# Calvin's Reformation in Geneva: Self and Social Signalling

Gilat Levy and Ronny Razin
London School of Economics

**Abstract**

As Weber (1904) recognized, Calvinistic beliefs about predestination may constitute a powerful incentive for good works; an individual wishes to receive assurances about her future prospects of salvation, and good works may provide a positive signal about such prospects. These beliefs can in turn create a social pressure to behave well, as good works can also signal to others that individuals belong to the "elect"and are therefore likely to behave well in social interactions. Moreover, the Consistory, an institution created by Calvin to monitor and publicize individuals' behaviour, can allow for such social signalling. We analyze these self and social signalling incentives, and show how religions affect levels of cooperation and coordination.

## 1 Introduction

Calvin's reformation in Geneva, and from there in other city-states in Switzerland and communities in France, has been recognized by some as an instrumental and practical model of religious organisation.[1] The main innovation in Calvin's reformation was to couple his version of the doctrines of predestination and sanctification with the creation of a real-world institution that governs social behaviour, the Consistory. In this paper we analyse a model of such a religious organization.

The doctrine of predestination implies that an individual's destiny in the afterlife, does not depend on his good deeds in this life. Both Martin Luther and Calvin intertwine this with sanctification and justification. Sanctification, according to both, implies that an individual who belongs

[1] MacCulloch (2010) argues that Calvin's achievements "led Protestantism out of stagnation in the 1550's" and had influenced religious practice to this day (pp. 632).

to the elect inhibits the traits of Jesus Christ and therefore follows his examples of good works, both in his private and social life. Justification, according to Calvin, implies that an individual may be in the elect and hence salvaged only upon the grace of God, as *"For all man are not created in equal condition rather eternal life is foreordained for some, eternal damnation for others" (Institutes, III. xxi 5).*[2]

These doctrines, as Weber (1904) first suggested, provide a strong incentive for good works through self signalling: Individuals might look at their own behaviour to be assured that they are part of the few who will be saved. Weber (1904) writes: "*The question: Am I one of the elect? must sooner or later have arisen for every believer and have forced all other interests into the background. And how can I be sure of this state of grace?".* Calvin has therefore made good works and moral behaviour the centre of religious life by creating a sense of anxiety over individuals' salvation.

But Calvin went beyond pure theology. In the *Ecclesiastical Ordinances* (1541), Calvin drew up the structure of his well-ordered church in Geneva. The most distinctive and controversial aspect of this organization was the Consistory.[3] It was formed in 1542, "*their office is to have oversight of the life of everyone...There were to be twelve of them, chosen from the members of the three councils, to keep an eye on everybody*".[4] To be able to do so, the Consistory members met once a week and visited the homes of each Genevan family twice a year, which allowed them to be well informed. Deviant behaviour -religious and civil alike- was punished by public scolding, sometimes by Calvin himself. When other communities in Switzerland and France decided to adopt Calvin's religion, he insisted on the formation of local Consistories, which may be better at collecting local information.[5]

In this paper we analyse a model of a Calvinistic religion which creates incentives for self and social signalling in the social arena. We compare the welfare effects of these signalling incentives. In particular, We consider a population whose members are randomly matched to play a Prisoners' Dilemma for two periods.[6] Individuals have different

[2] Luther in contrast suggested justification by faith, in which faith alone determines the destiny of the individual. This assures salvation to those who believe in Christ.

[3] At the time, institutions called Consistories already existed in Europe and in Geneva, mainly enforcing the canon marriage laws.

[4] *Ecclesiastical Ordinances* (1541), in Gilbert (1998).

[5] Calvin's emphasis on discipline is evident in his insistence that discipline is the third mark of a good Church (this was opposed by Lutherans) and is certainly a mark of his own reign in Geneva.

[6] More generally we can use any social interaction in which trust is important. A recent literature has analysed cooperation when players sustain (possibly different)

convictions about their prospects of belonging to the elect which are private information. We assume that individuals derive positive utility from holding such convictions. Good works, i.e., cooperation in the game, allow an individual to self signal that he is more likely to belong to the elect.[7] Individuals' behaviours in the first period might be publicly observed, depending on the level of dissemination of public information.

The model implies a *spiritual* as well as a *material* benefit from cooperative behaviour. When engaged in self signalling, individuals that cooperate in equilibrium obtain a spiritual utility as they increase their beliefs about salvation. When information about past behaviour is available, social signalling arises, and cooperation becomes a signal of religious conviction which is rewarded by more cooperative behaviour from others in future interactions. This implies a material benefit as well as an enhanced spiritual benefit as individuals are induced to cooperate even more. Some individuals will be tempted to cooperate first in order to take advantage of others, so that such social signalling may not be perfect. However, we show that the benefits from social signalling can outweigh such losses and individuals in society benefit from more public information.

This paper contributes to the economics literature on religious organisations. Iannaccone (1992, 1998) and Berman (2000) analyse models of religions in which religion is a club good and costly rituals serve to resolve free rider problems in the production of religious goods. Levy and Razin (2011a) analyse a model in which costly rituals serve to signal religious beliefs and intention of cooperative behaviour.

In Benabou and Tirole (2006), religious individuals have an incentive to maintain their beliefs about private actions which facilitate hard work.[8] Glaeser and Glendon (1998) propose that the Protestant Ethic induces individuals to shift their behaviour towards actions that are more visible and conducive for signalling. These two papers accord with Weber's (1904) Protestant Work Ethic, which proposes that self signalling of good works has evolved to attach value to hard work or personal success and in particular to entrepreneurial and risky behaviour. In contrast, we focus on the theological interpretation of good works as behaviour in the social arena. We therefore consider the "social ethic" induced by the Reformation and not the "work ethic".[9] In Scheve and Stasavage (2006),

norms of cooperation, see for example Dixit (2003), Tabellini (2008), and Andreoni and Samuelson (2006).

[7]For a discussion of self-signalling see Ainslie (1992). See Bodner and Prelec (2003) and Bénabou and Tirole (2004) for other formal models of self signalling.

[8]See also Benabou and Tirole (2011).

[9]For an empirical study of the Protestant "social ethic", see Arrunada (2010).

religious beliefs allow for a psychic benefit in bad times, and hence such beliefs negatively correlate with preferences for social insurance.

Our analysis shows how religious organizations orchestrate behaviour in the social sphere.[10] This view of the role religious organizations in consistent with recent empirical literature looking at the effect of religion on economic outcomes; see for example Barro and McCleary (2003) and Guiso *et al* (2003).

Section 2 below presents the model of self signalling, and Section 3 allows for social signalling. A welfare comparison among the two is in Section 4. An appendix contains all proofs.

## 2 The Model

The social interaction game consists of two periods. In each period, individuals are randomly paired to play a Prisoner's Dilemma (PD) game:

| | C | D |
|---|---|---|
| C | $d$,$d$ | $c$,$b$ |
| D | $b$,$c$ | $a$,$a$ |

where $b > d > a > c$. We assume strategic complements, i.e., that $d - b > c - a$. This assumption is standard in the literature on cooperation. Our choice for the social interaction game is affected by recent empirical evidence showing that religious beliefs or practice affect levels of trust in society, which can be captured in the PD environment.[11] For welfare comparisons, assume that society always values cooperation so that $2a < b + c$ and hence mutual defection is the worst outcome from a social welfare point of view.

We assume that each individual $i$ believes that he belongs to the elect with some probability, initially $\alpha_i \in [0, 1]$. Belonging to the elect provides some spiritual utility $\varepsilon$. We sometimes refer to $\alpha_i$ as the individual's conviction and often suppress the subscript $i$. We assume that the individual's conviction is private information. Let the initial types in the population be distributed uniformly on $[0, 1]$. The heterogeneity of beliefs in the population represents different reactions to exposure to theological material, such as reading the Bible or the publications of Calvin.

In Calvin's theology, Sanctification implies that individuals who are in the elect behave as Jesus Christ would. Accordingly, in our model, to hold and maintain their beliefs individuals look back to their previous social behaviour to infer the likelihood of belonging to the elect.

[10] Wilson (2002) and Stark (1996) provide numerous examples in which religions provide material benefits. These and others are discussed in Levy and Razin (2011a).

[11] The model can easily be extended to other types of public good games.

We adopt the following simple, reduced-form, model of self signalling. We assume that individuals believe that those who belong to the elect always cooperate, whereas all other individuals cooperate only with some probability, $\kappa \in [0,1]$. We will shortly discuss how $\kappa$ is determined in equilibrium but for now let us assume that it is given. Note that the assumption that a person who belongs to the elect always cooperates is a simplification and what is important is that his probability of cooperation is higher than a "normal" person.

Self signalling then amounts to an individual revisiting her previous actions and noting whether she had cooperated or not. If an individual with beliefs $\alpha$ cooperated in the first period her posterior beliefs are that she belongs to the elect with probability $\frac{\alpha}{\alpha+(1-\alpha)\kappa}$. On the other hand, if she defects she will believe that she does not belong to the elect with certainty.[12] In accordance with Calvin's writings, we assume that agents believe that the population share of types who do belong to the Elect is close to zero.[13]

We now discuss how $\kappa$ is determined. To this end, we adopt the concept of Cursed Equilibrium due to Eyster and Rabin (2005). According to that concept, an individual may fail to understand that the strategy of a player may depend on his type, and consider instead the average strategy in the population (which effectively is the average in the population of players, i.e., "normal" types).[14] We therefore assume that an individual who updates his posterior probability about belonging to the elect correctly conjectures the average probability of cooperation among all types in the population, which we assign to be $\kappa$.[15]

The level $\kappa$ represents how individuals perceive "normal" behaviour in the population and hence $1-\kappa$ can capture the strength of self signalling. A high level of $\kappa$ corresponds to a high discipline as individuals

[12] A possible interpretation of this updating rule is that it actually occurs within a dynasty by an individual's offspring, so that parents take actions that allow their children to learn about their family type. Another way to conceptualize this is to think of a model similar to Benabou and Tirole (2004) in which individuals may receive signals about their past behaviour but may choose to manipulate these signals.

[13] This is not necessary for the analysis, in fact, it does not enter the updating process as each individual cares only about his own probability of being in the elect. Note that the belief that only a marginal share belongs to the Elect is consistent with each individual's own belief but does not square with the distribution of beliefs, which can be interpreted as a case of non-common priors.

[14] This is due to the assumption that the share of those that belong to the elect is close to zero.

[15] In an alternative model one could assume that $\kappa$ is the individual's equilibrium probability of cooperation in the first period. The analysis of this model is complicated by the continuum of types and for tractability we focus on the fully cursed case.

are almost as cooperative as Christ would have been, i.e., tend to behave as they are told to by the scripts. This also implies that self signalling is a rather slow process. A lower level of $\kappa$ implies less disciplined societies, but faster self signalling.

Throughout we assume that $\varepsilon > a - c$. This means that some individuals in society, with strong enough beliefs, will always cooperate irrespective of their perception of others' behaviour. The expected utility of an individual in the beginning of the game is the sum over the individual's utility from his behaviour in the two periods where the second period's utility is discounted by $\delta$. Each period's utility includes the material payoff from the PD game, as well as the anticipation of the spiritual payoff, i.e., from his updated beliefs.

We look for an equilibrium in which: (i) at each period $t \in 1, 2$, individuals correctly conjecture the average probability of cooperation in the population, $\kappa_t$, and update their beliefs according to that and their own action; (ii) given their initial beliefs $\alpha$ and the expected behaviour in the population, individuals play optimally in the PD game.

## 3 The option value of self signalling

As a benchmark, consider first the case in which there is only one period. Let $\Delta_c = b - d$ and $\Delta_d = a - c$ (these are the incentives to defect conditional on the rival cooperating or defecting respectively). Given that a share $\kappa$ of the population cooperates and the rest defects, the relative payoff from cooperation vis-a-vis defection is

$$-\kappa\Delta_c - (1-\kappa)\Delta_d + \frac{\alpha}{\alpha + (1-\alpha)\kappa}\varepsilon$$

where the spiritual payoff is the anticipation of how beliefs will be updated following cooperation vis-a-vis defection.

This relative payoff is monotone in $\alpha$, which implies that there must be a cutoff $\hat{\alpha}$ above which all types cooperate and below which all defect. Thus $\kappa = 1 - \hat{\alpha}$ and $\hat{\alpha}$ solves

$$\frac{\hat{\alpha}}{\hat{\alpha} + (1-\hat{\alpha})^2}\varepsilon = (1-\hat{\alpha})\Delta_c + \hat{\alpha}\Delta_d$$

The left hand side measures the spiritual benefit of cooperation for the cutoff type $\hat{\alpha}$ given that he knows that all above him cooperate, and the right hand side measures the material cost of cooperation, which given the cutoff $\hat{\alpha}$, is the same for all types. Note that the right hand side is an increasing linear function in $\hat{\alpha}$ and the left hand side is an increasing function in $\hat{\alpha}$ as well. To see that an equilibrium value of $\hat{\alpha}$ must exist note that when $\hat{\alpha} = 0$, the *lhs* is smaller than the *rhs*, whereas when $\hat{\alpha} = 1$, the *lhs* is larger than the *rhs*.

Suppose now that the game has two periods of play. Again, the game must consist of two cutoffs, $\alpha_1$ and $\alpha_2$, such that all individuals above (below) $\alpha_t$ cooperate (defect) in period $t = 1, 2$. But all individuals who defected in the first period would have lost their beliefs and must defect again and hence it has to be that $\alpha_2 \geq \alpha_1$. In the Appendix we show that it cannot be that $\alpha_2 > \alpha_1$ and hence $\alpha_2 = \alpha_1 \equiv \tilde{\alpha}$. Note that cooperation in the second period is more attractive the higher are ones' beliefs about belonging to the elect. Therefore, all individuals who cooperated in the past will want to cooperate again.

In the first period, individuals take into consideration that they will also cooperate in the second period or in other words, individuals take into consideration that cooperation implies that they have the option value to self signal also in the second period. The first period cutoff is therefore a combination, weighted by $\delta$, of the myopic first and second period cutoffs.

The unique cutoff $\tilde{\alpha}$ solves:

$$\frac{\tilde{\alpha}}{\tilde{\alpha} + (1 - \tilde{\alpha})^2}\varepsilon + \delta\frac{\tilde{\alpha}}{\tilde{\alpha} + (1 - \tilde{\alpha})^3}\varepsilon = (1 + \delta)((1 - \tilde{\alpha})\Delta_c + \tilde{\alpha}\Delta_d) \qquad (1)$$

where arguments as above imply that the solution $\tilde{\alpha} \in (0, 1)$ exists. We can further show that:

**Proposition 1:** *There exists a unique cutoff $\tilde{\alpha} \in (0, 1)$ such that in equilibrium, in both periods, all individuals with beliefs above $\tilde{\alpha}$ cooperate and all with beliefs below $\tilde{\alpha}$ defect, The cutoff $\tilde{\alpha}$ decreases in $\delta$ and in $\varepsilon$, and increases in $\Delta_d$ and $\Delta_c$.*

There is more cooperation when the option value is larger (captured by $\delta$) and when the value from signalling is higher (captured by $\varepsilon$). When $\Delta_d$ or $\Delta_c$ are higher, the loss from cooperation when the opponent defects is larger, implying that the material cost from self signalling is higher, inducing individuals to cooperate less.

Note that the above analysis can be extended to more periods. The result above extends to any number of periods $T$, with a unique solution $\tilde{\alpha}_T > 0$ satisfying,

$$\varepsilon \sum_{t=1}^{T} \delta^{t-1} \frac{\tilde{\alpha}_T}{\tilde{\alpha}_T + (1 - \tilde{\alpha}_T)^{t+1}} = \sum_{t=1}^{T} \delta^{t-1}((1 - \tilde{\alpha}_T)\Delta_c + \tilde{\alpha}_T\Delta_d).$$

## 4 Discipline and cooperation

We have analysed a two-period model which has allowed us to focus on the option value of self signalling. But having more than one period also implies that agents may use their rivals' first period behaviour to glean information about their rivals' convictions. In particular, cooperation in the first period may signal strong beliefs and hence can be informative about second period behaviour.

In fact, one of Calvin's main innovations was the creation of an institution which monitored and made public individuals' behaviour in the social arena. A great deal of the Consistory's function was devoted to resolving disputes within families, neighbours, and among business associates. To be able to do so, the Consistory members met once a week and visited the homes of each Genevan family twice a year, which allowed them to be well informed. Deviant behaviour -religious and civil alike- was publicly punished. We now incorporate such an institution into our model. In particular, we assume that first period behaviour can be observed, so that players condition their second period behaviour on this information.[16]

Suppose that first-period behaviour is observed. In addition to the equilibrium notion specified above, it is now the case that individuals in the second period correctly anticipate the probability that an individual who cooperated in the first period will cooperate again and optimise accordingly.

To illustrate how social signalling and discipline may arise, consider a particular equilibrium in which information creates a strong discipline effect, so that all types cooperate in the first period. This implies that in the first period $\kappa_1 = 1$.[17] This behaviour is sustained by a fear that if one defects, some individuals will defect against him in the second period instead of cooperating, as we illustrate below. In the second period, cooperation carries no meaningful information, and behaviour will be as in a one-period game. In other words, in the second period, individuals above some $\hat{\alpha}_2$ cooperate and others defect, where $\hat{\alpha}_2$ solves

$$\frac{\hat{\alpha}_2}{\hat{\alpha}_2 + (1 - \hat{\alpha}_2)^2}\varepsilon = (1 - \hat{\alpha}_2)\Delta_c + \hat{\alpha}_2\Delta_d \qquad (2)$$

However, after an out-of-equilibrium behaviour in which an individual

[16] Such institution can induce social behaviour for fear of punishment. We instead take the view that its most important role was publicizing behaviour. See Kingdom (1992).

[17] There may be other equlibria with more meaningful social signalling behaviour in which only part of the population cooperates in the first period. The key intuition and results remain the same as the analysis below.

defects, all others believe, and rightly so, that he will defect again. Given that he defects, all types below $\bar{\alpha}_2$ will defect against him, where $\bar{\alpha}_2$ solves,

$$\frac{\bar{\alpha}_2}{\bar{\alpha}_2 + (1-\bar{\alpha}_2)(1-\hat{\alpha}_2)}\varepsilon = \Delta_d. \tag{3}$$

It is easy to see from the equations above that $\bar{\alpha}_2 > \hat{\alpha}_2$, and hence there exists an interval of types that will change their behaviour in response to a deviation. This creates the endogenous social discipline mechanism.

In the first period, by monotonicity, we need to check the incentives of the lowest type to cooperate. This type, with $\alpha = 0$, knows that the whole population cooperates in the first period, and that he will defect in the second period. To prefer to cooperate in the first period, we need:

$$-\Delta_c + \delta(\bar{\alpha}_2 - \hat{\alpha}_2)(b-a) > 0 \tag{4}$$

Note that $-\Delta_c$ is the loss in the first period from cooperation, conditional on all types cooperating. Cooperation vs. defection entails however a second-period gain from an interval of types who cooperate instead of defect. This type defects in the second period, and hence this provides a gain of $b-a$, weighed by the probability of meeting these types and by $\delta$. We then have:

**Proposition 2:** *Suppose that the incentive to defect given a rival's defection, $\Delta_c$, is sufficiently small. Then there exists a full discipline equilibrium in which in the first period all cooperate and in the second period all above (below) some cutoff $\hat{\alpha}_2 \in (0,1)$ cooperate (defect).*

# 5 Welfare

Note that given that cooperation is always beneficial, self signalling dominates a world in which all defect. But is social discipline, as characterized above, beneficial compared with self signalling?

Consider a comparison between the self signalling equilibrium and the social discipline equilibrium above. With only self signalling, cooperation is determined according to $\tilde{\alpha}$ as in equation (1). In the social discipline case, society fully cooperates in the first period, and second period cooperation is by all types above $\hat{\alpha}_2$, determined by (2). For all $\delta > 0$, $\tilde{\alpha} < \hat{\alpha}_2$; the motivation to self signal across two periods yield a higher cooperation level than in a one-period game (i.e., the second period in the social discipline case). This implies a possible trade-off; first period cooperation is higher in the social discipline equilibrium but second period cooperation levels are lower. We still find though that the average level of cooperation across the two periods can be higher under the discipline effect:

**Proposition 3:** *When $\varepsilon$ is sufficiently large, social welfare is higher in the social discipline equilibrium compared with the equilibrium with only self signalling.*

Note that the social discipline effect can only arise together with the self signalling mechanism; the point of Proposition 3 is that the combination of both the theology and the monitoring institution can increase social welfare, compared with an environment in which only the theology exists (self signalling), or only the institution (where no cooperation is possible). Intuitively, when $\varepsilon$ is sufficiently large, the cutoffs in the second period in the two different equilibria become sufficiently close, as both induce relatively high levels of cooperation. This allows the discipline equilibrium to dominate as it induces full cooperation in the first period.

## 6 Discussion

The role of religious organizations as providing information about ethical behaviour has been noted before. Adam Smith observed that religions tend to produce and distribute moral information about individual members, which allows traders to assess the risk involved in conducting business with these individuals.[18] Weber (1906) writes of the social pressure in American Protestant communities, "*Unqualified integrity, evidenced by, for example, a system of fixed prices in retail trade…appears as the specific, indeed, really the only, form by which one can demonstrate his qualification as a Christian and therewith his moral legitimation for membership in the sect…admittance into the Baptist congregation was primarily of decisive importance…because of the on-going inquiries about moral and business conduct*".

What is the mechanism that allows religious organization to signal ethical or moral qualities? In this paper we suggest one such mechanism. First, religious beliefs, for example those of predestination and justification, induce individuals to behave cooperatively in order to self signal. Without such a mechanism, there will be no cooperation even if behaviour is observed. Second, the institution of the consistory allows for social signalling, implying enhanced cooperation.

Naturally, other religious organizations have the theology and the institutions to allow for pro-social behaviour. Beliefs in rewards and punishments, whether in this life or the afterlife, are rife in many ancient and modern religions, and create a direct incentive to properly behave in a social context. The Roman Catholic Church assured salvation to individuals who had behaved properly and, in addition, accepted

[18] See Anderson (1988).

the Church's sacraments and submitted to the clerical authority. Participation in such public and costly rituals might also indicate a strong religious conviction; previous literature had focused on how costly religious rituals allow to signal religious beliefs or good behaviour to others (see Iannaccone 1992, Berman 2000, Chwe 2003 and Levy and Razin 2011a). In Levy and Razin (2011b) we compare between religious institutions according to the signalling mechanism they employ. How these relate to beliefs and religious governance seems a fruitful avenue for future research.

# 7 Appendix

**Proof of Proposition 1**: We show that a unique cutoff must arise. To proceed, suppose that $\alpha_2 > \alpha_1$. This implies that at the cutoff $\alpha_2$, the relative payoff of cooperation and defection is equal so the type at the cutoff is indifferent, and similarly the type at $\alpha_1$ disregards his second period behaviour which does not depend on his first period behaviour, and hence considers only the one period utility. Thus the following two equations have to be satisfied:

$$\frac{\alpha_1}{\alpha_1 + (1-\alpha_1)^2}\varepsilon = (1-\alpha_1)\Delta_c + \alpha_1\Delta_d$$
$$\frac{\alpha_2}{\alpha_2 + (1-\alpha_2)^2(1-\alpha_1)}\varepsilon = (1-\alpha_2)\Delta_c + \alpha_2\Delta_d$$

Note that the right hand side is linear for both expressions whereas the left hand side is first convex and then concave. Moreover in both equations the RHS is strictly larger than the LHS when $\alpha_1$ and $\alpha_2$, respectively, are zero, and the LHS strictly larger than the RHS when $\alpha_1$ and $\alpha_2$ are one, respectively. This implies that there is a unique cutoff $\alpha_1^*$ for the first equation and and $\alpha_2^*$ for the second. But note that the function on the left hand side of the second equation is higher than the function in the first, which implies that $\alpha_2 < \alpha_1$, a contradiction.

**Proof of Proposition 2:** Note that it is enough to show that $\delta(\bar{\alpha}_2 - \hat{\alpha}_2)(b-a)$ is bounded as $\Delta_c \mapsto 0$. Suppose that this is not the case, i.e., that $\lim_{\Delta_c \mapsto 0} \delta(\bar{\alpha}_2 - \hat{\alpha}_2)(b-a) = 0$. This implies that $\lim_{\Delta_c \mapsto 0} \hat{\alpha}_2 = \lim_{\Delta_c \mapsto 0} \bar{\alpha}_2 \equiv \alpha^*$ where $\bar{\alpha}_2$ and $\hat{\alpha}_2$ are given by,

$$\frac{\bar{\alpha}_2}{\bar{\alpha}_2 + (1-\bar{\alpha}_2)(1-\hat{\alpha}_2)}\varepsilon = \Delta_d$$
$$\frac{\hat{\alpha}_2}{\hat{\alpha}_2 + (1-\hat{\alpha}_2)^2}\varepsilon = (1-\hat{\alpha}_2)\Delta_c + \hat{\alpha}_2\Delta_d$$

As these two equations hold all along we take their limit to get

$$\frac{\alpha^*}{\alpha^* + (1-\alpha^*)^2}\varepsilon = \Delta_d$$
$$\frac{\alpha^*}{\alpha^* + (1-\alpha^*)^2}\varepsilon = \alpha^*\Delta_d$$

But as from the first equation $\alpha^* < 1$ (as $\varepsilon > \Delta_d$) we get a contradiction.■

**Proof of Proposition 3**: Note that when $\delta = 1$ this induces the highest chance for the self signalling equilibrium to improve social welfare as this puts more weight on the second period where there is an advantage for self signalling, as well as decreases the cutoff determined in (1). We will show that the discipline equilibrium dominates for $\delta = 1$ and thus it will dominate for all $\delta$.

For $\delta = 1$, social welfare is higher with social discipline if $\frac{\hat{\alpha}_2}{2} \leq \tilde{\alpha}$ (which implies a higher average level of cooperation across the two periods). Consider (2) where both sides are multiplied two:

$$\frac{2\hat{\alpha}_2}{\hat{\alpha}_2 + (1-\hat{\alpha}_2)^2}\varepsilon = 2((1-\hat{\alpha}_2)\Delta_c + \hat{\alpha}_2\Delta_d) \tag{5}$$

Plugging for $0.5\hat{\alpha}_2$ in (1) and comparing to (5), we have that the rhs of (1) is lower than that in (5) by $A = \hat{\alpha}_2(\Delta_d - \Delta_c)$ whereas the lhs is lower by $B = \frac{2\hat{\alpha}_2}{\hat{\alpha}_2+(1-\hat{\alpha}_2)^2}\varepsilon - \frac{0.5\hat{\alpha}_2}{0.5x+(1-0.5x)^2}\varepsilon - \frac{0.5\hat{\alpha}_2}{0.5x+(1-0.5x)^3}\varepsilon$. Note that if $B > A$ we will have that $\frac{\hat{\alpha}_2}{2} \leq \tilde{\alpha}$.

Note that

$$\begin{aligned}
\frac{B}{A} &= \frac{\varepsilon}{(\Delta_d - \Delta_c)}\left(\frac{2}{\hat{\alpha}_2 + (1-\hat{\alpha}_2)^2} - \frac{0.5}{0.5\hat{\alpha}_2 + (1-0.5\hat{\alpha}_2)^2} - \frac{0.5}{0.5\hat{\alpha}_2 + (1-0.5\hat{\alpha}_2)^3}\right) \\
&> \frac{\varepsilon}{\Delta_d}\left(2 - \frac{0.5}{0.75} - \frac{0.5}{0.625}\right) \\
&= \frac{\varepsilon}{\Delta_d}0.53333
\end{aligned}$$

and so we have that if $\varepsilon > \frac{\Delta_d}{0.53333}$, $\frac{B}{A} > 1$ ■

**References**


Ainslie, G. (1992), *Picoeconomics : The Strategic Interaction of Successive Motivational States Within the Person*, Cambridge, England and New York: Cambridge University Press.

Anderson, G. (1988), "Mr. Smith and the Preachers: The Economics of Religion in the Wealth of Nations", The Journal of Political Economy, Vol. 96, No. 5, pp. 1066-1088.

Andreoni, J. and L. Samuelson (2006), "Building rational cooperation", Journal of Economic Theory, vol. 127(1), pages 117-154.

Arrunada, B. (2010), "Protestants and Catholics: Similar Work Ethic, Different Social Ethic," Economic Journal 120 (547), 890–918.

Barro, R. and R. McCleary (2003), "Religion and Economic Growth across Countries", American Sociological Review, vol. 68(5), pp. 760-81.

Benabou, R. and J. Tirole (2011), "Identity, Morals and Taboos: Beliefs as Assets", discussion paper no. 6123, CEPR.

Benabou, R. and J. Tirole (2006), "Belief in a Just World and Redistributive Politics", Quarterly Journal of Economics, vol. 121(2), pp. 699-746.

Benabou, R. and J. Tirole (2004), "Willpower and Personal Rules", Journal of Political Economy, 112(4), PP. 848-886.

Berman, E., (2000), "Sect, Subsidy and Sacrifice: An Economist's View of Ultra-Orthodox Jews", Quarterly Journal of Economics, vol. 115(3), pp. 905-953.

Bodner, R. and D. Prelec (2003), "Self-Signaling and Diagnostic Utility in Everyday Decision Making" In *The Psychology of Economic Decisions*, Ed. Isabella Brocas and Juan Carillo, Oxford University Press.

Chwe, M.S. (2003), *Rational Ritual: Culture, Coordination, and Common Knowledge*, Princeton: Princeton University Press 2003.

Dixit, A. (2003), "Trade expansion and contract enforcement", Journal of Political Economy, Vol. 111, pp. 1293-1317.

Eyster, E. and M. Rabin (2005), "Cursed Equilibrium", *Econometrica,* 73(5), 1623-1672, September 2005.

Gilbert, W. (1998), *Renaissance and Reformation*, Electronic edition, Kansas University.

Glaeser E. and S. Glendon (1998), "Incentives and Predestination", Economic inquiry, vol 36.

Guiso, L. P. Sapienza, and L. Zingales (2003), "People's opium? Religion and economic attitudes," Journal of Monetary Economics 50, 225–282.

Huber, J.D. (2005), "Religious belief, religious participation, and social policy attitudes across countries", working paper, Columbia University.

Iannaccone, L. (1992), “Sacrifice and stigma: reducing free-riding in cults, communes and other collectives”, Journal of Political Economy, vol. 100(2), pp. 271-291.

Iannaccone, L. (1998), “Introduction to the Economics of Religion”, Journal of Economic Literature, vol. 36(3), pp. 1465-1495.

Kingdon, D. (1992), “Calvin and the Family: The Work of the Consistory in Geneva”, in Calvin’s *work in Geneva*, ed. R. Gamble, Graland.

Levy, G. and R. Razin (2011a), “Religious beliefs, religious participation, and cooperation”, forthcoming, American Economic Journals: Microeconomics.

Levy, G. and R. Razin (2011b), “Rituals or good behaviour? social signalling in religious organizations”, mimeo, LSE.

MacCulloch, D. (2010), *A History of Christianity*, Penguin Books.

Scheve, K. and D. Stasavage (2006), “Religion and Preferences for Social Insurance”, Quarterly Journal of Political Science, vol.1, pp.255-286.

Stark, R. (1996), *The Rise of Christianity*, Princeton University Press, Princeton.

Tabellini, G. (2008), “The Scope of Cooperation: Values and Incentives”, The Quarterly Journal of Economics, MIT Press, vol. 123(3), pages 905-950.

Weber, M. (1904), *The Protestant Ethic and the Spirit of Capitalism*, New York: Scribners (translated by T. Parsons, 1930).

Weber, M (1906), "“Churches” and “Sects” in Northern America: An Ecclesiastical Sociopolitical Sketch”, trans. Colin Loader. *Sociological Theory* 3: 7-13. (1986 translation).

Wilson, D.S. (2002), *Darwin’s Cathedral*, University of Chicago Press, Chicago.