Incentives, Socialization, and Civic Preferences [☆]

Sung-Ha Hwang[a], Samuel Bowles[b,*]

[a]*Korea Advanced Institute of Science and Technology (KAIST), Seoul, Korea*
[b]*Santa Fe Institute, U.S.A.*

Abstract

Change in social norms or motivations is typically studied as a process whereby preferences are updated under the influence of natural selection or some other decentralized process, or using models of cultural evolution in which parents inculcate values in their offspring. But preference change is sometimes an objective of deliberate policy, whether by religious orders, political parties, firms, or states. To study this process of deliberate preference manipulation, we consider a far-sighted social planner seeking to use material incentives to induce citizens to adopt what we term "civic preferences" that will motivate them to contribute unconditionally to a public good. A subsidy to contributors, for example, will encourage parents to raise their children to have civic preferences if, as is standard in cultural evolution models, the preference updating process favors higher payoff types. However, there is a second indirect and possibly offsetting effect that occurs if those with civic preferences are socially esteemed and contributing is a noisy signal of one's preferences. By inducing some self-interested types to contribute to the public good, the subsidy will diminish the social esteem value of really having civic preferences and this will lead parents to place a lesser weight on inculcating civic preferences in their offspring than they would in the absence of incentives. We characterize optimal incentives that would be selected by the planner who is cognizant of this cultural crowding-out process, and identify conditions under which greater use of incentives will be called for than would be the case of the absence of this adverse indirect effect on cultural transmission (rather than the opposite as would be expected).

Keywords: Social preferences, social planner, motivational crowding out, cultural evolution, explicit incentives, endogenous preferences
JEL Classification Numbers: D64 (Altruism); D78 (Policy making and implementation); D03 (Behavioral Economics); Z18 (Cultural economics, public policy)

[☆] This version: October 14, 2017.

[*] Corresponding author

Email addresses: sungha@kaist.ac.kr (Sung-Ha Hwang), samuel.bowles@gmail.com (Samuel Bowles)

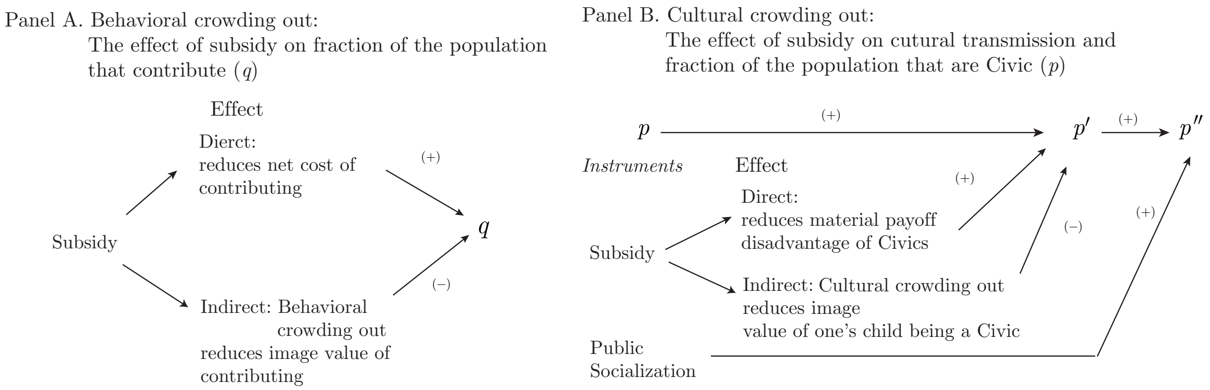
1. Introduction

For the past century, economists have extended the public economics paradigm initiated by Alfred Marshall and A.C. Pigou, devising incentives to induce self-interested individuals acting non-cooperatively to implement socially preferred allocations in cases where incomplete markets or impediments to efficient bargaining prevent the private economy from accomplishing this result. Modern mechanism design continues this tradition. In this approach, preferences are exogenous and incentives work by altering the economic costs or benefits of some targeted behavior such as contributions to a public good. However, policy makers may also seek to advance their objectives by deliberately altering preferences.

Here, we extend the public economics paradigm to consider the problem facing a social planner seeking to use incentives to motivate citizens to adopt civic-minded preferences that will induce them to contribute to a public good. We term as Civics, those individuals who place a positive intrinsic value on contributing to the public good sufficient to motivate them to contribute unconditionally, a character virtue that is socially admired. An example of civic preferences is a lexical commitment to abide by one's society's laws.

A subsidy paid to contributors may affect the evolution of the population fraction that are Civics in two ways, the first one intended, and the second not. First, the subsidy reduces the payoff disadvantage of Civics (who always contribute to the public good) relative to Non-civics, some of whom do not contribute (because the subsidy falls short of the cost of contributing). This will encourage the adoption of civic preferences if, as is standard in cultural evolution models, the cultural transmission process favors higher payoff types.

However, there may be a second possibly offsetting effect because there is a social esteem value associated with one's type, that is, being a Civic. By inducing some self-interested people to contribute to the public good, the subsidy will diminish the image value of really having civic preferences. This will lead parents who care about their child's well-being as adults to place a lesser weight on inculcating civic



Panel A. Behavioral crowding out:
The effect of subsidy on fraction of the population that contribute (q)

Panel B. Cultural crowding out:
The effect of subsidy on cutural transmission and fraction of the population that are Civic (p)

Figure 1: **Cultural transmission and crowding-out effects of incentives.**

preferences in their offspring. This second effect, operating through the effect of the social esteem motive on parenting practices, is what we term cultural crowding out.

Cultural crowding out is distinct from framing or other information effects that account for the fact that for any given citizen, the social esteem motive for contribution may also be reduced by the subsidy (as explored theoretically, for example, by Roland Benabou and Jean Tirole (2006)). When, as a result, the effect of a subsidy on contribution is diminished, we say that behavioral crowding out has occurred.

We illustrate the two forms of crowding out—behavioral and cultural—in Figure 1 and Table 1. Panel A of Figure 1 reproduces the causal logic of the behavioral crowding out of Benabou and Tirole (2006), while Panel B presents an overview of our model of endogenous preferences under the influence of cultural transmission and the indirect negative effects of a subsidy on the stationary fraction of the population who are Civics.

Behavioral crowding out is best represented by state-dependent preferences in which changes in the nature and extent of an incentive define different states, while cultural crowding out is a case of endogenous preferences. The key difference between the two panels in Figure 1, then, is that while the preferences involved in behavioral crowding out are time invariant but state-dependent, when preferences are culturally transmitted across generations, the effect of the incentive endures in the long run because the updating process, on which cultural transmission is based, typically

Type of *Crowding out*	*Actor*	*Action*	*Crowding mechanism* The subsidy diminishes:	*Representation*
Behavioral	Citizen	Contribute to public good	Image value of contributing	Equation (5)
Cultural	Parent	Raise child as a Civic	Image value of one's child being a Civic	Equation (11)

Table 1: **Cultural and behavioral crowding-out effects.**

occurs during youth and its effect persists over decades if not the entire lifetime.

The present study makes three contributions to the literature: two methodological and one substantive. First, we extend the standard replicator equation dynamics, which has provided the basis for modeling cultural evolution, to include the joint effects of incentives, social image motivations, and publicly-supported socialization[1](Section 2). Second, we incorporate a standard model of behavioral crowding out with a given distribution of preference types in the population into a cultural model to represent the effect of incentives on the distribution of preferences in the long run (Sections 2 and 3). Third, we use this model of endogenous preferences to characterize optimal incentives that would be adopted by a sophisticated social planner who is aware that both behavioral and cultural crowding out occurs (Sections 4 and 5).

2. Social esteem, incentives, and cultural transmission: Model setup

In our model, citizens' decisions on whether to contribute to a public project lead to a behavioral equilibrium in the short run, while the evolution of preferences takes place through parental upbringing and socialization, resulting in a cultural equilibrium in the long run. We present a schematic representation of our model in Figure 2.

We begin with citizens' decisions on whether to contribute to a public project when they are concerned about the image value of contribution, using a simple Bayesian signal extraction model (Benabou and Tirole, 2006; Ali and Lin, 2013).

[1] Extending Bisin and Verdier (2011); Boyd and Richerson (1985); Cavalli-Sforza and Feldman (1981); Bowles (1998).

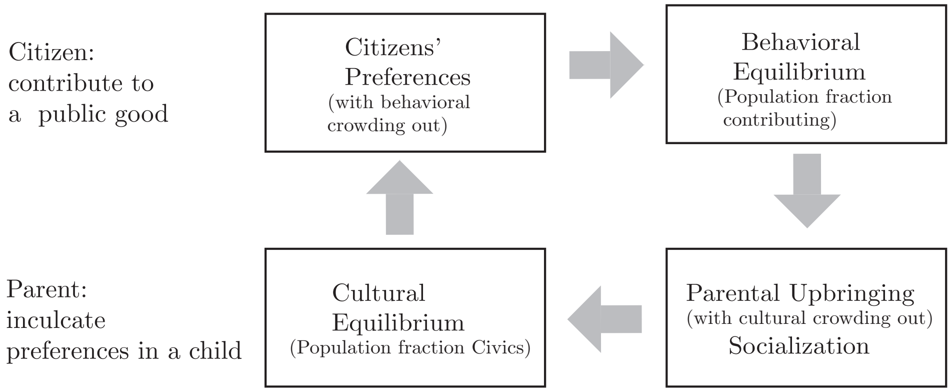


Figure 2: **The formation of cultural equilibrium.**

2.1. The image value of being a contributor

Consider a continuum of individuals who engage in two activities. In period 1, as citizens, they may contribute to a public good, and in period 2, as parents, they inculcate preferences in their children. We suppose that there are two preference types, called *Civics* (denoted by C) and *Non-civics* (denoted by N) and that citizen i bears the cost of contribution, g_i, which is distributed according to a distribution function $F(\cdot)$ with support $[0, \bar{g}]$. Thus, agents are heterogenous with respect to the cost of contribution, responding differently to the subsidy, $s \in [0, \bar{s}]$ so that the net cost of contribution is $g_i - s$. To take account of the fact that some Non-civics contribute, we denote by p and q the population fractions of Civics and contributors, respectively, to be determined endogenously. The citizens' contributions produce a pure public good, resulting in a benefit to each citizen of $\phi(q)$, where $\phi(\cdot)$ is a positive and increasing function.

We assume that Civics always contribute (see Assumption A1). In deciding whether to contribute (in period 1) Non-civic citizens are concerned about the image value of being (considered to be) a Civic ($v = 1$) or not ($v = 0$) by taking action (a). Following Benabou and Tirole (2006) and Ali and Lin (2013), we model the image value as a posterior expectation of being regarded as a Civic conditional on having taken the action, $\mathbb{E}[v|a]$, whose explicit expression will be derived shortly. Thus, a Non-civic citizen i's payoff from taking action $a_i = 1, 0$ (i.e., whether to contribute

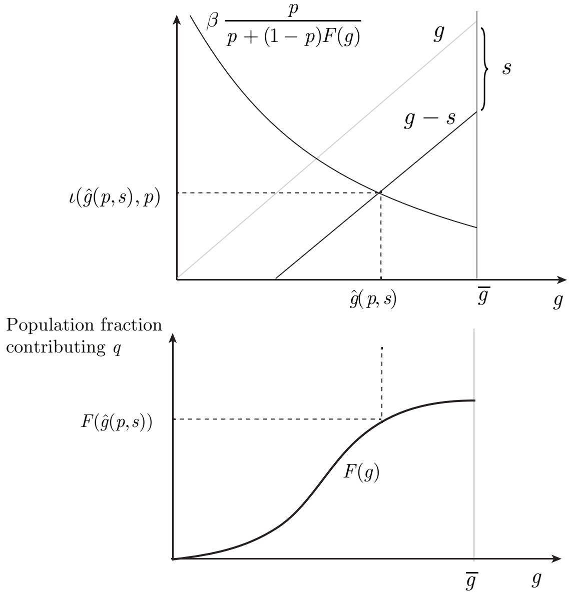


Figure 3: **Determination of the fraction of Non-civic contributors.** In the upper figure, if $g < \hat{g}(p,s)$, then $\mathbb{E}[v|a] > g - s$. In the bottom figure, we show the fraction of Non-civics who contribute.

or not, respectively) is

$$\phi(q) + \beta\mathbb{E}[v|a_i] - (g_i - s)a_i, \tag{1}$$

where the first term represents the public goods benefit function. The second term denotes the image value of contributing, with β representing the subjective value of having a good image relative to the material costs and benefits measured by s, g_i and $\phi(\cdot)$. The third term is the net contribution cost when contributing. Thus, from (1), a Non-civic agent i contributes if

$$\beta\mathbb{E}[v|a_i = 1] - (g_i - s) > \beta\mathbb{E}[v|a_i = 0]. \tag{2}$$

We assume that those who do not contribute are never considered to be Civics, so, $\mathbb{E}[v|a_i = 0] = 0$, and this choice of normalization is taken for simplicity. We look for an equilibrium in which Non-civic citizens contribute if the cost of contribution is less than $\hat{g}$, otherwise, they do no. Specifically, at a behavioral equilibrium, we require that

(i) Individual optimization: for a given $\hat{g}$, $\begin{cases} i \text{ contributes }, & \text{if } g_i < \hat{g} \\ i \text{ does not contribute}, & \text{if } g_i > \hat{g}. \end{cases}$

(ii) Consistency: $\mathbb{E}[v|a_i = 1] = \dfrac{p}{p + (1-p)F(\hat{g})}$ for all i

where the expression in the consistency requirement is obtained by the Bayes'rule, and the numerator in the expression denotes the population fraction of Civics and the denominator denotes the population fraction of contributors consisting of both Civics and Non-civic contributors. Then, the payoff condition in (2) ensures that there exists the threshold cost of contribution, $\hat{g}$, satisfying these two requirements (see Figure 3), namely

$$\beta \frac{p}{p + (1-p)F(\hat{g})} + s - \hat{g} = 0. \tag{3}$$

Thus, to the Non-civic citizen with the threshold cost, $\hat{g}$, the image value is equal to the net costs of contribution, $\hat{g} - s$, and the Non-civic citizen is indifferent between contribution and non-contribution. Also, clearly from (2), a Non-civic agent i with the cost g_i contributes if and only if $g_i < \hat{g}$ where $\hat{g}$ is given by (3).

We will denote the (equilibrium) image value by $\mathbb{E}[v|a]$, whose expression is given by

$$\mathbb{E}[v|a] := \frac{p}{p + (1-p)F(\hat{g})} \tag{4}$$

Note also that the image value would be maximal (i.e., $\mathbb{E}[v|a] = 1$) when every citizen is a Civic ($p = 1$) or when there are no Non-civic contributors ($F(\hat{g}) = 0$). Thus, the first two terms in (3) are bounded above by $\beta + \bar{s}$. We assume that some Non-civics contribute, while the others do not; thus, we require that $\bar{g} > \beta + \bar{s}$, which ensures $\bar{g} > \hat{g}$ from (3). This means that the subsidy does not fully offset the contribution cost of the Non-civic agent with the largest contribution cost, $\bar{g}$, who therefore does not contribute.

A1 Civics always contribute. Some Non-civics contribute, while other Non-civics do not; i.e., $\bar{g} > \beta + \bar{s}$:

We may consider a situation in which some Civics may not contribute as well. Under this alternative setting, we can also find the threshold cost level for Civics by specifying Civic citizen i's utility similar to (1) and modify the equilibrium image value in (4) properly and study the problem. However, this new setting will not change our results qualitatively.[2]

In sum, the population fraction of contributors (q) is given by the sum of the Civic and Non-civic contributor fractions, $q = p + (1-p)F(\hat{g})$, and the aim of the subsidy is to increase q and therefore, for a given p to reduce $\mathbb{E}[v|a]$. This is what we call

$$\text{Behavioral crowding out: } \text{sign}(\frac{\partial \mathbb{E}[v|a]}{\partial s}) = -\text{sign}(F'(\hat{g})\frac{\partial \hat{g}}{\partial s}) < 0 \tag{5}$$

This result reproduces the logic of Benabou and Tirole (2006).

2.2. *Parental upbringing, differential payoffs, and cultural crowding out*

Each citizen is also a parent of a single child who will be a citizen in the next generation. So, in period 2, a two-stage preference adaptation process takes place (see Panel B in Figure 1 and Figure 2): parental upbringing of the next generation followed by socialization by public interventions such as schooling (the latter is considered in the

[2] More precisely, we could define Civic citizen i's payoff by

$$\phi(q) + \beta\mathbb{E}[v|a_i] - (g-s)a_i + C$$

where $C > 0$ represents an additional valuation attached to the provision of the public good as a Civic. Then, a new image value term becomes as follows:

$$\mathbb{E}[v|a] := \frac{pF(\hat{g_C})}{pF(\hat{g_C}) + (1-p)F(\hat{g}_N)}$$

where $\hat{g}_C$ and $\hat{g}_N$ are threshold costs for Civics and Non-civics, respectively and we can find the equilibrium values for $\hat{g}_C$ and $\hat{g}_N$ similar to Equation (3). Under this setup, the fraction of contributor is given by $q = pF(\hat{g}_C) + (1-p)F(\hat{g}_N)$ and a similar analysis is possible.

next subsection). In the first stage, the parent inculcates preferences in her child, seeking to maximize the child's expected payoffs as an adult based on the payoffs that the parent expects her offspring to obtain when grown up depending on whether he is a Civic or not. Using Equation (1), we find the expected payoffs for Civics (π_C) and Non-civics (π_N) as follows:

$$\pi_C(p,s) := \phi(q) + \beta\mathbb{E}[v|a] + s - \mathbb{E}[g] \tag{6}$$

$$\pi_N(p,s) := \underbrace{(\phi(q) + \beta\mathbb{E}[v|a] + s - \mathbb{E}[g|g \leq \hat{g}])}_{\text{contributors' expected payoff}} F(\hat{g}) + \underbrace{\phi(q)}_{\substack{\text{non-contributors'}\\ \text{expected payoffs}}} (1 - F(\hat{g})). \tag{7}$$

Since the contribution cost (g) of the child is unknown to the parent, the expression $\pi_C(p,s)$ in (6) is the expected adult payoff for the child as a Civic, where $\mathbb{E}[g]$ is the unconditional expectation of g. In Equation (7), $\pi_N(p,s)$, is the expected adult payoff for a child as a Non-civic. The first underbraced term in (7) is the payoff to a Non-civic contributor, where $\mathbb{E}[g|g \leq \hat{g}]$ is the conditional expectation of the contribution cost, conditional on being a Non-civic contributor. This term is multiplied by the probability of being a contributor ($F(\hat{g})$). The last term is the payoff to a Non-civic non-contributor (i.e., a free-rider) multiplied by the probability of being a free-rider $(1 - F(\hat{g}))$. Alternatively, π_C and π_N can be interpreted as the average of payoffs for Civics and Non-civics in the population.

We study cultural crowding, first by subtracting π_N from π_C in (6) and (7) to find the explicit expression for the payoff advantage for Civics, $\Delta\pi(p,s) := \pi_C - \pi_N$, composed of two parts: the difference in image value (the first term on the right hand side below); and the difference in material payoffs, namely, the cost of contribution minus the subsidy (the second term):

$$\Delta\pi(p,s) = \underbrace{\beta[\mathbb{E}[v|a] - \mathbb{E}[v|a]F(\hat{g})]}_{\substack{\text{the expected image value payoff difference:}\\ \text{the image value of being raised as a Civic}}} - \underbrace{[(\mathbb{E}[g] - s) - (\mathbb{E}[g|g \leq \hat{g}] - s)F(\hat{g})]}_{\substack{\text{the expected material payoff difference:}\\ \text{the material payoff disadvantage for a Civic}}} \tag{8}$$

The material payoff disadvantage for Civics relative to Non-civics in Equation (8) can be rearranged as

$$
\begin{aligned}
(\mathbb{E}[g]-s)-(\mathbb{E}[g|g\leq\hat{g}]-s)F(\hat{g}) =& \mathbb{E}[g-s]-(\mathbb{E}[g-s|g\leq\hat{g}]F(\hat{g})) \\
=& \int_{\hat{g}}^{\bar{g}}(g-s)dF(g)>0. \qquad (9)
\end{aligned}
$$

Thus, as expected, the material payoff disadvantage for Civics is just the material payoff advantage for Non-civic non-contributors (i.e., free-riders) who avoid the net contribution cost $(g-s)$. From Equation (3) and Assumption A1, the subsidy s is always less than or equal to the threshold level of contribution cost $\hat{g}$ (i.e., $\hat{g}\geq s$) and cannot completely offset the cost of contribution for Non-civics with the cost g greater than $\hat{g}$ (i.e., $g>s$ for all $g>\hat{g}$.) Thus, Equation (9) is positive and Civics always experience a material payoff disadvantage relative to Non-civics. Note that in the absence of free-riding Non-civics (i.e., if $\hat{g}$ were equal to $\bar{g}$), Civics would not have a material payoff disadvantage.

The effect of the subsidy on the extent of material payoff disadvantages for Civics is

$$
\frac{\partial}{\partial s}(\int_{\hat{g}}^{\bar{g}}(g-s)dF(g)) = -\underbrace{(1-F(\hat{g}))}_{\substack{\text{reduction in the pay off advantages}\\ \text{of those civics who free ride}}} - \underbrace{(\hat{g}-s)\frac{\partial\hat{g}}{\partial s}F'(\hat{g})}_{\substack{\text{fewer Non-civics enjoy}\\ \text{the benefits of free-riding}}} < 0, \qquad (10)
$$

which confirms that an increase in subsidy will reduce the material payoff disadvantage of being raised as Civics in two ways. First, an increase in the subsidy directly reduces the payoff advantages of free-riding Non-civics (the magnitude of this effect is $1-F(\hat{g})$ in (10), namely the fraction of free-riding Non-civics). Second, an increase in the subsidy also induces fewer Non-civics to free ride, thereby further mitigating the Civics' cost disadvantage. The magnitude of this second effect is given (in the second term in (10)) by the net contribution cost of the new marginal contributors $(\hat{g}-s)$ multiplied by the effect of subsidy on $\hat{g}$ $(\frac{\partial\hat{g}}{\partial s})$ and the resulting mass of new contributors $(F'(\hat{g}))$. In other words, the effect of the subsidy on the material payoff *advantage* for Civics is positive $(\frac{\partial}{\partial s}(\int_{\hat{g}}^{\bar{g}}(s-g)dF(g))>0$ from (10)).

The effect of the subsidy on the advantage in image value for Civics advantage, however, is negative.

$$\text{Cultural crowding out: } \frac{\partial}{\partial s}(\mathbb{E}[v|a] - \mathbb{E}[v|a]F(\hat{g})) = \\ \frac{\partial \mathbb{E}[v|a]}{\partial s}(1 - F(\hat{g})) - \mathbb{E}[v|a]F'(\hat{g})\frac{\partial \hat{g}}{\partial s} < 0 \qquad (11)$$

As (11) makes clear, cultural crowding out occurs for two reasons. First, the subsidy will reduce the image value of contributing, $\mathbb{E}[v|a]$, as we have already seen from Equation (5); and, second, it will decrease the behavioral difference between Civics and Non-civics by inducing more of the latter to contribute and hence to enjoy the same image value as Civics. Thus, we have (11).

Now we find the total effect of the subsidy on the payoff differential, $\Delta\pi(p, s)$, by subtracting (10) from (11):

$$\frac{d\Delta\pi}{ds} = \frac{\partial \mathbb{E}[v|a]}{\partial s}(1 - F(\hat{g})) - \underbrace{\mathbb{E}[v|a]F'(\hat{g})\frac{\partial \hat{g}}{\partial s}}_{\text{(i)}} + (1 - F(\hat{g})) + \underbrace{(\hat{g} - s)F'(\hat{g})\frac{\partial \hat{g}}{\partial s}}_{\text{(ii)}}. \qquad (12)$$

Recall that to the Non-civic agent with the threshold cost $\hat{g}$, the image value $\mathbb{E}[v|a]$ is equal to the net cost of contribution $\hat{g} - s$. Thus, two terms, designated by the brackets (i) and (ii), in (12) cancel out. By rearranging, we obtain

$$\frac{d\Delta\pi}{ds} = (\frac{\partial \mathbb{E}[v|a]}{\partial s} + 1)(1 - F(\hat{g})) = \frac{\partial \hat{g}}{\partial s}(1 - F(\hat{g})) > 0 \qquad (13)$$

where we again use the fact that $\mathbb{E}[v|a] = \hat{g} - s$ to derive the second equality.

Equation (13) shows that the positive effect of subsidy on the material payoffs dominates its negative effect on the image value. Thus in parental practices for raising their offspring, the adverse effect of the subsidy on the image value cannot completely crowd out the positive effect of the subsidy on the material payoff advantage of Civics. This means that the sometimes observed counter-productive effect of incentives termed as “strong crowding out” cannot occur in this setting.[3] Cultural crowding out, in our model, diminishes but does not reverse the intended effect of the subsidy.

Civics on average enjoy a higher image value than Non-civics. However, a question that arises is whether there exists a sufficiently large value of β (the subjective value of having a good image) that would compensate for the greater contribution costs borne by Civics. A second natural question is whether there exists some level of subsidy less than fully compensating for the cost of contribution of the citizen facing the highest cost, which would motivate parents to raise their children as Civics.

To answer these questions, using Equations (8) and (9) and $\mathbb{E}[v|a] = \hat{g} - s$ we explicitly find that

$$\Delta\pi(p, c) = (\hat{g} - s)(1 - F(\hat{g})) + \int_{\hat{g}}^{\bar{g}} (s - g)dF(g) = \int_{\hat{g}}^{\bar{g}} (\hat{g} - g)dF(g) \leq 0, \tag{14}$$

which shows that the payoffs to Civics cannot exceed those to Non-civics even if β is sufficiently high. This is because when the subjective weight placed on the image value is sufficiently large, the threshold cost $\hat{g}$ becomes $\bar{g}$ and Civics and Non-civics are behaviorally indistinguishable. Hereafter, to avoid notational clutter, and without loss of generality, we assume that β, the payoff from the image value $\mathbb{E}[v|a]$, is 1.

If the payoffs to Civics must fall short of those to Non-civics, parental upbringing alone cannot support the evolution of civic preferences. We therefore introduce the second stage affecting the preferences of the next generation: public socialization.

2.3. Cultural evolution with public socialization

In our cultural transmission model, the payoff differential, $\Delta\pi = \pi_C - \pi_N$, can be considered a cultural fitness differential, taking account of the effect of the subsidy and both the image value and material payoffs associated with being a Civic or

[3] Bowles and Polania Reyes (2012); Gneezy and Rustichini (2000)

(Random) matching probabilities

	Civic	Non-civic
Civic	p^2	$p(1-p)$
Non-civic	$(1-p)p$	$(1-p)^2$

Switching probabilities

From C to C	No switch
From C to N	$\mu[\pi_C - \pi_N]_+$
From N to C	$\mu[\pi_N - \pi_C]_+$
From C to C	No switch

Table 2: **Matching and switching probabilities**. The left table shows matching probabilities between two types, Civics and Non-civics. The right table shows switching probabilities between matched pairs.

not. Each parent is paired with a cultural model chosen randomly from the parent's generation in the population. If the model and the parent have the same preference type, the parent inculcates her own preference in the child (thus, no change in the population fraction of Civics in the next generation). However, if the model and the parent have different preference types, the parent may inculcate a preference different from her own in the child (change in the population fraction of Civics).

In the parental upbringing stage the child of a Non-civic will become a Civic (we term this as "switch") with a probability equal to $\mu[\pi_C - \pi_N]_+$, while the opposite switch occurs with a probability of $\mu[\pi_N - \pi_C]_+$, where μ is a positive coefficient converting payoff differences into switch probabilities and ensuring that $\mu[\pi_C - \pi_N]_+$ and $\mu[\pi_N - \pi_C]_+$ are not greater than 1 and the operator $[]_+$ is defined by $[t]_+ = \max\{t, 0\}$, ensuring that the term in $[]_+$ is non-negative (see Table 2).

Following Bowles (2004), the population fraction of Civics at the end of the first stage p' (see Panel B in Figure 1) is just the prior frequency of Civics plus those Non-civic parents who have inculcated a civic preference in their children minus those Civic parents who have inculcated a non-civic preference in their children, or

$$p' := p + \underbrace{(1-p)p\mu[\pi_C - \pi_N]_+}_{\text{from a Non-civic to a Civic}} - \underbrace{p(1-p)\mu[\pi_N - \pi_C]_+}_{\text{from a Civic to a Non-civic}} = p + \mu p(1-p)(\pi_C - \pi_N) \tag{15}$$

where we use $[t]_+ - [-t]_+ = t$ in the second equality and the adaptation process returns a value of p', the fraction of the next generation that are Civics following the parental upbringing stage. The term, $p(1-p)\mu[\pi_N - \pi_C]_+$, in (15), for example, is the fraction of the population who are Civics paired with a Non-civic cultural model ($p(1-p)$) and hence who may inculcate non-civic rather than civic preferences in their offspring, multiplied by the probability of a switch given by the difference in cultural fitness of Non-civics and Civics ($\mu[\pi_N - \pi_C]_+$). Note that under our specification of π_C and π_N, $\pi_N \geq \pi_C$ (Equation (14)); thus, $\mu[\pi_C - \pi_N]_+ = 0$ and $\mu[\pi_N - \pi_C]_+ = \mu(\pi_N - \pi_C)$.

The accounting equation (15), also called the "in and out" equation in the literature, can be motivated by an alternative theory—the scenario of the cultural transmission model developed by Bisin and Verdier (2001). In Appendix A, we present this version for interested readers, which shows that Equation (15) remains invariant under plausible specifications of switching or transmission and thus that the cultural dynamic equation (15) is not model specific.

In the second stage, a public socialization signal is observed by everyone, and a fraction, m, of the $1-p'$ which constitutes Non-civics is converted to Civics, where $m < 1$. The resulting fraction of the population who are Civics as a result of both parental upbringing and public socialization is thus

$$p'' = p' + m(1-p'). \tag{16}$$

Note that the public socialization effect on p'' is zero when $p' = 1$ (there are no Non-civics to socialize), and large when p' is close to 0. Thus, the subsidy and public socialization are effectively competing to convert the Non-civics, and the more effective the subsidy, the fewer Non-civics there are for the signal to socialize publicly. There are many plausible alternative representations of the public socialization process; we adopt this formulation because it is a simple way to ensure an interior stationary state in the cultural evolution dynamic, consistent with the empirical observation that the population is heterogeneous when it comes to non-economic motivations such as those represented by our Civics.[4]

[4] Camerer (2003); Fehr and Gaechter (2000); Henrich et al. (2005); Loewenstein and Bazerman

Putting the two stages of cultural transmission—parental inculcation and public socialization—together by substituting p' in (15) into (16), we obtain the following cultural evolution dynamics:

$$\Delta p = p'' - p = (1-m)(\mu p(1-p)(\pi_C - \pi_N)) + m(1-p). \tag{17}$$

Then setting $dp/dt \approx \Delta p = p'' - p$, we have

$$\frac{dp}{dt} = (1-m)(\mu p(1-p)(\pi_C - \pi_N)) + m(1-p). \tag{18}$$

Finally, inserting the expressions for the payoff terms, (8) and (9), into (18) yields an explicit expression for the cultural evolution dynamics:

$$\begin{aligned}\frac{dp}{dt} =& (1-m)\mu p(1-p)\Delta\pi(p,s) + m(1-p) \qquad (19)\\ =& (1-m)\mu p(1-p)\left\{\underbrace{\mathbb{E}[v|a](1-F(\hat{g}))}_{\text{image value>0}} - \underbrace{\int_{\hat{g}}^{\bar{g}}(g-s)dF(g)}_{\text{material payoffs<0}}\right\} + \underbrace{m(1-p)}_{\text{socialization>0}}\end{aligned}$$

Observe that in the absence of the socialization effect ($m = 0$) Equation (19) reproduces the replicator dynamic—the equations that are most frequently adopted in studying the population dynamics in the standard literature on evolutionary games (Weibull, 1995). In this way, our cultural dynamic equation (19) extends the existing dynamics to study the cultural evolution of preferences under the influences of image value, incentives, and public socialization.

3. Implementation by equilibrium preferences under crowding out

We now use the above results to examine how the social planner's choice of a subsidy supports differing stationary distributions of preferences in the population. From

(1989).

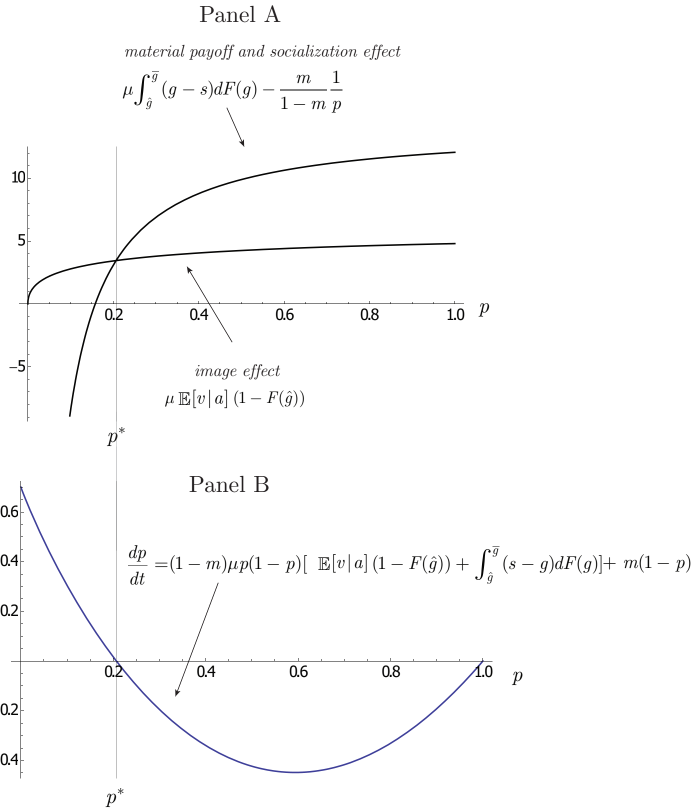


Figure 4: **Determination of the stable stationary state.** These figures show the determination of the fraction of citizens who are *Civics* in a cultural equilibrium, namely p^* when $s = 0$. Pane A shows the left hand side of (20) and the right hand side of (20) and Panel B shows $\frac{dp}{dt} = (1-m)\mu p(1-p)\Delta\pi(p,s) + m(1-p)$ in (19).

(19), an interior equilibrium preference distribution (i.e., an interior stationary point $p^* \in (0,1)$ so that $dp/dt = 0$) requires that at p^*, the image effect in parenting practices favoring Civics equals the material payoff net of socialization effects favoring Non-civics, or

$$\mu(\hat{g}(p^*,s)-s)(1-F(\hat{g}(p^*,s))) = \mu\int_{\hat{g}(p^*,s)}^{\bar{g}}(g-s)dF(g) - \frac{m}{1-m}\frac{1}{p^*}, \qquad (20)$$

where we again use $\mathbb{E}[v|a] = \hat{g} - s$.

Under the dynamic equation (19), note that when no citizens are Civics ($p = 0$), public socialization always induces some positive fraction of Civics, leading to an increased fraction of Civics. Thus, to have a stable interior equilibrium p^* under the dynamic equation (19), it is sufficient that when all citizens are Civics, the material payoff (net of socialization) advantage favoring Non-civics is greater than the image effect favoring Civics, leading to decreases in the fraction of Civics in the population (see Figure 4).

Recall that when all citizens are Civics, $p = 1$, the image value of contribution is maximal, and in this case, the threshold value of the contribution cost is equal to the sum of the subsidy and the maximal image value, i.e., $\hat{g} = s + \beta = s + 1$ from (3). Thus, by substituting $\hat{g} = s + 1$ and $p = 1$ into (20), we obtain the following condition:

$$\mu(1 - F(s + 1)) < \mu \int_{s+1}^{\bar{g}} (g - s) dF(g) - \frac{m}{1 - m} \tag{21}$$

(see Appendix B for the derivation of (21)).

Proposition 1 (The existence of a stable cultural equilibrium)**.** *Suppose that the condition in* (21) *holds. Then there exists a stable stationary state* p^**.*

Proof. See Appendix B. □

In Appendix B, we also illustrate that when F follows a uniform distribution, under plausible parametric values, there exists a unique stable stationary point p^*.

Also observe that Equation (20) implicitly defines the social planner's implementation function for it gives for each value of s the resulting stationary fraction of Civics in the population. To study how variations in s affect the evolution of civic preferences in the presence or absence of crowding out, we define κ, the evolutionary advantage of Civics satisfying $dp/dt = p(1-p)\kappa$. Note that κ is just the time derivative of p, normalized by the speed of adjustment of the replicator dynamic, namely, the fraction of the population that is matched with cultural models of a different type from individual types $(p(1 - p))$.

Because κ is a measure of the extent to which the combined effects of image value, the subsidy, and socialization will cause the fraction of the population switching from

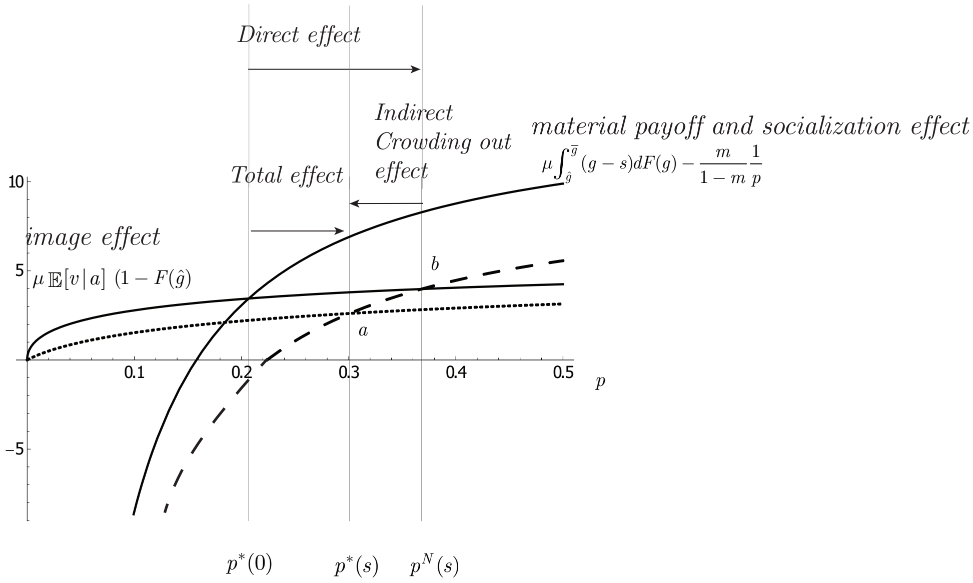


Figure 5: **Cultural crowding out.** The solid curves are the same as in Panel A of Figure 4, namely when $s = 0$. The subsidy reduces the material payoff difference between Non-civics and Civics, shifting down the material payoff and socialization effect curves (the dashed curves). However, the subsidy also reduces the image value, shifting down the image effect curve too (the dotted curves). This effect is called cultural crowding out (the indirect effect) and the resulting population fraction of Civics is given by $p^*(s)$, rather than $p^N(s)$ the population fraction of Civics that would have occurred in the absence of cultural crowding out.

Non-civics to Civics to exceed the fraction switching in the opposite direction,

$$\begin{aligned}\kappa(p,s,m) &= (1-m)\mu\Delta\pi(p,s)+\frac{m}{p}\\ &= (1-m)\mu[\underbrace{(\hat{g}-s)(1-F(\hat{g}))}_{\text{image value effect}}-\underbrace{\int_{\hat{g}}^{\bar{g}}(g-s)dF(g)]}_{\text{payoff \& social. effect}}+\frac{m}{p}, \qquad (22)\end{aligned}$$

we know that at the cultural equilibrium, p^* , $\kappa = 0$. We now illustrate the direct and indirect effects of a subsidy. The direct effect of the introduction of a subsidy $s > 0$ is to reduce the material payoff (net of socialization) disadvantage of Civics, increasing κ and shifting down the material payoff and socialization effect function in Figure 5, shown by the dashed curved line (Equation (10)). However, the subsidy also affects the image effect on parenting, shifting downward the curve shown by the dotted curved line (Equation (11)). The result is that the total effect of the subsidy on the evolutionary advantage of Civics is less than the direct effect. The indirect effect on the population fraction of Civics is consequently less than the direct effect; i.e., following the introduction of the subsidy, the implemented population fraction of Civics, $p^*(s)$, is less than the hypothetical population fraction of Civics, $p^N(s)$, that would be expected if one were to consider the direct effect only.

From $\kappa(p,s,m) = (1-m)\mu\Delta\pi(p,s) + \frac{m}{p}$ and Equation (13), the effect of the subsidy on the evolutionary advantage of the Civics is given by

$$\frac{\partial \kappa}{\partial s} = \mu(1-m)(1-F(\hat{g}))\frac{\partial \hat{g}}{\partial s} > 0, \tag{23}$$

thus, as we have already seen, the direct effect always dominates the indirect effect. This is because the crowding-out effect occurs entirely via the positive effect of the subsidy on contributing behavior. By similar computation, the evolutionary impact of public socialization evaluated at the status quo population distribution p^* is

$$\left.\frac{\partial \kappa}{\partial m}\right|_{p=p^*} = \frac{1}{1-m}\frac{1}{p^*} > 0 \tag{24}$$

which diminishes with greater use of incentives because incentives raise p^*. This discussion leads to the following proposition.

Proposition 2 (Effects of subsidy and public socialization)**.** *Given the cultural transmission process described in (19), at a stable stationary state, the following holds:*
(a) The subsidy increases the equilibrium fraction of Civics: i.e.,

$$\frac{\partial p^*}{\partial s} > 0$$

(b) Socialization increases the equilibrium fraction of Civics: i.e.,

$$\frac{\partial p^*}{\partial m} > 0.$$

Proof. From (22), we find that $\kappa(p^*(s,m),s,m)=0$. Thus, we have

$$\frac{\partial p^*}{\partial s} = -\frac{\partial\kappa/\partial s}{\partial\kappa/\partial p}, \qquad \frac{\partial p^*}{\partial m} = -\frac{\partial\kappa/\partial m}{\partial\kappa/\partial p}$$

and since $p=p^*$ is stable, $\partial\kappa/\partial p<0$ (see Appendix B) and (23) and (24) provide us the results stated. □

From (23), we know that the greater is the extent of socialization, the less effect the subsidy will have. Thus, in our setting, incentives and public socialization are substitutes in the sense that an increase in the level of one diminishes the other's marginal effect on the evolutionary advantages of Civics:

$$\left.\frac{\partial^2\kappa}{\partial s\partial m}\right|_{p=p^*} = -\mu(1-F(\hat{g}))\frac{\partial\hat{g}}{\partial s} < 0.$$

4. Optimal incentives with endogenous preferences

Using the results on the implementation function just derived, the far-sighted social planner wishes to select a level of subsidy to increase the fraction of Civics in the population and a level of public goods contribution among Non-civics that will maximize what we term social welfare, namely, the benefits of the public good vis-à-vis the net of costs of provision borne by citizens and the costs of the policies she adopts. For reasons of tractability we assume that the planner is sufficiently far sighted so as to abstract from the benefits and costs incurred on the path from the status quo to the implemented optimal outcome, treating the problem comparative statically rather than dynamically.

While, as we will see, the problem can be addressed using standard optimization techniques, it is far from simple, because the subsidy affects public goods provision directly (by inducing Non-civics to contribute) and indirectly (by inducing parents to raise their children to be Civics, who always contribute). We simplify the problem substantively in two ways; we return to these two issues in our final section.

First, we assume that while the implementation problem takes account of the full range of motivations affecting the citizens' behavior, the planner's definition of social welfare does not include the image value of contributors or intrinsic motivation of Civics. Second, we abstract from reasons other than the provision of the public good modeled here that may encourage a planner (or society) to promote civic mindedness among citizens.

The planner seeks to maximize the benefits of a public good net of both the costs incurred by the Civics and Non-civics in contributing to the public good, and the costs of the use of the subsidy for affecting the population fraction of Civics. More precisely, we first introduce a net benefit function of the public good ω:

$$\omega(p, s) := \phi(p + (1-p)F(\hat{g}(p, s)) - p\int_0^{\bar{g}} gdF(g) - (1-p)\int_0^{\hat{g}(p,s)} gdF(g) \qquad (25)$$

where the first term on the right-hand side is the value of the public good produced by those contributing either because they are Civics or because the subsidy more than offsets their cost of contribution, while the second and third terms are the costs incurred by the Civics and contributing Non-civics, respectively. As mentioned, the instrument, s, involves costs, $c(s)$, that are increasing and convex: i.e., $c(0) = 0$ and $c'(s)$, $c''(s) \geq 0$ for $s > 0$.

The planner varies s to maximize the social welfare, ω, net of the above cost, $c(s)$, where p is given by the implementation technology $p^*(s)$ satisfying (20). Thus we consider the following maximization problem of the social planner:

$$\max_{s\geq 0,\, p\in[0,1]} \omega(p, s) - c(s) \;\; \text{s.t. } p = p^*(s) \qquad (26)$$

We wish to study the case where an increase in the Civic fraction enhances social welfare, $\frac{\partial \omega}{\partial p}(p, s) > 0$, which requires that the net benefit function of the public good, defined in (25), is increasing in p and that the problem in (26) is well-defined which we ensure by:

A2 We require that $\frac{\partial \omega}{\partial p} > 0$ and that the second order condition for maximization

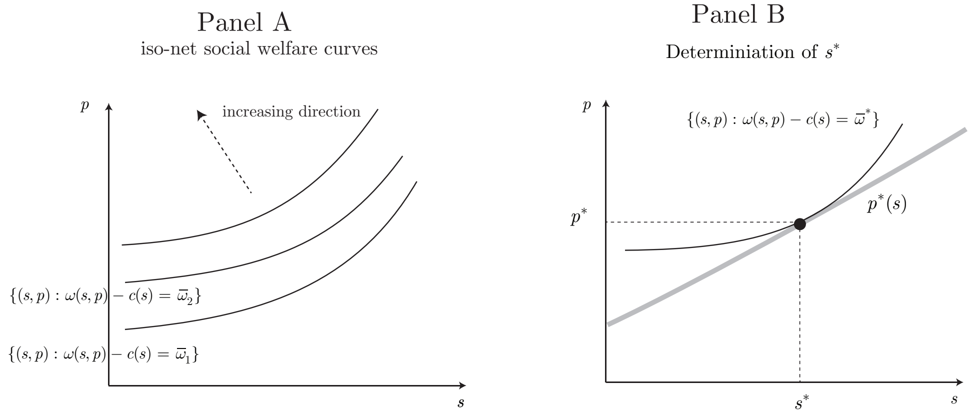

Panel A
iso-net social welfare curves
p
increasing direction
$\{(s,p) : \omega(s,p) - c(s) = \bar{\omega}_2\}$
$\{(s,p) : \omega(s,p) - c(s) = \bar{\omega}_1\}$
s
Panel B
Determiniation of s^*
p
$\{(s,p) : \omega(s,p) - c(s) = \bar{\omega}^*\}$
p^*
$p^*(s)$
s^*
s

Figure 6: **Iso-net social welfare curves and the determination of the optimal subsidy.** Panel A shows iso-net social welfare curves, defined by $\omega(s,p) - c(s) = \bar{\omega}$. The dotted arrow shows the direction in which the social welfare, $\bar{\omega}$, is increasing. In Panel B, we show the implementation function $p = p^*(s)$ as the constraint (the thick gray curve). The optimal choice in the problem of (26) occurs at the point where the iso-net social welfare curve is tangent to the implementation function $p = p^*(s)$.

is satisfied.

To study the problem in (26) in the (s,p) plane (see Figure 6), we define the marginal rate of (negative) substitution (MRS) between the subsidy and Civic fraction, $\sigma(p,s)$, as follows:

$$\sigma(p,s) := -\frac{\frac{\partial \omega}{\partial s}(p,s) - c'(s)}{\frac{\partial \omega}{\partial p}(p,s)}. \tag{27}$$

Similarly, we will call the slope of $p^*(s)$ with respect to s, $\frac{dp^*}{ds}$, a marginal rate of transformation(MRT) of the subsidy into Civic fraction. When $\sigma(p,s)$ is positive, the subsidy s is a "bad" ($c'(s) > \frac{\partial \omega}{\partial s}(p,s)$ meaning that its marginal benefit in raising public goods contributions for a given value of p falls short of its marginal cost). Thus, to leave the social planner indifferent, an increase in the subsidy must be accompanied by a higher fraction of Civics, p. Using the marginal rate of substitution, σ, we can express the first-order condition for the problem in (26) as the following tangency condition:

$$\frac{\partial p^*}{\partial s}(s^*) = \sigma(p^*(s^*), s^*) \tag{28}$$

which requires that an iso-net social benefit locus should be tangent to the implementation function curve (see Panel B in Figure 6) or, as expected, the marginal rate of substitution should equal the marginal rate of transformation.

5. Effect of crowding out on optimal incentives

To identify the effect of crowding out on optimal incentives we need to know what subsidy would the planner have implemented if she had suppressed the adverse effect of the subsidy on the image effect associated with contributing. Call the subsidy in this thought experiment s^N for "no crowding out." We then determine the conditions under which s^N can be greater or smaller than s^*, the optimal subsidy taking account of the crowding-out effect determined in (26).

Suppressing the crowding-out effect alters the effect of the subsidy and thus entails a new implementation function. We let $p^*(s)$ and $p^N(s)$ be the two implementation functions, respectively taking account of and suppressing the crowding-out effect. Then $p^*(s)$ and $p^N(s)$ are defined, respectively, as follows:

$$\begin{aligned} p^* : \qquad & \mu(\hat{g}(p^*, s) - s)(1 - F(\hat{g}(p^*, s))) = \mu \int_{\hat{g}(p^*,s)}^{\bar{g}} (g - s) dF(g) - \frac{m}{1-m} \frac{1}{p^*} \\ p^N : \qquad & \mu(\hat{g}(p^N, 0))(1 - F(\hat{g}(p^N, 0))) = \mu \int_{\hat{g}(p^N,s)}^{\bar{g}} (g - s) dF(g) - \frac{m}{1-m} \frac{1}{p^N} \end{aligned}$$

Observe that p^N is obtained by ignoring the effect of subsidy s on the image effect by setting $s = 0$ in $\hat{g}(p, s)$, but still taking into account the material payoff net of socialization effect. Then, unsurprisingly, it can be shown that (see Appendix D)

$$p^N(s) > p^*(s) \text{ for all } s, \text{ and } \left.\frac{dp^N}{ds}\right|_{s=0} > \left.\frac{dp^*}{ds}\right|_{s=0} \tag{29}$$

that is to say, a given subsidy sustains a higher level of Civics in the population in the absence of the crowding-out effect.

By examining Equation (27), we see that if marginal benefits of the fraction Civics is increasing in the fraction of Civics itself ($\frac{\partial^2 \omega}{\partial p^2} > 0$) and in the subsidy ($\frac{\partial^2 \omega}{\partial s \partial p} > 0$),

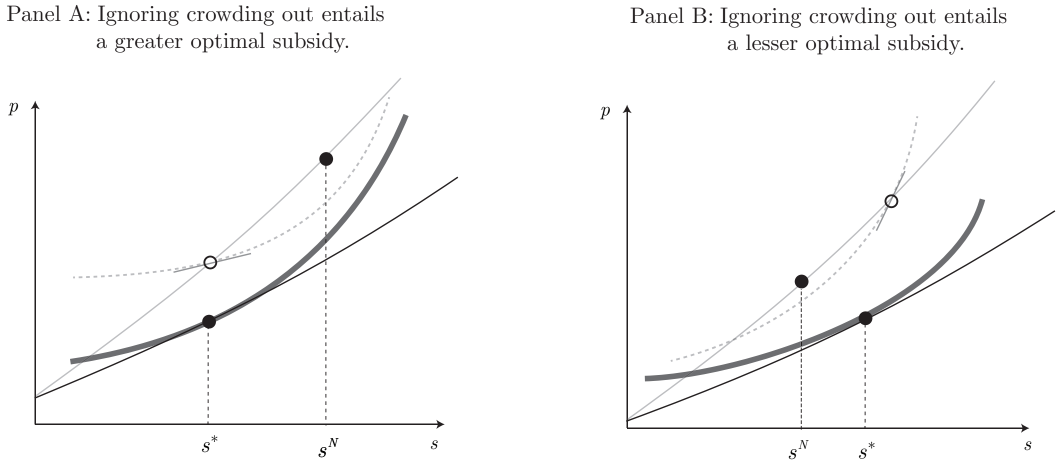

Panel A: Ignoring crowding out entails a greater optimal subsidy.
p
s^*
s^N
s
Panel B: Ignoring crowding out entails a lesser optimal subsidy.
p
s^N
s^*
s

Figure 7: **The effect of crowding out on the optimal subsidy.** Shown in both panels, the implementation function at $s = s^*$ is steeper due to suppressing the crowding-out effect (this need not be the case). In the left panel, the MRS at $s = s^*$ is flatter and in the right panel, it is steeper when no account is taken of the crowding-out effect.

then from $p^N(s) > p^*(s)$,

$$\sigma(p^*(s^*), s^*) > \sigma(p^N(s^*), s^*) \tag{30}$$

holds. In this case, the planner, recognizing the cultural crowding-out effect, would implement a greater subsidy than she would have done had she ignored the negative effect of the subsidy on the image effect.

In Appendix C, we show that if $\phi(\cdot)$ is sufficiently convex (thus, marginal benefits of the fraction Civics are increasing both in the fraction of Civics and in the subsidy ($\frac{\partial^2 \omega}{\partial p^2} > 0$ and $\frac{\partial^2 \omega}{\partial s \partial p} > 0$)), then the inequality in (30) holds. Conversely, if $\phi(\cdot)$ is sufficiently concave, then the opposite inequality in (30) holds.

The commonly expected result—that in the presence of crowding out, the optimal use of the subsidy will be less than in its absence—can be readily illustrated by the following graphical representation of two possible consequences of suppressing the crowding-out effect:

$$\text{flattening the iso-social benefit loci} : \sigma(p^*(s^*), s^*) > \sigma(p^N(s^*), s^*) \tag{31}$$

$$\text{steepening the implementation function } : \frac{dp^N}{ds}(s^*) > \frac{dp^*}{ds}(s^*) \tag{32}$$

when crowding out occurs at s^* and $p^*(s^*)$.

From the graphical examination (Figure 7), it is clear that if suppressing the crowding-out effect makes the subsidy appear to be more effective in increasing p and increases in p to be more valuable, then the planner in this hypothetical thought experiments will implement a larger subsidy than is optimal under crowding out. The conditions in (31) and (32) are sufficient for this to occur.

However, while this is a possible effect of suppressing the crowding-out effect, it does not follow that the optimal subsidy s^* must be less than s^N, the subsidy that would be adopted in the absence of crowding out (Panel B in Figure 7). There are two reasons for this, which we will demonstrate below.

First, in the case that the marginal contribution of the fraction of Civics to social welfare is decreasing (the benefit function is concave in p) the marginal benefit of raising p will be less when crowding out is suppressed because for any given level of $s > 0, p^*(s) < p^N(s)$. This means that having suppressed the crowding-out effect, increases in p will appear to be less valuable and, as a result, the iso social welfare loci may steepen rather than flatten.

Second, the marginal effectiveness of the subsidy need not be greater when crowding out is suppressed; the marginal effectiveness of the subsidy may fall, flattening the implementation function rather than steepening it as might be expected. Suppressing the crowding-out effect does raise the "average product" of the subsidy $\frac{p(s)}{s}$, but it need not increase its marginal effectiveness (that is, the slope of the implementation function, i.e. the MRT).

The reason is that the effect of crowding out on $\frac{dp}{ds}$ depends on the distribution of costs of contributing $F(g)$ and the fraction of Civics that the planner anticipates being in the population, the latter of which differs depending on whether crowding-out effects are suppressed or accounted for. Depending on the cost of the marginal contributor, a reduction in the net cost of contributing (due to the subsidy) may increase the number of Non-civics whose costs are lower than the threshold by more at $p = p^*$ than at $p = p^N$. This would occur for example if at p^* the marginal contributor's cost was below that indicated in Figure 3 while at p^N the reverse were true, so that the $F(g)$ function was steeper when the crowding-out problem is considered than when it is suppressed.

Thus, the conditions for the opposite signs in both (31) and (32) are possible; and should this be the case, it is clearly sufficient for the optimal subsidy under crowding out to exceed the subsidy implemented when crowding is ignored or $s^* > s^N$. Which case is considered depends on which of the two effects in (31) and (32) are greater, as the following proposition shows.

Proposition 3. *Suppose that* ***A2*** *holds. Then $s^* > s^N$ occurs if and only if*

$$\frac{dp^N}{ds}(s^*) - \frac{dp^*}{ds}(s^*) < \sigma(p^N(s^*), s^*) - \sigma(p^*(s^*), s^*) \tag{33}$$

Proof. See Appendix D. □

6. Discussion

In this study, using now-standard models of cultural evolution, we have advanced the idea that the type and extent of a society's use of economic incentives may affect the process of cultural transmission from parents, other elders, or peers, by which individuals acquire new tastes or social norms that will persist over a long period.

We modeled the evolution of preferences, not as a decentralized process resulting from natural selection or uncoordinated parental socialization of the young, but rather as a mechanism design problem that might be addressed as our fictive social planner acting on behalf of a far-sighted religious order, political party, or national government. For this purpose, we extended the standard implementation-by-Nash-equilibrium paradigm to include endogenous preferences, thereby imposing a cultural stationarity condition on the equilibrium outcome.

We characterized these equilibrium preferences in a cultural evolutionary process and showed how they are affected by a subsidy implemented with the intention of altering two types of behavior: citizens contributing to a public good, and parents socializing children to have public-spirited preferences motivating them to contribute unconditionally. Finally, we showed that if the subsidy has an adverse effect on the social esteem value of contributing, and thus diminishing parental efforts to raise public-spirited children—cultural crowding out—the optimal subsidy may be either greater than or less than that in the absence of crowding out.

Here we comment on possible alternative formulations of the problem and relationship to the relevant literature in economics and psychology.

In Section 4, we discussed two alternatives to the social planner maximizing the benefits of the public goods project, net of the costs of provision and of the planner's policies. The first is to include in the planner's maximand the utility experienced as a result of the ethical or social esteem values of citizens when these are subject to modification by public policy. (Recall that in our model, while the planner takes account of the effect of the citizens' civic-minded preferences on their behaviors, the objective function of the planner did not include whatever intrinsic pleasure or other subjective benefits that contributors to the public good may experience as a result of their civic-mindedness.)

Including these subjective effects in the planners' maximand naturally raises difficult philosophical and economic issues (Diamond, 2006; Bergstrom, 2006; Hwang and Bowles, 2014; Chaloupka et al., 2014): should the planner count as a cost the foregone pleasure of the drug high of a once addicted target of intervention? Without taking a position on this difficult question, we have studied the case of the thorough-going utilitarian planner who includes the full range of subjective effects in her maximand(Bowles and Hwang, 2008). Based on this earlier work, we do not believe that including in the planner's objective function the image value and the effect that the subsidy may have in diminishing this would alter the qualitative results we have derived.

The second alternative formulation is to recognize that the planner (or the entity whose objectives she is seeking to advance) might have reasons beyond supporting contributions to this particular public good to value having a substantial fraction of Civics in the citizenry. An example is provided by cases in which contributions to some other public good are not observable by the planner (or are observable only at great cost) and hence cannot be motivated by subsidies. Our sense is that a great many public goods are of this type, including observing informal norms facilitating coordination of everything from traffic to interactions with strangers in public spaces. The planner might value inducing parents to raise their children as Civics as an effective means of ensuring public goods provision in such cases of non-observability. Alternatively, the planner might value a civic-minded population for additional reasons unrelated to public goods provision. While it would be simple to accommodate this more expansive notion of the planners' objectives (by adding a term to $\omega(p, s)$) it would not alter the model qualitatively.

Our analysis of cultural crowding out has used a standard signal extraction model to represent the effects of explicit incentives on the long-run evolution of preferences. However, the fundamental idea could have been motivated independently of Bayesian reasoning.

The basic intuition captured by this model—that the presence of explicit incentives may lead an observer to interpret a contribution to the public good as self-interested rather than socially motivated—is termed the "over justification effect" in psychology. In this literature the subsidy supplies a competing justification for the contribution: "he did it for the money."

The psychologist Mark Lepper and his co-authors, write: "When an individual observes another person engaging in some activity, he infers that the other is intrinsically motivated... to the extent that he does not perceive salient, unambiguous, and sufficient extrinsic contingencies to which to attribute the other's behavior...."(Lepper and Greene, 1982).

The implication is that when these extrinsic contingencies are present—as would be the case in the presence of a subsidy—the attribution of intrinsic motivation—such as really being a Civic—is less likely. There is some neurological evidence that this inference would not be incorrect. The presence or absence of an incentive is associated with a basic shift in cognitive processing: in an experimental Trust Game the introduction of fines to be levied on those who failed to reciprocate a trusting action shifted neural activity to a different brain region(Li et al., 2009). We suspect that developing our model to take greater account of these neurological and other psychological dimensions might yield results beyond those using a simple signal extraction model as we have done.

A related literature in economics considers the behavioral effects of the information that incentives provide (Benabou and Tirole, 2003; Fehr and Rockenbach, 2003; Schotter et al., 1996; Falk and Kosfeld, 2006; Bowles and Hwang, 2008; Hwang and Bowles, 2014; Bowles and Polania Reyes, 2012). In this case, crowding out is behavioral and results from the state-dependence of preferences, the absence, presence, and nature of incentives representing different states.[5] Neither the over-justification literature in psychology nor the economic literature on behavioral crowding out has considered the long-term effects of policy interventions on the evolution of preferences.

Our model of this process is new, but the idea is definitely not. The economic literature on the appropriate use of incentives when preferences are endogenous dates back to Jeremy Bentham's *An Introduction to the Principles of Morals and Legislation* (1789:1907). However, with few exceptions (Hirschman (1985), Aaron (1994), and others cited in Bowles (2016)) economists have not considered the way that incentives affect the process of preference formation.

Oren Bar-Gill and Chaim Ferstman (2005) modeled a case of strong crowding out in which a subsidy for a pro-social action increases the likelihood that altruists will be

[5] Psychologists refer to this mechanism as framing and term the preferences subject to framing as situation dependent (Ross and Nisbett, 1991). We addressed the case of optimal incentives with state dependent preferences in Bowles and Hwang (2008) and Hwang and Bowles (2014).

taken advantage of by non-altruists, leading to a decrease in the fraction of the population that are altruists. Iris Bohnet and her coauthors (2001) modeled the influence of legal policy on the process of preference formation, and found that the effect of incentives on the evolution of a preference for contract performance is non-monotonic. Felix Bierbrauer, Axel Ockenfels and their co authors (2017) introduced the notion of a social-preference-robust mechanism, which can perform well, irrespective of specific assumptions about the nature and intensity of selfish and social preferences. None of these important studies (nor any others, to our knowledge) addresses the question of optimal subsidies or taxes where preferences are endogenous.

The policy advice given here by our sophisticated planner has been based on her recognition that an incentive may alter the cultural transmission process so as to diminish the incentives of parents to inculcate their children to have public-spirited preferences. But she has taken as given the extent of cultural crowding out.

A super-sophisticated planner would not stop at simply taking account of crowding out; she would let the extent of crowding out itself be a target for public policy manipulation. An example is framing material incentives as prizes, a mechanism that was adopted by Athenians at the time of Aristotle to mobilize both the material interests and the moral sentiments of the citizens to support contributions to public goods (Ober (2008):124-134).

Acknowledgments: We would like to thank (to be completed) for their contributions to this research. We would also like to thank the Behavioral Sciences Program of the Santa Fe Institute, the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2016S1A5A8019496), and the U.S. National Science Foundation for support on this project.

Appendix

A. Alternative derivation of the cultural evolution equation in (15)

The story (derivation) is based on Bisin and Verdier (2001). Suppose that a family consists of a parent and a child and a child is born without a trait, Civic or Non-civic. Within family a child, who is exposed to a parent with type i, is raised as type i with probability d_i, for $i = C, N$ (called a vertical or direct cultural transmission). We assume that for a Civic parent, d_C is $\mu\pi_C(p)$ and for a Non-civic parent, d_N is $\mu\pi_N(p)$, where we assume that π_C and π_N are positive and μ ensures the whole expressions are less than 1. If a child from a parent with type i fails to be raised as type i, then he or she picks the type of the role model chosen randomly from the population (oblique transmission). This probability that a child adopts trait C (or trait N, resp.) from the population is given by p or $(1-p)$. This setting gives the transition probabilities that a child from a family with type i become trait i (or j):

$$P_{ii} = d_i + (1-d_i)p, \quad P_{ij} = (1-d_i)(1-p),$$

for $i = C, N$. More specifically, we find that

$$P_{CC} = d_C + (1-d_C)p, P_{CN} = (1-d_C)(1-p), P_{NN} = d_N + (1-d_N)(1-p), P_{NC} = (1-d_N)p.$$

and using the accounting equality $\Delta p = (1-p)P_{NC} - pP_{CN}$, $d_C = \mu\pi_C(p)$, and $d_N = \mu\pi_N(p)$ we find that $\Delta p = (1-p)p\mu(\pi_C - \pi_N)$, which is Equation (15).

B. The existence and uniqueness of stable stationary preferences

In this appendix, we provide the conditions for the existence and uniqueness of stable stationary preferences and some sufficient conditions under which the domain for the implementation tools, subsidy and socialization, is well-defined (see Proposition 4).

B.1. Proof of Proposition 1

We let

$$\Phi(p) := (1-m)\mu p(1-p)\int_{\hat{g}}^{\bar{g}}(\hat{g}-g)dF(g)+m(1-p)$$

Then we have $dp/dt = \Phi(p)$ and $\Phi(0) = m > 0$ and $\Phi(1) = 0$. Thus if $\Phi'(1) > 0$, there exists a stable stationary preference since Φ is a continuous function. Then we find

$$\Phi'(p) = (1-m)\mu(1-2p)\int_{\hat{g}}^{\bar{g}}(\hat{g}-g)dF(g)+(1-m)\mu p(1-p)(1-F(\hat{g}))\frac{\partial\hat{g}}{\partial p}-m \tag{B.1}$$

and thus if

$$\Phi'(1) = -(1-m)\mu\int_{s+1}^{\bar{g}}(g-(s+1))dF(g)-m > 0, \tag{B.2}$$

there exists an interior stable stationary preference, where we use $\hat{g}(1,s) = s+1$. Thus if (B.2) holds, then there exists $p^* \in [0,1]$ such that $\Phi'(p^*) < 0$. By rearranging (B.2), we obtain (21) in the text. Also recall that we define κ in (22):

$$\begin{aligned}\kappa(p,s,m) :=& (1-m)\mu((\hat{g}-s)(1-F(\hat{g}))+\int_{\hat{g}}^{\bar{g}}(s-g)dF(g))+\frac{m}{p}\\ =&(1-m)\mu\int_{\hat{g}}^{\bar{g}}(\hat{g}-g)dF(g)+\frac{m}{p}\end{aligned}$$

Then we find that

$$\frac{\partial\kappa}{\partial p} = (1-m)\mu(1-F(\hat{g}))\frac{\partial\hat{g}}{\partial p}-\frac{m}{p^2} \tag{B.3}$$

Since $\Phi(p^*) = 0$ or equivalently $(1-m)\mu\int_{\hat{g}}^{\bar{g}}(\hat{g}-g)dF(g) = -\frac{m}{p^*}$, using (B.1) and (B.3) we find that

$$\Phi'(p^*) < 0 \text{ if and only if } \left.\frac{\partial\kappa}{\partial p}\right|_{p=p^*} < 0$$

B.2. Uniqueness

For the stable stationary preference to be unique, we need the following condition:

$$\Phi''(p) > 0$$

The expression for $\Phi''(p)$ is generally complicated and is given by

$$\begin{aligned}\Phi''(p) =& 2(1-m)\mu \int_{\hat{g}}^{\bar{g}} (g-\hat{g})dF(g) + 2(1-m)\mu(1-2p)(1-F(\hat{g}))\frac{\partial \hat{g}}{\partial p} \\ &+ (1-m)\mu p(1-p)(-F'(\hat{g}))(\frac{\partial \hat{g}}{\partial p})^2 + (1-m)\mu p(1-p)(1-F(\hat{g}))(\frac{\partial^2 \hat{g}}{\partial p^2}) > 0\end{aligned} \tag{B.4}$$

Then using the expressions for derivatives in Appendix C, we can check the condition in (B.4).

B.3. Uniform Distribution Case

Now suppose that $F(g) \sim [0, \bar{g}]$. Then we find the existence condition as follows:

$$\mu(\frac{(\bar{g}-(s+1))^2}{2\bar{g}}) > \frac{m}{1-m} \tag{B.5}$$

We have the following result.

Proposition 4. *Suppose that $F(g) \sim$ **Unif**$[0, \bar{g}]$. Let $\bar{m} < 1$ and $\bar{s}$ be given. Then there exists $\bar{\mu}$ such that for all $\mu > \bar{\mu}$, for all $0 \leq s \leq \bar{s}$ and for all $0 \leq m \leq \bar{m}$ there exists a stable stationary state.*

Proof. Choose $\bar{\mu}$ such that

$$\bar{\mu} > \frac{\bar{m}/(1-\bar{m})}{\frac{(\bar{g}-(\bar{s}+1))^2}{2\bar{g}}}$$

Then we find that

$$\mu(\frac{(\bar{g}-(s+1))^2}{2\bar{g}}) > \bar{\mu}(\frac{(\bar{g}-(\bar{s}+1)^2}{2\bar{g}}) > \frac{\bar{m}}{1-\bar{m}} \geq \frac{m}{1-m}$$

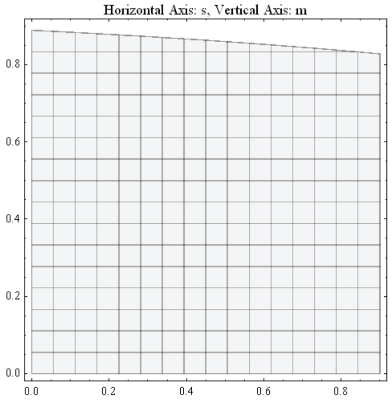


Figure B.8: **The example of $\bar{s}$ and $\bar{m}$ in Proposition 4.** The shaded region shows the area in which the condition in Proposition 4 is satisfied for s and m. We can choose $\bar{s} = 0.9$ and $\bar{m} = 0.75$ in Proposition 4. We also checked that $\kappa''(p) > 0$ for all $0 < p < 1$ and for all s, m. We use $\bar{g} = 5, \mu = 5$.

and the condition in (21) is satisfied. □

The following figure shows the range of s and m such that the unique stable stationary state exists.

C. Derivatives

Here we find various expressions for derivatives. To do this, we define the following function:

$$\varpi(p, \hat{g}) := \frac{p}{p + (1-p)F(\hat{g})}$$

Then $\hat{g}(p, s)$ satisfies

$$\varpi(p, \hat{g}(p, s)) + \theta s = \hat{g}(p, s)$$

Then we find that

$$\frac{\partial \varpi}{\partial \hat{g}} = -\varpi(1-\varpi)\frac{F'(\hat{g})}{F(\hat{g})} < 0, \quad \frac{\partial \varpi}{\partial p} = \frac{\varpi(1-\varpi)}{p(1-p)}$$

and

$$\begin{aligned}\frac{\partial^2\varpi}{\partial\hat{g}\partial p} =&\frac{1}{p(1-p)}(2\varpi-1)\varpi(1-\varpi)\frac{F'(\hat{g})}{F(\hat{g})}\\ \frac{\partial^2\varpi}{\partial\hat{g}^2} =&-(2\varpi-1)\varpi(1-\varpi)(\frac{F'(\hat{g})}{F(\hat{g})})^2-\varpi(1-\varpi)\frac{F''(\hat{g})F(\hat{g})-(F'(\hat{g}))^2}{(F(\hat{g}))^2}\end{aligned}$$

Using these we find that

$$\frac{\partial\hat{g}}{\partial p}=\frac{\partial\varpi/\partial p}{1-\partial\varpi/\partial\hat{g}}>0,\quad 0<\frac{\partial\hat{g}}{\partial s}=\frac{1}{1-\partial\varpi/\partial\hat{g}}<1$$

And

$$\frac{\partial^2\hat{g}}{\partial p\partial s}=\frac{\frac{\partial^2\varpi}{\partial\hat{g}\partial p}+\frac{\partial^2\varpi}{\partial\hat{g}^2}\frac{\partial\hat{g}}{\partial p}}{(1-\frac{\partial\varpi}{\partial\hat{g}})^2},\quad \frac{\partial^2\hat{g}}{\partial s^2}=\frac{\frac{\partial^2\varpi}{\partial\hat{g}^2}\frac{\partial\hat{g}}{\partial s}}{(1-\frac{\partial\varpi}{\partial\hat{g}})^2}$$

Next recall that

$$\omega(p,s):=\phi(q(p,s))-p\int_0^{\bar{g}}gdF(g)-(1-p)\int_0^{\hat{g}(p,s)}gdF(g)$$

where $q(p,s)=p+(1-p)F(\hat{g}(p,s))$. We find that

$$\begin{aligned}\frac{\partial\omega}{\partial p} =&\phi'(q(p,s))\frac{\partial q}{\partial p}(p,s)-\int_0^{\bar{g}}gdF(g)+\int_0^{\hat{g}(p,s)}gdF(g)-(1-p)\hat{g}(p,s)F'(\hat{g}(p,s))\frac{\partial\hat{g}}{\partial p}(p,s)\\ \frac{\partial\omega}{\partial s} =&\phi'(q(p,s))\frac{\partial q}{\partial s}(p,s)-(1-p)\hat{g}(p,s)F'(\hat{g}(p,s))\frac{\partial\hat{g}}{\partial s}(p,s)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2\omega}{\partial p^2} &= \phi''(q)(\frac{\partial q}{\partial p})^2 + \phi'(q)\frac{\partial^2 q}{\partial p^2}\\ &\quad + 2\hat{g}F'(\hat{g})\frac{\partial \hat{g}}{\partial p} - (1-p)\hat{g}F'(\hat{g})(\frac{\partial \hat{g}}{\partial p})^2 - (1-p)\hat{g}F''(\hat{g})(\frac{\partial \hat{g}}{\partial s})^2 - (1-p)\hat{g}F'(\hat{g})\frac{\partial^2 \hat{g}}{\partial p^2}\\ \frac{\partial^2\omega}{\partial s\partial p} &= \phi''(q)\frac{\partial \hat{q}}{\partial p}\frac{\partial q}{\partial s} + \phi'(q)\frac{\partial^2 \hat{q}}{\partial s\partial p}\\ &\quad + \hat{g}F'(\hat{g})\frac{\partial \hat{g}}{\partial s} - (1-p)F'(\hat{g})(\frac{\partial \hat{g}}{\partial s})^2 - (1-p)\hat{g}F''(\hat{g})(\frac{\partial \hat{g}}{\partial s})^2 - (1-p)\hat{g}F'(\hat{g})\frac{\partial^2 \hat{g}}{\partial s^2}\end{aligned}$$

where we suppress the arguments of functions in the second order derivatives. Since $(\frac{\partial q}{\partial p})^2 > 0$ and $\frac{\partial \hat{q}}{\partial p}\frac{\partial q}{\partial s} > 0$, we see that if ϕ'' is positive and sufficiently large , then $\frac{\partial^2\omega}{\partial p^2} > 0$ and $\frac{\partial^2\omega}{\partial s\partial p} > 0$ and if ϕ'' is negative and sufficiently small, then $\frac{\partial^2\omega}{\partial p^2} < 0$ and $\frac{\partial^2\omega}{\partial s\partial p} < 0$.

D. the proof of Proposition 3 and results in (29)

Proof of Proposition 3. Recall that s^N and s^* satisfy the following FOCs:

$$\begin{aligned} s^N: &\quad \frac{\partial w}{\partial p}(p^N(s^N), s^N)\frac{dp^N}{ds}(s^N) + \frac{\partial w}{\partial s}(p^N(s^N), s^N) - c'(s^N) = 0\\ s^*: &\quad \frac{\partial w}{\partial p}(p^*(s^*), s^*)\frac{dp^*}{ds}(s^*) + \frac{\partial w}{\partial s}(p^*(s^*), s^*) - c'(s^*) = 0\end{aligned}$$

We define

$$\Psi(s) := \frac{\partial w}{\partial p}(p^N(s), s)\frac{dp^N}{ds}(s) + \frac{\partial w}{\partial s}(p^N(s), s) - c'(s).$$

Since $\Psi(s^N) = 0$ and $\Psi'(s) < 0$ (by SOC), we have

$$\Psi(s^*) < 0 \iff s^* > s^N.$$

Then we find that

$$\begin{aligned}\Psi(s^*) =&\frac{\partial w}{\partial p}(p^N(s^*), s^*)\frac{dp^N}{ds}(s^*) + \frac{\partial w}{\partial s}(p^N(s^*), s^*) - c'(s^*)\\ =&\frac{\partial w}{\partial p}((p^N(s^*), s^*)[\frac{dp^N}{ds}(s^*) - \sigma(p^N(s^*), s^*)]\\ <&\frac{\partial w}{\partial p}((p^N(s^*), s^*)[\frac{dp^*}{ds}(s^*) - \sigma(p^*(s^*), s^*)] = 0\end{aligned}$$

since $\frac{\partial w}{\partial p}((p^N(s^*), s^*) > 0$. Similarly, we find that $\Psi(s^*) > 0$ if the opposite inequality in (33) holds.

□

Proof of results in (29). Let ι and η be the image effect and the non-image effect, respectively. That is,

$$\iota(p, s) = \mu(\hat{g}(p, s) - s)(1 - F(\hat{g}(p, s))), \quad \eta(p, s) = \mu\int_{\hat{g}(p,s)}^{\bar{g}}(g - s)dF(g) - \frac{m}{1-m}\frac{1}{p} \tag{D.1}$$

Then we have

$$\iota(p^*(s), s) = \eta(p^*(s), s), \;\; \iota(p^N(s), 0) = \eta(p^N(s), s)$$

Then because of the stability condition, we have $\frac{\partial \iota}{\partial p}(p^*, s) < \frac{\partial \eta}{\partial p}(p^*, s)$. Thus if we can show that $\iota(p, s)$ is decreasing in s, then we have $p^N(s) > p^*(s)$. This follows from that

$$0 < \frac{\partial \hat{g}}{\partial s}(p, s) < 1$$

and $\hat{g} - s > 0$ and $1 - F(\hat{g}) > 0$.

To study the sign of $\frac{dp^*}{ds} - \frac{dp^N}{ds}$, we differentiate (D.1) and find that

$$\frac{dp^*}{ds} \lesseqgtr \frac{dp^N}{ds} \iff \frac{-\frac{\partial\eta}{\partial s}(p^*(s),s) + \frac{\partial\iota}{\partial s}(p^*(s),s)}{\frac{\partial\eta}{\partial p}(p^*(s),s) - \frac{\partial\iota}{\partial p}(p^*(s),s)} \lesseqgtr \frac{-\frac{\partial\eta}{\partial s}(p^N(s),s)}{\frac{\partial\eta}{\partial p}(p^N(s),s) - \frac{\partial\iota}{\partial p}(p^N(s),0)} \tag{D.2}$$

Then at $s = 0$

$$\frac{\partial\eta}{\partial p}(p^*(s),s) - \frac{\partial\iota}{\partial p}(p^*(s),s) = \frac{\partial\eta}{\partial p}(p^N(s),s) - \frac{\partial\iota}{\partial p}(p^N(s),0) \text{ and } -\frac{\partial\eta}{\partial s}(p^*(s),s) = -\frac{\partial\eta}{\partial s}(p^N(s),s)$$

Thus $\frac{\partial\iota}{\partial s} < 0$ (crowding out effect) implies that

$$\left.\frac{dp^*}{ds}\right|_{s=0} < \left.\frac{dp^N}{ds}\right|_{s=0}$$

□

References

Aaron, H., 1994. Public policy, values, and consciousness. Journal of Public Economic 8, 3–21.

Ali, S. N., Lin, C., 2013. Why people vote: Ethical motives and social incentives. American Economic Journal: Microeconomics 5, 73–98.

Bar-Gill, O., Fershtman, C., 2005. Public policy with endogenous preferences. Journal of Public Economic Theory 7 (5), 841–857.

Benabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. Review of Economic Studies 70 (33), 489–520.

Benabou, R., Tirole, J., 2006. Incentives and prosocial behavior. American Economic Review 96 (5), 1652–78.

Bentham, J., 1789:1907. An Introduction to the Principles of Moral and Legislation. Oxford University Press.

Bergstrom, T., 2006. Benefit-cost in a benevolent society. American Economic Review 96 (1), 339–51.

Bierbrauer, F., Ockenfels, A., Pollak, A., Rückert, D., 2017. Robust mechanism design and social preferences. Journal of Public Economics 149, 59–80.

Bisin, A., Verdier, T., 2001. The economics of cultural transmission and the dynamics of preferences. Journal of Economic Theory 97, 298–319.

Bisin, A., Verdier, T., 2011. The economics of cultural transmission and socialization. In: Benhabib, J., Bisin, A., Jackson, M. (Eds.), Handbook of Social Economics. Elsevier Science.

Bohnet, I., Frey, B. S., Huck, S., 2001. More order with less law: On contractual enforcement, trust, and crowding. American Political Science Review 95, 131–44.

Bowles, S., 1998. Endogenous preferences: The cultural consequences of markets and other economic institutions. Journal of Economic Literature 36, 75–111.

Bowles, S., 2004. Microeconomics. Princeton University Press.

Bowles, S., 2016. The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens. Yale University Press.

Bowles, S., Hwang, S.-H., 2008. Social preference and public economics: Mechanism design when preferences depend on incentives. Journal of Public Economics 92 (8-9), 1811–20.

Bowles, S., Polania Reyes, S., 2012. Economic incentives and social preferences: Substitutes or complements. Journal of Economic Literature 50, 368–425.

Boyd, R., Richerson, P. J., 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.

Camerer, C., 2003. Behavioral Game Theory: Experimental Studies of Strategic Interaction. Princeton University Press.

Cavalli-Sforza, L. L., Feldman, M. W., 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton, N.J.

Chaloupka, F. J., Warner, K. E., Acemoglu, D., Gruber, J., Laux, F., Max, W., Newhouse, J., Schelling, T., Sindelar, J., 2014. An evaluation of the FDA's analysis of the costs and benefits of the graphic warning label regulation. Tobacco Control 0, 1–8.

Diamond, P., 2006. Optimal tax treatment of private contributions for public goods with and without warm glow preferences. Journal of Public Economics 90, 897–919.

Falk, A., Kosfeld, M., 2006. The hidden costs of control. American Economic Review 96 (5), 1611–30.

Fehr, E., Gaechter, S., 2000. Fairness and retaliation: The economics of reciprocity. Journal of Economic Perspectives 14, 159–81.

Fehr, E., Rockenbach, B., 2003. Detrimental effects of sanctions on human altruism. Nature 422 (13), 137–40.

Gneezy, U., Rustichini, A., 2000. A fine is a price. Journal of Legal Studies 29 (1), 1–17.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Hill, K., Gil-White, F., Gurven, M., Marlowe, F., Patton, J. Q., Smith, N., Tracer, D., 2005. Economic man in cross-cultural perspective: Behavioral experiments in fifteen small-scale societies. Behavioral and Brain Sciences 28, 795–815.

Hirschman, A. O., 1985. Against parsimony: Three ways of complicating some categories of economic discourse. Economics and Philosophy 1, 7–21.

Hwang, S.-H., Bowles, S., 2014. Optimal incentives with state-dependent preferences. Journal of Public Economic Theory 16 (5), 681–705.

Lepper, Mark R., G. S. J. D., Greene, D., 1982. Consequences of superfluous social constraints: Effects on young children's social inferences and subsequent intrinsic interest. Journal of Personality and Social Psychology 42, 51–65.

Li, J., Xiao, E., Houser, D., Montague, P. R., 2009. Neural responses to sanction threats in two-party economic exchanges. Proceedings of the National Academy of Science 106 (39), 16835–16840.

Loewenstein, George F., L. T., Bazerman, M. H., 1989. Social utility and decision making in interpersonal contexts. Journal of Personality and Social Psychology 57, 426–41.

Ober, J., 2008. Democracy and Knowledge: Innovation and Learning in Classical Athens. Princeton University Press.

Ross, L., Nisbett, R. E., 1991. The Person and the Situation: Perspectives of Social Psychology. Temple University Press, Philadelphia.

Schotter, A., Weiss, A., Zapater, I., 1996. Fairness and survival in ultimatum and dictatorship games. Journal of Economic Behavior and Organization 31, 37–56.

Weibull, J., 1995. Evolutionary Game Theory. MIT Press.