UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 488

Naïve Beliefs and the Multiplicity of Social Norms

Amrish Patel and Edward Cartwright

February 2011

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

Department of Economics
School of Business, Economics and Law at University of Gothenburg
Vasagatan 1, PO Box 640, SE 405 30 Göteborg, Sweden
+46 31 786 0000, +46 31 786 1326 (fax)
www.handels.gu.se info@handels.gu.se

Naïve Beliefs and the Multiplicity of Social Norms*

Amrish Patel
Department of Economics,
University of Gothenburg,
Box 640, SE 405 30,
Gothenburg, Sweden.
Amrish.Patel@economics.gu.se

Edward Cartwright
Department of Economics,
University of Kent,
Canterbury, Kent,
United Kingdom, CT2 7NP
E.J.Cartwright@kent.ac.uk

February 10, 2011

Abstract

In a signalling model of conformity, we demonstrate that naïve observers, those that take actions at face value, constrain the set of actions that can possibly be social norms. With rational observers many actions can be norms, but with naïve observers only actions close to that preferred by the ideal type can be norms. We suggest, therefore, that the naïvety or inexperience of observers is an important determinant of norms and how they evolve.

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

JEL codes: D82, D83, Z13.

*We would like to thank the referee for some very constructive and useful suggestions on how to improve an earlier version of the paper.

1 Introduction

Many have argued that conformity implies a *multiplicity* of possible social norms. For example, Akerlof (1980, p. 751) writes 'such multiplicity is the essence of social custom: it is inherent in the adage "When in Rome, do as the Romans do"'. In models of conformity it is not uncommon, however, to see a *unique* equilibrium that is consistent with only one possible norm at any one time (e.g. Chamley 1999, Lindbeck et al. 1999, Azar 2008 and Cartwright 2009). In each case the norm is determined by preferences and/or available information. For instance, norms which are intrinsically costly, but followed so as to avoid social disapproval, may be unstable and gradually erode over time (Azar 2004).

So, how indeterminate are social norms? This question is relatively unexplored in the literature but fundamental in understanding social norms and how they change over time. Intuition might suggest that the indeterminacy of norms will depend on salient features of the environment, and our objective in this paper is to provide a specific, and we think very important, example of that. The example we shall use is that of naïve observers. We shall show that if some observers are naïve then the more the esteem of such naïve observers is valued the less indeterminate are norms. This suggests that the naïvety of observers is an important ingredient in determining social norms and how they evolve.

We say that an observer is naïve if he takes actions at face value and so does not try to second guess why an agent may have chosen a particular action. Related notions of naïvety in signalling games are considered by Bodner and Prelec (2003), Eyster and Rabin (2007) and Cartwright and Patel (2010). Inexperience, the fundamental attribution error and trust are three of many reasons why an observer may be naïve. For example, someone with less experience with which to judge why an agent may choose a particular action may be naïve while someone with more experience of observing others may more rationally infer incentives from actions. This distinction is suggested up by the experimental literature on signalling games where subjects typically begin somewhat naïve but become 'more rational' with experience (Camerer 2003 and Cooper and Kagel 2008).

We show that if some observers are naïve then the set of actions that could be possible norms is reduced, and potentially much reduced, compared to if they are rational. The intuition behind this is that a naïve observer looks at what *action the agent undertakes* and ignores whether or not he is conforming. This is in contrast to the rational observer who considers *whether the agent conforms* and ignores what action he chooses. This difference means that when observers are rational anything can be a norm, but when they are naïve only actions sufficiently close to the 'ideal action' can be norms. Between the extremes of everyone being naïve and everyone being rational, norms become more and more indeterminate.

To give an example, consider a town with a norm against tipping at restaurants. A rational observer dining with someone who did not tip would not read too much into the lack of tip because this is the norm. By contrast, a naïve observer who saw someone not tipping might think this person ungenerous. With sufficiently many naïve observers the norm against tipping could not be sustained. Similarly, consider a dress code norm of wearing jeans and a t-shirt to work on Fridays. A naïve observer might be surprised by the lack of shirt and tie and draw a negative inference from that. The norm would thus be hard to sustain in companies that regularly have important visitors on a Friday.[1]

We proceed as follows: Section 2 introduces a signalling model of conformity, section 3 demonstrates the effects of naïvety, and we conclude in section 4.

2 Model

We shall use the seminal model of Bernheim (1994).[2] An agent chooses an action x from set $X = [0, 2]$ and a set of observers observe the action chosen. The *type* t of the agent is randomly drawn from the set $T = [0, 2]$ according to a continuous probability distribution function f. The agent knows his type but observers do not.

The payoff of the agent is a weighted sum of intrinsic and esteem utility. The *intrinsic utility* he receives if he is type t and chooses x is given by $g(x - t)$ where $g(\cdot)$ is a strictly concave, symmetric, twice continuously differentiable function that achieves a maximum at 0. Thus, if of type t he maximizes his intrinsic utility by choosing $x = t$. The *esteem utility* he receives is based on 'inferred type'. If inferred to be of type b he receives esteem $h(b)$ where $h(\cdot)$ is a strictly concave, symmetric, twice continuously

[1] Just telling the visitor that Friday is dress down day, or that the norm in this town is to not tip, may not be enough. The person might still draw their negative inferences because 'first impressions matter' and so on.

[2] Interestingly, Bernheim (1994) does not emphasise the multiplicity of norms that results from his model, instead focussing on the characteristics of equilibria. Only at the end of the conclusion (p. 865) does the multiplicity of norms really get mentioned. In Cartwright and Patel (2010) we show that naïve inferences change the characteristics of equilibria quantitatively but not qualitatively.

differentiable function that achieves a maximum at 1. The *ideal type* thus intrinsically prefers to choose 1.

An *action function* μ maps the set of types to the set of actions. In interpretation $\mu(t)$ is the choice made by the agent if of type t. An *inference function* $\phi(b,x)$ assigns a probability distribution over types for any choice x. Informally, think of $\phi(b,x)$ as the probability the agent is inferred to be type b if he chooses x. We split the set of observers into two subsets: *naïve* and *rational*, where ϕ_N and ϕ_R denote their respective inference functions. The payoff of the agent if he chooses x and is of type t is

$$U(t,x,\phi_N,\phi_R) = g(x-t) + E\left(\theta \int_T h(b)\,\phi_N(b,x)\,db + (1-\theta)\int_T h(b)\,\phi_R(b,x)\,db\right)$$

where real numbers $E > 0$ and $0 \le \theta \le 1$ are the weights, respectively, on esteem relative to intrinsic utility and on naïve observer esteem relative to rational observer esteem.

We shall return shortly to the interpretation of θ. Before that we should clarify what we mean by naïve and rational. We shall assume that observers with *naïve inferences* infer the agent must be of type t if he chooses action $x = t$, that is, they take things at face value. This is naïve because the agent may have an incentive to choose action $x = t$ when not of type t. A rational observer would take into account such incentives. We capture this distinction in our definition of a naïve equilibrium.

A *naïve (signalling) equilibrium* consists of an action function μ and inference functions ϕ_N and ϕ_R, such that (i) actions are optimal given inferences, (ii) inferences ϕ_R are Bayes rational[3] and satisfy the D1 Criterion[4] given action function μ, and (iii) inferences ϕ_N are naïve, that is satisfy

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

If $\theta = 0$, and so weight is only given to the esteem of rational observers, a naïve signalling equilibrium is equivalent to the standard definition of

[3] 'Bayes rational' is hard to pin down to a simple definition. But, for instance, it does imply that positive probability should be put on the agent being type b when choosing x if $\mu(b) = x$. Also, this probability should be one if $\mu(b') \neq x$ for all $b' \neq b$.

[4] The D1 Criterion (Cho and Kreps, 1987) imposes assumptions on out of equilibrium beliefs. Informally, it requires for any x where there does not exist t such that $\mu(t) = x$ that $\phi_R(b,x) > 0$ if and only if there does not exist some type b' such that b' has a greater incentive to deviate from $\mu(b')$ than type b has to deviate from $\mu(b)$.

signalling equilibrium as used, for example, by Bernheim (1994). If $\theta > 0$ then a naïve equilibrium will typically not be a signalling equilibrium. In particular, if the agent has an incentive to choose an action $x \neq t$ when of some type t then a naïve equilibrium is not a signalling equilibrium. Others who consider equilibria where individuals suffer systematic belief biases include Eyster and Rabin (2005) and Jehiel (2005).

Bernheim demonstrates that any signalling equilibrium must have the property that $\mu(t) = x_p$ for all $t \in [t_l, t_h]$ and some $x_p \in X$ and $t_h \geq 1 \geq t_l$. In interpretation, the agent conforms to a *norm* x_p when of type $t \in [t_l, t_h]$. There always exists a signalling equilibrium with norm $x_p = 1$ but may or may not exist a signalling equilibrium with norm $x_p \neq 1$.[5] Showing the same applies for naïve equilibria is simple. Of primary interest to us is *what actions, other than* 1*, can be norms in a naïve equilibrium*?

In order to answer this question we shall vary θ, the relative weight put on the esteem of naïve observers, from 0 to 1 and track changes in the set of equilibria. This clearly makes θ of crucial importance. The simplest interpretation of θ is as the proportion, or perceived proportion, of naïve observers in the population.[6] This interpretation implies that the agent values the esteem of all observers equally. The agent may, however, value the esteem of some observers more heavily than that of others. This should be reflected in θ. For example, the diner in a town with a norm against tipping may value more the esteem of the rational observers with whom he lives than the naïve observers who are passing through town. In this case θ would be lower than suggested by the proportion of locals in the restaurant. By contrast, the worker in a firm with a norm of dressing down on Fridays may value more the esteem of the naïve observers visiting the firm than the rational observers with whom he works. In this case θ would be higher than suggested by the proportion of visitors. Taking things one stage further, it may be that the agent only values the esteem of one other observer. The value of θ should now be interpreted as the probability that the observer is naïve. The interpretation of θ will depend, therefore, a lot on context. A positive value of θ does, however, always imply two basic things: the agent

[5] We are including here the special case of a separating signalling equilibrium where $t_l = t_h = 1$ and $x_p = 1$. The distinction between separating and pooling equilibria is not crucial for our purposes. Note, however, that if $x_p \neq 1$ then $t_l < t_h$ and so there is pooling.

[6] Perceptions are all that really matter here as they are what will determine the agent's incentives. The perceived proportion need not equate with the actual proportion because of limited or non-random past interaction with observers and psychological biases in memory and proportion estimation.

thinks that observers may be naïve and that he values, to some extent, the esteem of such naïve observers.

3 The Effect of Naïve Inferences

The basic point we wish to make in this paper can be made most clearly with the following simple but very stark result. We show that when a lot of weight is put on esteem: *if the esteem of rational observers is highly weighted any action can be a norm*, but *if the esteem of naïve observers is highly weighted only actions arbitrarily close to that preferred by the ideal type can be a norm.*

Proposition 1: For any action $x \in X$ there exists a naïve equilibrium with norm x for θ sufficiently small and E sufficiently large. For any arbitrarily small $\varepsilon > 0$ and any action $x \notin [1-\varepsilon, 1+\varepsilon]$ there does not exist a naïve equilibrium with norm x for θ sufficiently large and E sufficiently large.

Proof: We shall start by proving the first part of the proposition. To do so consider a possible norm of $x_p \neq 1$ and suppose that $t_l = 0$ and $t_h = 2$. We shall check whether the agent has an incentive to deviate from the norm. If the agent does deviate he receives esteem $h(0)$ from any rational observers for any $x \neq x_p$ that he chooses, because of the D1 Criterion (see Bernheim 1994 or Cartwright 2009). The esteem that he receives from any naïve observer is at most $h(1)$ and his intrinsic utility at most $g(0)$. We know, therefore, that his payoff from conforming is greater than that from deviating if

$$g(x_p - t) + E\left(\theta h(x_p) + (1-\theta)\int_0^2 f(b)h(b)db\right) > g(0) + E\left(\theta h(1) + (1-\theta)h(0)\right).$$

Rearranging this becomes

$$E(1-\theta)\left(\int_0^2 f(b)h(b)db - h(0)\right) > g(0) - g(x_p - t) + E\theta\left(h(1) - h(x_p)\right). \quad (1)$$

This must be satisfied for E sufficiently large and θ sufficiently close to 0 since $\int_0^2 f(b)h(b)db - h(0) > 0$. Thus there exists a naïve equilibrium with a norm of x_p for E sufficiently large and θ sufficiently close to 0.

We shall now prove the second part of the proposition. Again, consider a possible norm of $x_p \neq 1$ with some t_l and t_h. We shall check whether a type

1 agent has an incentive to deviate to $x = 1$. Without loss of generality[7], assume $x_p > 1$. An agent deviating to 1 is inferred by a rational observer to be type t_l (because of the D1 Criterion). The agent, therefore, has an incentive to deviate if,

$$g(0)+E\left(\theta h(1)+(1-\theta)\,h(t_l)\right) > g(x_p-1)+E\left(\theta h(x_p)+(1-\theta)\int_{t_l}^{t_h} f(b)h(b)db\right).$$

Since $h\,(t_l) \geq h\,(0)$ and $\int_{t_l}^{t_h} f(b)h(b)db \leq h\,(1)$, a sufficient but not necessary condition for the agent to deviate is then

$$g(0)+E\left(\theta h(1)+(1-\theta)\,h\,(0)\right) > g(x_p-1)+E\left(\theta h(x_p)+(1-\theta)\,h\,(1)\right), \quad (2)$$

or,

$$g\,(0)-g\,(x_p-1)+E\theta\,(h\,(1)-h\,(x_p)) > E\,(1-\theta)\,(h\,(1)-h\,(0))\,. \quad (3)$$

Fixing x_p, the condition is satisfied if E and θ are sufficiently large since $g(0) > g(x_p-1)$ and $h\,(1) > h\,(x_p)$. If an agent of type 1 deviates then there is no conformist naïve equilibrium (see Bernheim (1994), Theorem 3, page 856). To illustrate why, consider a candidate equilibrium with $x_p, t_l, t_h > 1$. If a type t_l agent deviates to an action slightly lower than x_p, i.e. closer to the action choice of a type 1 agent he potentially could incur a small decrease in his intrinsic utility, but would receive an increase in esteem from naïve observers and a discrete increase in esteem from rational observers because he is no longer pooling with less desirable types. Type t_l therefore has an incentive to deviate and the potential conformist naïve equilibrium unravels. There does not, therefore exist a naïve equilibrium if E and θ are sufficiently large.■

This result shows the big contrast between the norms that can exist with naïve as opposed to rational inferences. Naïve inferences clearly and significantly constrain the set of norms that can be equilibria. To obtain a result as clear and stark as this we needed to focus on high values of E, meaning situations where esteem is given a relatively large weight. We shall now complement this by showing that the result naturally extends to smaller values of E. Obtaining general results in this case is a much more complicated task and so we shall focus on an interesting special case where[8]

$$g\,(x-t) = -\,(x-t)^2\,; \qquad h\,(b) = -\,(1-b)^2\,; \qquad f(b) = 0.5.$$

[7] We could equally assume $x_p < 1$ and use an out-of-equilibrium inference of t_h.

[8] We have considered a generalisation of this in which a general quadratic formula is assumed for $g\,(\cdot)$ and $h\,(\cdot)$. The conclusion remains the same, but summarising those results becomes cumbersome given the increase in parameters.

Applying (1) we see that in this case there exists a conformist signalling equilibrium with any norm if

$$E \geq \frac{6}{1-\frac{5}{2}\theta}.$$

Applying (3) we see that there does not exist a naïve equilibrium with norm $x_p \neq 1$ if[9]

$$x_p \geq 1+\sqrt{\frac{E\,(1-\theta)}{1+E\theta}} \quad \text{or} \quad x_p \leq 1-\sqrt{\frac{E\,(1-\theta)}{1+E\theta}}.$$

Analysis of the two boundaries of x_p are analogous, thus we focus on only one, the maximum possible norm.

To illustrate what happens for smaller values of E we need to solve for the equilibrium action function. In Cartwright and Patel (2010) we derive the equilibrium action function $\mu(t)$ for those types where the agent does not conform. It remains to determine for what types the agent will conform and what the norm can be. There is no simple analytical way (that we know of) to find the values of t_l and t_h for any potential norm x_p. We can, however, check whether a particular combination of t_l, t_h and x_p is consistent with equilibrium by checking whether types t_l and t_h are indifferent between separating and conforming, that is,

$$\begin{aligned} & -(t_f-\mu(t))^2 - E\left(\theta(1-\mu(t))^2 + (1-\theta)\,(1-t_f)^2\right) \\ = \quad & -(t_f-x_p)^2 - E\left(\theta(1-x_p)^2 + (1-\theta)\int_{t_l}^{t_h}\frac{1}{2}(1-b)^2 db\right) \end{aligned}$$

for $t_f = t_l, t_h$. Thus, given a potential value for x_p one can search for t_l and t_h and see whether there exists a conformist equilibrium with norm x_p.

Figure 1 shows how the maximum possible norm varies with θ for five different values of E. We see that the maximum is lower the more weight is put on the esteem of naïve observers. Figure 2 plots the maximal possible norm as E varies for five different values of θ. The more the agent values esteem, the higher is the maximum norm.

[9] The lower constraint to x_p cannot be found directly from (3) as we assumed $x_p > 1$ when deriving (3). As noted in footnote 7, we could have equally assumed $x_p < 1$; the resulting condition gives the lower constraint to x_p.

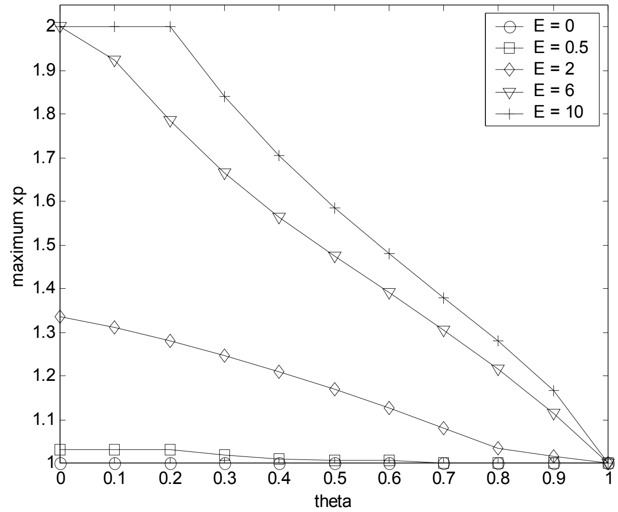

Figure 1: The maximum norm for different values of θ when $E = 0, 0.5, 2, 6$ and 10.

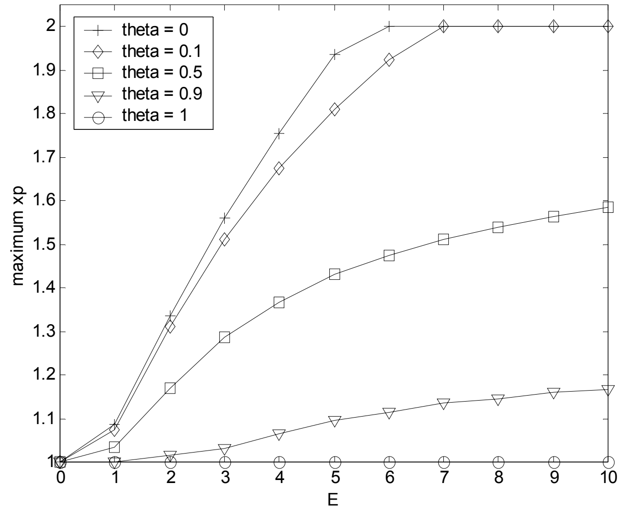

Figure 2: The maximum norm for different values of E when $\theta = 0, 0.1, 0.5, 0.9$ and 1.

The clear message from Proposition 1 and Figures 1 and 2 is that fewer actions can be norms if observers are naïve. More surprising is the relative importance of rational observer esteem relative to naïve observer esteem. For the most part, a small increase in the weight on esteem increases the maximum norm significantly if θ is small, but only increases the maximum norm a little if θ is large. This, however, can be explained by the concavity of the esteem function. This concavity means that the agent does not lose much if naïve observers infer him to be of, say, type 1.1 rather than 1. By contrast, the esteem the agent gets from rational observers can differ a lot if he chooses 1.1 rather than 1 if it means he has not conformed. Increasing the weight on the esteem of rational observers can, therefore, have big consequences. Put another way, it only takes a little bit of weight on the esteem of rational observers to obtain a large set of norms.

The crux behind these results is whether observers take into account the norm. A rational observer *does* acknowledge the existence of a norm and so judges esteem relative to *whether the agent conforms*. A rational observer gives significantly less esteem to the agent if he does not conform because such deviation is seen to signal an 'extreme type'. This provides the incentive to conform and is independent of what the norm is and how close it is to the action preferred by the ideal type. By contrast, a naïve observer *does not* take into account the existence of a norm and so judges esteem relative to *whether the agent chooses the action preferred by the ideal type*. An agent who conforms to a norm far from the action preferred by the ideal type gets *less* esteem from a naïve observer than someone who does not conform but chooses the action preferred by the ideal type. It immediately follows that the more naïve are observers, the more difficult any norm other than the preferred action of the ideal type is to sustain.

4 Conclusion

We began this paper by raising the open and intriguing question of how indeterminate social norms typically are. The suggestion of this paper is that norms may not be as indeterminate as seems typically perceived, particularly if there does exist an 'ideal type'. The extent of any indeterminacy will clearly, however, depend on context, and what we have argued here is that the presence of naïve observers is one important feature of context. We have shown that if observers may be naïve, then the more the esteem of naïve observers is valued the fewer possible social norms there are. The existence of naïve observers essentially breaks the self-reinforcing process whereby an action is the norm because everyone knows it is the norm. Naïve observers do not take account of the norm and judge on the basis of what the agent does. This means that only those actions close to that preferred by the ideal type can be sustained.

This explanation motivates the issue of naïvety as important and relevant in understanding the norms that can possibly emerge from conformity. In particular, the more naïve observers there are, or the more the opinions of naïve observers are esteemed, the fewer actions we would expect could be norms. We obtain, therefore, a testable prediction. We also obtain a prediction of how norms might evolve over time. In particular an increased presence of inexperienced or naïve observers may lead to changes in the norm. The interesting thing here is that this does not follow because the newcomers have different preferences, but because they are unaccustomed to the norm. Workplace norms or fairness norms may, for example, change with an influx of newcomers even if these newcomers share the same preferences as those they join or replace.

One final issue we shall highlight is the potential trade-off an agent faces between doing something that is esteemed by naïve observers and doing something that is esteemed by rational observers. Unless the norm is the action preferred by the ideal type the agent will face this trade-off. If he conforms he gets less esteem from the naïve and if he does not conform he gets less esteem from the rational. So, the agent has to make a choice who's esteem is valued more. The worker, for example, may have to trade the esteem of an informed manager with naïve co-workers (or a naïve manager with informed co-workers), the diner the esteem of an informed friend with a naïve visitor, or the amateur musician the esteem of naïve friends with informed professionals. This trade-off is reflected in the value of θ. That means we can infer something about θ from actions that are norms. In particular, if the norm is distant from the action preferred by the ideal type then it suggests θ is relatively low and that the agent weights highly the esteem of rational observers. Otherwise such a norm could not be sustained. We can, therefore, potentially infer something interesting about preferences by comparing the norm with the preferred action of the ideal type.

References

[1] Akerlof, G. A. (1990) "A theory of social custom, of which unemployment may be one consequence," *Quarterly Journal of Economics*, 94: 749-775.

[2] Azar, O.H. (2008) 'Evolution of social norms with heterogeneous preferences: A general model and an application to the academic review

process,' *Journal of Economic Behavior and Organization* 65: 420-435.

[3] 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.

[4] Bernheim, D. B. (1994) "A theory of conformity," *Journal of Political Economy*, 102(5): 841-877.

[5] 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.

[6] Camerer, C. (2003) *Behavioral Game Theory: Experiments in strategic interaction*, Princeton University Press.

[7] Cartwright, E. (2009). "Conformity and out of equilibrium beliefs," *Journal of Economic Organization and Behavior* 70: 164-185.

[8] Cartwright, E. and Patel, A. (2010) "Public goods, social norms and naïve beliefs," *Journal of Public Economic Theory*, 12(2): 199-223.

[9] Chamley, C. (1999) 'Coordinating regime switches', *The Quarterly Journal of Economics* 114: 869-905.

[10] Cho, I.-K. and Kreps, D. M. (1987) "Signaling games and stable equilibria," *The Quarterly Journal of Economics*, 102: 179-222.

[11] Cooper, D. and J. Kagel, (2008) "Learning and transfer in signaling games," *Economic Theory*, 34: 415-439.

[12] Eyster, E. and Rabin, M. (2005) "Cursed equilibrium," *Econometrica*, 73(5): 1623-1672.

[13] Eyster, E. and Rabin, M. (2007) "Notes on inferential naïvety in games," *working paper.*

[14] Jehiel, P. (2005) "Analogy-based expectation equilibrium," *Journal of Economic Theory*, 123: 81-104.

[15] Lindbeck, A., S. Nyberg and J. Weibull (1999) 'Social norms and Economic Incentives in the Welfare State,' *The Quarterly Journal of Economics*, 114: 1-35.