, 48(1): 19-42.
- [12] Bernheim, D. B., (1994). "A theory of conformity," *Journal of Political Economy*, 102(5): 841-877.
- [13] Bernheim, D. B. and Serverinov, S., (2003). "Bequests as signals: an explanation for the puzzle of equal division," *Journal of Political Economy*, 111(4): 733-764.
- [14] Benabou, R. and Tirole, J., (2003). "Intrinsic and extrinsic motivation," *Review of Economic Studies*, 70(3): 489-520.
- [15] Benabou, R. and Tirole, J., (2006). "Incentives and prosocial behavior," *American Economic Review*, 96(5): 1652-1678.
- [16] Bicchieri, C. (2006). *The Grammar of Society: The nature and dynamics of social norms* Cambridge University Press, Cambridge.
- [17] Bikhchandani, S., Hirshleifer, D. and Welch, I., (1998). "Learning from the behaviour of others: conformity, fads and informational cascades," *Journal of Economic Perspectives*, 12(3): 151-170.
- [18] Bodner, R. and Prelec, D., (2003). "Self-signaling and diagnostic utility in everyday decision making," *The Psychology of Economic Decisions: Vol 1: Rationality and Well-Being*, eds. Brocas, I. and Carrillo, J. D., Oxford University Press, Oxford.
- [19] Brandts, J. and C. Holt, (1992) "An experimental test of equilibrium dominance in signaling games", *American Economic Review* 82: 1350-1365.
- [20] Brennan, G. and Petit, P., (2004). *The Economy of Esteem: An essay on civil and political society*. Oxford University Press, Oxford.
- [21] Cadsby, C., M. Frank and V. Maksimovic, (1990). "Pooling, separating, and semiseparating equilibria in financial markets: some experimental evidence," *Review of Financial Studies* 3: 315-342.

- [22] Camerer, C., (2003). *Behavioral Game Theory: Experiments in strategic interaction* Princeton University Press.
- [23] Cartwright, E. J., (2009). "Conformity and out of equilibrium beliefs," *Journal of Economic Organization and Behavior*.
- [24] Cartwright, E. J. and Patel, A., (2008). "Public goods, social norms and naïve beliefs," *UKC Economics Discussion Paper 08/07*.
- [25] Chamley, C., (2003). *Rational Herds: Economic Models of Social Learning*. Cambridge University Press, Cambridge.
- [26] Charness, G. and Levin, D., (2005). "When optimal choices feel wrong: a laboratory study of Bayesian updating, complexity and affect," *American Economic Review*, 95(4): 1300-309.
- [27] Cho, I. K. and Kreps, D. M., (1987). "Signaling games and stable equilibria," *Quarterly Journal of Economics*: 102, 179-221.
- [28] Cialdini, R. B. and Goldstein, N. J., (2004). "Social influence: compliance and conformity," *Annual Review of Psychology*, 55: 591-621.
- [29] Cialdini, R. B. and Trost, M. R., (1998). "Social influence: social norms, conformity and compliance," *The Handbook of Social Psychology 4th edition*, ed. Gilbert, D. T., Fiske, S. T. and Lindzey, D., 2: 151-192. MacGraw-Hill, Boston.
- [30] Coleman, J., (1990). *Foundations of Social Theory*. Havard University Press, Cambridge, MA.
- [31] Cooper, D., S. Garvin and J. Kagel, (1997). "Signaling and adaptive learning in an entry limit pricing game," *RAND Journal of Economics* 28: 662-683.
- [32] Cooper, D. and J. Kagel, (2003). "The impact of meaningful context on strategic play in signaling games," *Journal of Economic Behavior and Organization*, 50: 311-337.

- [33] Cooper, D. and J. Kagel, (2008). "Learning and transfer in signaling games," *Economic Theory*, 34: 415-439.
- [34] DellaVigna, S., (2007). "Psychology and Economics: Evidence from the Field," *NBER working paper*, 13420.
- [35] Deutsch, M. and Gerard, H. B., (1955). "A study of normative and informational social influences upon individual judgement," *Journal of Abnormal Social Psychology*, 51: 629-636.
- [36] Dufwenberg, M. and Lundholm, M., (2001). "Social norms and moral hazard," *Economic Journal*, 111: 506-525.
- [37] Duxbury, N., (2001). "Signalling and social norms," *Oxford Journal of Legal Studies*, 21(4): 719-736.
- [38] Elster, J., (1989). "Social norms and economic theory," *Journal of Economic Perspectives*, 3(4): 99-117.
- [39] Esponda, I., (2008). "Behavioral equilibrium in economies with adverse selection," *American Economic Review*, 98(4): 1269-1291.
- [40] Eyster, E. and Rabin, M., (2005). "Cursed equilibrium," *Econometrica*, 73(5): 1623-1672.
- [41] Eyster, E. and Rabin, M., (2007). "Notes on inferential naïvety in games," *mimeo*.
- [42] Fehr, E. and Gächter, S., (2002). "Do incentives contracts undermine voluntary cooperation," Institute of Empirical Research in Economics, *University of Zurich*, working paper no. 34.
- [43] Fiske, A. P., Kitayama, S., Markus, H. R. and Nisbett, R. E., (1998). "The cultural matrix of social psychology." In *Handbook of Social Psychology*, D.T. Gilbert, S. T. Fiske and G. Lindzey Eds. New York: McGraw-Hill.
- [44] Fudenberg, D. and Levine, D. K., (1993). "Self-confirming equilibrium," *Econometrica*, 61(3): 523-545.

- [45] Fudenberg, D. and Tirole, J., (1991). *Game Theory*. MIT Press, Cambridge, MA.
- [46] Frey, B.S. (1993). "Does monitoring increase work effort? The rivalry between trust and loyalty," *Economic Inquiry*, 31: 663-670.
- [47] Grossman, Z. (2009). "Self-signaling versus social-signaling in giving," mimeo.
- [48] Hayakawa, H., (2000). "Bounded rationality, social and cultural norms, and interdependence via reference groups," *Journal of Economic Behavior and Organization*, 43: 1-34.
- [49] Hechter, M. and Opp, K. D., (2001). eds. *Social Norms*. Russell Sage Foundation, New York.
- [50] Huck, S. and Oechssler, J., (2000). "Informational cascades in the laboratory: do they occur for the right reasons?" *Journal of Economic Psychology*, 21: 661-671.
- [51] Jehiel, P., (2005). "Analogy-based expectation equilibrium," *Journal of Economic Theory*, 123: 81-104.
- [52] Jones, E. E. and Harris, V. A. (1967). "The attribution of attitudes," *Journal of Experimental Social Psychology*, 3: 1-24.
- [53] Knack, S. and Keefer, P., (1997). "Does social capital have an economic payoff? A cross-country investigation," *Quarterly Journal of Economics*, 112(4): 1251-1288.
- [54] Lindbeck, A., (1997). "Incentives and social norms in household behavior," *American Economic Review*, 87(2): 370-377.
- [55] Lindbeck, A., Nyberg, S. and Weibull, J. W., (2003). "Social norms and welfare state dynamics," *Journal of the European Economic Association*, 1(2-3): 533-542.
- [56] Marriott, M. (1990). *India through Hindu Categories*. Newbury Park, CA: Sage.

- [57] McAdams, R. H., (2001). "Signalling discount rates: law norms and economic methodology," *Yale Law Journal*, 110: 625-689.
- [58] Miller, D. and C. McFarland (1987). "Pluralistic ignorance: When similarity is interpreted as dissimilarity," *Journal of Personality and Social Psychology* 53: 298-305.
- [59] Miller, D. and C. McFarland (1991). "When social comparison goes awry: The case of pluralistic ignorance," in *Social Comparison: Contemporary Theory and Research* J. Suls and T. A. Wills, Eds. Hillsdale, NJ. Erlbaum.
- [60] Miller, D. and D. A. Prentice (1994). "Collective errors and errors about the collective," *Personality and Social Psychology Bulletin* 20: 541-550.
- [61] Nöth, M. and Weber, M., (2003). "Information aggregation with random ordering: cascades and overconfidence," *Economic Journal*, 113: 166-189.
- [62] Nyborg, K. and Rege, M., (2003). "On social norms: the evolution of considerate smoking behavior," *Journal of Economic Behavior and Organization*, 52: 323-340.
- [63] Posner, E. A., (1998). "Symbols, signals and social norms in politics and law," *Journal of Legal Studies*, 27: 765-298.
- [64] Posner, E. A., (2000). *Law and Social Norms*. Harvard University Press, Cambridge, MA.
- [65] Putnam, R., (1993). *Making Democracy Work*. Princeton University Press, Princeton.
- [66] Rabin, M., (2002). "Inference by believers in the law of small numbers," *Quarterly Journal of Economics*, 117(3): 775-816.
- [67] Rabin, M. and Schrag, J. L., (1999). "First impressions matter: a model of confirmatory bias," *Quarterly Journal of Economics*, 114(1): 37-82.

- [68] Rege, M., (2004). "Social norms and private provision of public goods," *Journal of Public Economic Theory*, 6(1): 65-77.
- [69] Ross, L., (1977). "The intuitive psychologist and his shortcomings: Distortions in the attribution process," In *Advances in Experimental Social Psychology* L. Berkowitz, Ed. New York: Academic Press.
- [70] Shweder, R. A. and Bourne, E. J., (1982). "Does the concept of the person vary cross-culturally?" In *Cultural conceptions of mental health and therapy*, A. J. Marsella and G. M. White Ed. New York: Reidel.
- [71] Spence, M., (1974). *Market Signaling*. Harvard University Press, Cambridge, MA.
- [72] Sugden, R., (1986). *The economics of rights, cooperation and welfare*. Basil Blackwell, Oxford.
- [73] Thaler, R. H., (1988). "Anomalies: the winner's curse," *Journal of Economic Perspectives*, 2: 191-202.
- [74] Tversky, A. and Kahneman, D., (1971). "Belief in the law of small numbers," *Psychological Bulletin*, 76(2): 105-110.
- [75] Van Winden, F., (1998). "Experimental studies of signaling games" in L. Luini, Ed. *Uncertain Decisions, Bridging Theory and Experiments*. Boston, Kluwer.
- [76] Veblen, T., (1899). *The theory of the leisure class*. The Viking Press, New York.
- [77] Young, P. H., (1996). "The economics of convention," *Journal of Economic Perspectives*, 10(2): 105-122.
- [78] Young, P. H., (2001). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton, NJ.
- [79] Young, P. H., (2008). "Social norms," *New Palgrave Dictionary of Economics 2nd edition*, eds. Durlauf, S. N. and Blume, L. E.