Interventions with Sticky Social Norms: A Critique[☆]

Rohan Dutta[1], David K. Levine[2], Salvatore Modica[3]

Abstract

We study the consequences of policy interventions when social norms are endogenous but costly to change. In our environment a group faces a negative externality that it partially mitigates through social norms enforced through peer pressure. In this setting policy interventions can have unexpected consequences. When the cost of norm redesign is high introducing a Pigouvian tax can increase output and when the cost of norm redesign is low intervention may lead instead to tax repeal.

JEL Classification Numbers: A1, D7, D9

Keywords: endogenous social norms, natural experiments, randomized control trials

[☆] First Version: October 14, 2017. We would like to thank Heski Bar-Isaac, Marco Casari, Andrea Ichino, Andrea Mattozzi, Rohini Somanathan and seminar audiences at WUSTL, Warwick, Queen's, the Paris Institutions Conference, the Zurich Political Economy Conference, Delhi School of Economics and the University of Trento. We gratefully acknowledge support from the EUI Research Council.

[*] Corresponding author David K. Levine

Email addresses: `rohan.dutta@mcgill.ca` (Rohan Dutta), `david@dklevine.com` (David K. Levine), `salvatore.modica@unipa.it` (Salvatore Modica)

[1] Department of Economics, McGill University

[2] Department of Economics, EUI and WUSTL

[3] Università di Palermo, Dipartimento SEAS

1. Introduction

We study the consequences of policy interventions in an environment where social norms are endogenous but sticky. The environment is one in which a group engages in production that generates a negative externality. Following Olson (1965) and Ostrom (1990) peer pressure is used to mitigate this externality. Following Townsend (1994) and Levine and Modica (2016) we model the endogeneity of the norms as a mechanism design problem for the group. This setup has a distinctively Coasian flavor as the group is able partially to contend with externalities on its own. The new feature we introduce is the idea that redesigning social norms is costly: this introduces a stickiness in which social norms may be maintained when they are no longer optimal.[4]

We study a simple environment with two periods. In the first period the group designs a norm (mechanism) anticipating the second period will likely be the same as the first. In the second period an unanticipated intervention may take place - for example, the introduction of a Pigouvian tax. If it does the group may, at a cost, design a new norm to cope with changed circumstances. It may at no cost choose to maintain the existing norm. Finally, at low cost, it may simply abandon any effort to police itself and revert to the "law of the jungle." Our environment is constructed so that absent any social norm - with purely individualistic behavior - increases in the size of the intervention always reduce output.

What we find is this. If the size of the intervention is small the group does not respond at all. There is a threshold at which output jumps. If the cost of norm redesign is small output jumps down with the new norm and remains lower than in the first period. This is the same as we would expect if individuals faced adjustment costs as in the widely used menu cost model of Calvo (1983). However if the cost of norm redesign is large output jumps up - with the old norm abandoned and not replaced by a new one - and then declines eventually becoming lower than in the first period. Here as the intervention increases the first period norm becomes increasingly dysfunctional until it is better simply to revert to the law of the jungle. This is a counterintuitive outcome: output moves in the wrong direction in response to an intervention. We argue that two anomalous experimental/empirical results in the literature accord well with this model: the field experiment of Gneezy and Rustichini (2000) in which parents respond to a fine for picking up children late from day-care by picking them up even later; and the natural experiment of Card and Krueger (1994) in which fast food employment rose in response to an increase in the minimum wage. In particular, we believe that the cost of norm redesign was relatively high because these experiments were conducted before the advent of social media.

In the Pigouvian tax setting we examine the consequences of whether or not the tax is rebated lump sum to the group. In standard theory this makes no difference to behavior. Here it does: in particular if an outside agency intervenes to set a naive Pigouvian tax and keeps the proceeds there

[4] Levine (2012) gives evidence that social norms change very quickly when incentives for such a change are strong, while Bigoni et al (2016) and Dell et al (2018) give evidence that social norms can be sticky when incentives for change are weak.

will be underproduction. We also study welfare. If the group keeps the tax revenue - which we believe is effectively the case in the school setting of Gneezy and Rustichini (2000) - and production increases, this is evidence of a welfare improvement: it means the policy is a success notwithstanding the increase in production it brings about.

Finally, we study the attitude of the group towards Pigouvian taxes. Even if the outside agency keeps all the tax revenue the group generally prefers a non-zero tax rate, but less than the Pigouvian level. Consequently when the cost of norm redesign is low - as we believe it is today with the advent of social media - an increase in the tax to the Pigouvian level may prompt the group to reorganize. There is, however, a catch: a group unlike an individual may organize not only to adjust output, but may have options to repeal the tax. We argue that this is what was documented by Boyer et al (2019) when a modest decrease in the speed-limit in France in 2018 resulted in the destruction of most of the speed cameras in that country.

2. The Model

We consider a large organized group over two periods $t = 1, 2$. In each period identical group members $i \in [0, 1]$ engage in production choosing a real valued level of output $X \geq x_t^i \geq 0$. The utility of a member i in period t depends upon their common real valued public characteristics $\omega_t \geq 0$, their own output, and the average output of the group $\overline{x}_t = \int x_t^i di$ according to a smooth function $u(\omega_t, \overline{x}_t, x_t^i)$. We always assume that as function of individual output x_t^i utility is well-behaved in the sense that it is strictly differentiably concave.[5] In the second period there are two possibilities: it may be the same as the first period with $\omega_2 = \omega_1$, or an *intervention* may take place in which case $\omega_2 > \omega_1$ with $\omega_2 < \overline{\omega}$. At the beginning of the second period it is known whether or not an intervention has taken place. Our focus will be on the case where the chance of intervention is *a priori* regarded as low, that is, the intervention is "unanticipated."

The presence of $\overline{x}_t$ represents an externality that we assume is negative. Because of the externality the group collectively faces a mechanism design problem, and we assume that incentives can be given to group members in the form of individual punishments based on monitoring of individual behavior: the group can set production quotas y_t^i, receives signals of whether or not these quotas were violated and based on these signals it can impose punishments.[6] Specifically monitoring generates a noisy signal $z_t^i \in \{0, 1\}$ of whether member i exceeded the quota ($x_t^i > y_t^i$) where 0 means "good, respected the quota" and 1 means "bad, exceeded the quota." If the quota was honored ($x_t^i \leq y_t^i$) the bad signal occurs with probability $\pi > 0$; if the quota was violated the probability of the bad signal is higher $\pi_1 > \pi$. We define the *monitoring difficulty* as $\theta = \pi/(\pi_1 - \pi)$. There

[5] For functions of a single variable $f(x)$ we denote derivatives as $f'(x), f''(x)$ and so forth. For functions of several variables $f(x, y)$ we denote partial derivatives by $D_1 f(x, y) \equiv \partial f/\partial x, D_{12} f(x, y) \equiv \partial^2 f/\partial x \partial y$ and so forth. By strict differentiable concavity we mean that the second derivative is strictly negative, or in the multi-variate case that the matrix of second derivatives is negative definite - this is sufficient, but not necessary, for strict concavity. We shall also say that a solution is "weakly interior" when upper and lower bounds do not strictly bind.

[6] Note that in a large group with noisy signals collective punishments, such as price wars, are useless. See Fudenberg, Levine and Pesendorfer (1998).

is limited commitment so that punishments must take place in the period in which the signal is received, and when the signal is bad[7] the group imposes an endogenous utility penalty of P_t^i. This may be in the form of social disapproval or even in the form of monetary penalties.[8]

The tools available for mechanism design, in other words, consist of quotas y_t^i, together with punishments for a bad signal P_t^i. The overall period t utility of a member i who abides by the quota ($x_t^i \leq y_t^i$) is therefore $u(\omega_t, \overline{x}_t, x_t^i) - \pi P_t^i$ and for one who violates the quota ($x_t^i > y_t^i$) is $u(\omega_t, \overline{x}_t, x_t^i) - \pi_1 P_t^i$. These utilities define a game for the group members. If the mechanism designer chooses $\left(y_t^i, P_t^i\right)$ we denote by $X\left(y_t^i, P_t^i\right)$ the set of pure strategy Nash equilibria of this game. In the Appendix Theorem 7 we show that $X\left(y_t^i, P_t^i\right)$ is closed and non-empty. We refer to a triple $\left(x_t^i, y_t^i, P_t^i\right)$ with $\left(x_t^i\right) \in X\left(y_t^i, P_t^i\right)$ as an *incentive compatible social norm.* If a social norm issues no punishments ($P_t^i = 0$) we call it a *default social norm.* The mechanism designer is benevolent and cares about average expected utility, so receives period t utility from a social norm $\left(x_t^i, y_t^i, P_t^i\right)$ of

$$W\left(x_t^i, y_t^i, P_t^i\right) \equiv \int \left[u(\omega_t, \int x_t^i di, x_t^i) - \mathbf{1}[x_t^i > y_t^i]\pi_1 P_t^i - \mathbf{1}[x_t^i \leq y_t^i]\pi P_t^i\right] di.$$

For any type ω_t and average output $\overline{x}$ we say $x^b(\omega_t, \overline{x})$ is a best response if it is a maximizer of $u(\omega_t, \overline{x}, x^i)$. It is shown in the Appendix, Theorem 2, that it is sufficient to restrict attention to incentive compatible common quotas y_t and that the optimal quota is the unique solution of

$$\max u(\omega_t, y_t, y_t) - \theta\left[u(\omega_t, y_t, x^b(\omega_t, y_t)) - u(\omega_t, y_t, y_t)\right].$$

The term subtracted from individual utility represents the *monitoring cost* due to the need for the quota to be incentive compatible. We denote by $\overline{y}^c(\omega_t)$ the solution to this problem.

Adjustment Costs and the Mechanism Design Problem

In the initial period $t = 1$ the group solves the mechanism design problem of choosing an incentive compatible *simple social norm* $\left(x_1^i, y_1^i, P_1^i\right)$ as if the second period will be the same as the first. As there is limited commitment and no connection between the two periods, this amounts to ignoring the second period and maximizing period 1 designer utility over incentive compatible social norms. We denote by $\overline{y}^s = \overline{y}^c(\omega_1)$ the resulting quota.

In period 2 after observing whether or not there is an intervention the mechanism designer maximizes second period utility, but there are now three possibilities:

1. The initial design $\left(y_1^i, P_1^i\right)$ can be costlessly maintained, with the designer choosing any $\left(x_2^i\right)$

[7] In principle punishments could be issued even for a good signal: as incentives depend only on the difference in punishment between the good and bad signal and punishments are costly this will not be part of an optimal mechanism, so for notational simplicity we rule it out.

[8] In principle the group might also impose downward quotas against underproduction. In the Appendix we show that because the externality is negative if there is any cost associated with this then the group strictly prefers not to do so, so for notational simplicity we do not consider this possibility in the text.

such that $\left(x_2^i, y_1^i, P_1^i\right)$ is incentive compatible.

2. For a fixed cost of $f \geq 0$ an incentive compatible default social norm $\left(x_2^i, y_2^i, 0\right)$ may be chosen.
3. For a fixed cost of $F > f$ an arbitrary incentive compatible social norm $\left(x_2^i, y_2^i, P_2^i\right)$ may be chosen.

The fixed costs of adjustment are in the spirit of menu costs in the macroeconomic literature as in Calvo (1983).[9] Here our basic presumption, subsequently to be reflected in a specific assumption, is that f is small while F need not be. The idea is that reverting to an incentive compatible default social norm is a decentralized decision in the spirit of the ethical voter model or rule utilitarianism:[10] if it is evident that the default social norm is superior to the alternatives there is no need to get together to discuss this and reach an agreement, implicitly everyone has agreed in advance that in this case they will all go their own way. By contrast developing a new social norm cannot be decentralized and the group must be reconvened to reach a collective agreement on the new norm.

To be specific about f let M denote the monitoring cost of implementing the simple norm:

$$M \equiv \theta \left[u(\omega_1, \overline{y}^s, x^b(\omega_1, \overline{y}^s)) - u(\omega_1, \overline{y}^s, \overline{y}^s) \right].$$

It is shown in the Appendix that $M > 0$. The assumption on f, maintained throughout, is that $f < M$.

Costly Contemplation. In models of costly adjustment of plans the question always arises: why not plan for the contingency in advance? Instead of choosing a simple social norm $\left(x_1^i, y_1^i, P_1^i\right)$ in period one and waiting to see if there is reason to change in period two, why not also choose at the same time a plan $\left(x_2^i, y_2^i, P_2^i\right)$ conditional on whether or not there is an intervention? In this case there would be no stickiness.

This issue is not a new one: it is closely connected to the literature on incomplete contracts and on rational inattention. The incomplete contracting literature, such as Hart and Moore (1988), deals with a situation where it is expensive to specify a contingency in a way that can be enforced in court. The situation here is different in that the agreement is informal, so it is enough that everyone understands what the contingencies are. The literature on rational inattention, such as Sims (2003), recognizes that it is costly to acquire information about the right decision in the second period in order to make a plan. This differs from our model in one important respect: in a rational inattention model there is information, albeit noisy, about what the second period will be like and it will in general be optimal in the first period not to choose a simple social norm that is optimal in the absence of intervention but rather to hedge a little and choose a social norm that would do a little better in case of an intervention and a little less well in case of no intervention. We think that this type of hedging is costly because it requires active contemplation of what the future is like.

[9] The fixed costs might well depend on the size of the group: for example Levine and Modica (2017) assume it is proportional to group size. Here we are keeping the size of the population fixed.

[10] See, for example, Fedderson and Sandroni (2006).

Our model is one of unawareness in the sense of Modica and Rustichini (1994) and in the spirit of Tirole (2009) and Dye (1985). It is based on the idea of costly contemplation studied by Ergin and Sarver (2010). The idea is that the second period is like a box: we can either open it and take account of what is in it, or we can simply leave it closed and ignore it. In the model of Ergin and Sarver (2010) as the subjective perception that the box contains an intervention[11] grows small the subjective benefit of opening the box goes to zero. Since contemplation is costly it is best not to open the box at all: just optimize in the first period as if the second period will be the same. This is our model.

A Motivating Example: Pigou

In our first example we consider a simple negative externality where the intervention is a Pigouvian tax. Output x_t^i brings an individual benefit $U(x_t^i)$ which is strictly concave $U''(x_t^i) < 0$ and a social cost $L(\overline{x}_t^i)$ strictly increasing $L'(\overline{x}_t) > 0$ and weakly convex $L''(\overline{x}_t) \geq 0$. In one specific application output will be driving speed, with higher speed bringing an individual reduction in commuting time but a social cost in the form of more accidents.

The intervention is a Pigouvian tax $\omega_t x_t^i$ imposed by an outside agency. Here the type corresponds to the tax rate.[12]

We impose two boundary conditions, the first that the individual marginal benefit is large at the lower bound in the sense that $U'(0) > L'(0) + \overline{\omega}$ and the second that the upper bound is sufficiently large that individual benefit is no longer strictly increasing $U'(X) \leq 0$.

A portion of the tax $\alpha \in [0, 1]$ is returned to the group as an equally distributed lump sum, with the remainder going to the outside agency imposing the tax; so member i's utility function can be written as

$$u(\omega_t, \overline{x}_t, x_t^i) = U(x_t^i) - \omega_t x_t^i - L(\overline{x}_t) + \alpha \omega_t \overline{x}_t.$$

We also want to ensure that there is an adequate range of policy interventions. To do so we will require that the initial tax rate not be too high and that the maximum possible tax rate be "high enough." Specifically, we allow but do not require that the initial tax is zero, but we do require that the initial marginal tax is not too large in the sense that $\omega_1 < L'(X)$.

Under these assumptions we show in the Appendix that there is a unique solution x^* to the problem of maximizing $u(\omega_1, \overline{x}_t, \overline{x}_t)$, and that it lies in the interior. Our final assumption is that the highest tax rate is high enough that it becomes individually optimal to implement x^*, that is, $U'(x^*) - \overline{\omega} = 0$. In the case where ω_1 corresponds to no tax and there is a full rebate $\alpha = 1$, this says that $\overline{\omega}$ is "the" Pigouvian tax. We do not examine the consequences of setting tax rates higher than this.

[11] Or contains some other change: There may be many other things in the box besides an intervention to φ, ω_2 - trade wars, new products, and so forth, and the intervention may not be the most important.

[12] Our applications involve limits Λ and a fine Φ if realized output exceeds the limit. In the Appendix we show that if x_t^i represents a minimum intended output and realized output is the product of x_t^i and a standard Pareto distributed random variable the tax is proportional to $\Phi\Lambda^{-1}x_t^i$. In this case ω_t is the ratio of the fine to the limit.

General Assumptions on Individual Utility

We next consider general assumptions that assure that utility $u(\omega^i, \overline{x}_t, x_t^i)$ is well-behaved and that capture the idea of a negative externality present in the Pigou example. First we require that individual optimization problems be well behaved. In addition to assuming that as a function of individual output x_t^i utility $u(\omega^i, \overline{x}_t, x_t^i)$ is strictly differentiably concave, as in the Pigou example, it should satisfy the boundary conditions $D_3u(\omega^i, \overline{x}_t, 0) \geq 0$ and $D_3u(\omega^i, \overline{x}_t, X) \leq 0$. Second we assume that the social planner problem is well-behaved: as in the Pigou example, $u(\omega^i, \overline{x}_t, \overline{x}_t)$ should be strictly differentiably concave in $\overline{x}_t$ with $D_2u(\omega^i, 0, 0) + D_3u(\omega^i, 0, 0) > 0$.

Our interest is in the case of negative externalities. One interpretation of this is that that for $x_t^i > 0$ average output $\overline{x}_t$ reduces utility $D_2u(\omega^i, \overline{x}_t, x_t^i) < 0$, but this is not satisfied in the Pigou example, so we make the weaker assumption that $D_2u(\omega^i, \overline{x}_t, x_t^i) \geq 0$ and $x_t^i > 0$ imply $D_3u(\omega^k, \overline{x}_t, \overline{x}_t) > 0$, that is, the externality can only be positive if individuals also want to increase output. Taken together with earlier assumptions this assures that there is a region where the externality is negative and conflicts with the individual incentives. We assume also that the externality does not increase the individual marginal benefit from increasing output: $D_{32}u(\omega^i, \overline{x}_t, x_t^i) \leq 0$. (This is zero in the Pigou example.)

The types of interventions we are interested in are those like Pigouvian taxes that reduce the individual incentive to over-produce. Hence higher types are assumed to have less benefit from output in the sense that both the individual incentive to produce is reduced $D_{31}u(\omega^i, \overline{x}_t, x_t^i) < 0$ and the social incentive to produce is not increased $D_{21}u(\omega^i, \overline{x}_t, \overline{x}_t) + D_{31}u(\omega^i, \overline{x}_t, \overline{x}_t) \leq 0$. Moreover, to assure an adequate range of interventions, we assume that very high types have little individual incentive to produce in the following sense: if average output is high enough that the social incentive to produce is negative with the lowest type, $D_2u(\omega_1, \overline{x}_t, \overline{x}_t) + D_3u(\omega_1, \overline{x}_t, \overline{x}_t) < 0$, then the highest type prefers not to produce so much, or $D_3u(\overline{\omega}, \overline{x}_t, \overline{x}_t) < 0$.

Finally we make two technical assumptions that assure that even for very great monitoring difficulty the objective function in the presence of monitoring costs maintains the same basic properties as the underlying objective function: specifically we assume that $D_{223}u(\omega_j, \overline{x}_t, x_t^i) \geq 0$ and $D_{123}u(\omega_j, \overline{x}_t, x_t^i) \geq 0$.

In the Appendix we verify that all these assumptions are satisfied in the Pigou example (all the above assumptions are numbered and presented in a table there for convenience before proofs).

Another Example: Cournot

As a second, rather different, example we will consider a group of firms that operates as a cartel, where the intervention is the imposition of (or an increase in) a minimum wage. Here we take $u(\omega^i, \overline{x}_t, x_t^i) = p(\overline{x}_t)x_t^i - c(\omega^i, x_t^i)$ where $p(\overline{x}_t)$ is the market price and $c(\omega^i, x_t^i)$ is the cost for a firm facing a minimum wage of ω^i. We may define revenue as $r(\overline{x}_t) \equiv p(\overline{x}_t)\overline{x}_t$.

We make relatively standard assumptions that assure that the monopoly problem is well-behaved: price is positive, strictly downwards sloping and marginal revenue is also strictly downwards sloping. We also make the plausible although less common assumption that price is convex, that is, $p''(\overline{x}_t) \geq 0$. (We note that all these assumption are satisfied by linear demand.) Marginal cost $D_2c(\omega^i, x_t^i)$ is positive, strictly upwards sloping and strictly increasing in the minimum wage ω^i. We assume that a type $\overline{\omega}$ facing the highest possible minimum wage is still willing to enter at the lowest price $p(X)$, and that the type facing the lowest minimum wage ω_1 is not willing to produce to capacity X at the highest possible price $p(0)$. Under these assumptions at ω_1 there is a unique level of monopoly output x^m that maximizes $r(\overline{x}_t) - c(\omega_1, \overline{x}_t)$, and this lies in the interior (see the Appendix). To ensure an adequate range of interventions we assume that at x^m the marginal cost for highest type $\overline{\omega}$ is higher than the monopoly price: $D_2c(\overline{\omega}, x^m) > p(x^m)$.

In the Appendix we verify that under these assumptions the Cournot model satisfies the general assumptions above.

3. Comparative Statics of Intervention

We examine the comparative statics of increasing ω_2 from the initial level ω_1 towards the upper bound $\overline{\omega}$ in the different regions of the parameter space θ, f, F: how does optimal average output vary?

Recall that there are three alternatives: stick to the existing quota plan, revert to an incentive compatible default social norm, or choose a new social norm, so that optimal average output - which we denote by $\overline{y}^o(\omega_2)$ - may be any of them. For the default social norm, which is just Nash equilibrium in $\left(x_t^i\right)$, we denote that output by $\overline{y}^d(\omega_2)$. Recall that $\overline{y}^c(\omega_2)$ denotes output if a new optimal social norm is chosen and that the quota in the initial simple norm is denoted by $\overline{y}^s$.

The Classical Case

Our first result, Theorem 6 in the Appendix, characterizes the default social norm.

Theorem 1. *The default social norm $\overline{y}^d(\omega_2)$ is well-defined, weakly interior, smooth and strictly decreasing.*

Theorem 1 shows that in a standard setting without social norms, that is, for a Nash equilibrium of the game played by group members without quotas or punishments, we have the expected result: increasing the cost ω_2 of producing output reduces the output $\overline{y}^d(\omega_2)$. This is the source of our basic intuition about the effect of Pigouvian taxes, minimum wages, and the like. In the current setting it may be regarded as the limiting case where $\pi_1 = \pi$ and the signal is uninformative, in which case there is no room for peer-disciplined social norms.

The Main Result

The main result of this paper concerns the case in which monitoring is feasible, that is $\pi_1 > \pi$. The basic picture is that as ω_2 increases output is initially constant due to stickiness of the existing norm. There is a threshold at which it jumps and then (in a general sense) starts to decline and ultimately for high enough ω_2 is lower than in the first period. The anomaly relative to classical theory is that for large F output may jump up rather than down. We next report the details, breaking the presentation in three parts. In the Appendix the main ideas leading to the formal proofs are explained at some length.

I. Small Scale Interventions

Theorems 9 and 10 in the Appendix show that

Theorem 2. *For generic f, F there exists a threshold $\underline{\omega}_2 > \omega_1$ such that $\omega_2 < \underline{\omega}_2$ implies $\overline{y}^o(\omega_2) = \overline{y}^s$.*

In the region where the intervention is small $\omega_2 < \underline{\omega}_2$ optimal output does not change from the period 1 level. This is what one would expect but it is useful to understand what is happening here. Quotas do not constrain individual members from reducing output in response to the increased cost. However, at ω_1 the quotas bind - all would like to produce more - so with a small increase in costs they prefer to stick at the quota and not reduce output. This consequence is the same as would be the case if there were individual adjustment costs - a small change in incentives would not make it worthwhile to bear a fixed cost of adjusting output - but the reason for the "stickiness" is different.

II. Low Renegotiation Cost: Small F

Theorems 8 and 10 in the Appendix show that

Theorem 3. *There exist cost thresholds $\underline{F} > F^- > 0$ such that for generic f, F there exists a threshold $\hat{\omega}_2 > \underline{\omega}_2$ and for all sufficiently small $\epsilon > 0$*

(i) $F < F^-$ implies $\overline{y}^o(\omega_2) \leq \overline{y}^s$ and for $\omega_2 > \underline{\omega}_2$ we have $\overline{y}^o(\omega_2) < \overline{y}^s$

(ii) $F < \underline{F}$ and $\omega_2 > \hat{\omega}_2$ imply $\overline{y}^o(\omega_2) < \overline{y}^s$

(iii) $F < \underline{F}$ and $\underline{\omega}_2 < \omega_2 < \underline{\omega}_2 + \epsilon$ imply $\overline{y}^o(\omega_2) = \overline{y}^c(\omega_2) < \overline{y}^s - \epsilon$ and decreasing in ω_2

As ω_2 increases initially output is flat until we reach $\underline{\omega}_2$. At this point it jumps down (iii). If F is small enough (i) it remains strictly smaller than in the first period, and also if ω_2 is large enough (ii). Here because F is small the choice is basically between sticking at the previous level of output and re-optimizing - there is little advantage to be found in the default social norm. The result is homologous to what we expect in the case that there are individual adjustment costs: after a threshold is reached, it is best to re-optimize and output drops.

III. High Renegotiation Cost: Large F

Theorems 8 and 9 in the Appendix show that

Theorem 4. *There exists a cost threshold $F^+ > \underline{F}$ such that for generic f, F there exists a threshold $\hat{\omega}_2 > \underline{\omega}_2$ and for all sufficiently small $\epsilon > 0$*

(i) $F > F^+$ and $\omega_2 > \underline{\omega}_2$ imply $\overline{y}^o(\omega_2)$ is decreasing

(ii) $F > \underline{F}$ and $\underline{\omega}_2 < \omega_2 < \underline{\omega}_2 + \epsilon$ imply $\overline{y}^o(\omega_2) = \overline{y}^d(\omega_2) > \overline{y}^s + \epsilon$ and decreasing in ω_2

(iii) $F > \underline{F}$ and $\omega_2 > \hat{\omega}_2$ imply $\overline{y}^o(\omega_2) < \overline{y}^s$

As ω_2 increases initially output is flat until we reach $\underline{\omega}_2$. At this point it jumps up and begins to decline (part ii). If F is large enough (i) it declines for all $\omega_2 > \underline{\omega}_2$, but regardless (iii) for large enough ω_2 it is strictly smaller than in the first period. In this case because F is large the choice is basically between sticking at the previous level of output or reverting to the default social norm. And at the switch point, where the group is indifferent between simple and default norms, $\overline{y}^d(\omega_2) > \overline{y}^s$ so output jumps up. This is because as ω_2 increases eventually the default output level $\overline{y}^d(\omega_2)$ falls to y^s. However, when $\overline{y}^d(\omega_2) = y^s$ the default norm is strictly better than the simple norm since it yields the same output at cost f rather than $M > f$. Hence the switch in fact takes place when $\overline{y}^d(\omega_2) > \overline{y}^s$ which is why output jumps up.

This result has no analog when there are individual adjustment costs and is unique to a model of social mechanisms. Here we have initial stickiness followed by a discrete change in output - but in the wrong direction. Moreover, in this case a modest intervention provides misleading information about the consequences of a large intervention: a large enough intervention always lowers output, but the modest intervention has the opposite effect.

4. Increased Externality in the Face of Intervention

One unique prediction of our model is the possibility of output jumping upwards in response to an increase in a Pigouvian tax (Theorem 4), which as already noticed is counter-intuitive and not accounted for by the classical theory. In this section we apply the theorem to interpret the findings of two papers - Gneezy and Rustichini (2000) and Card and Krueger (1994) - reporting observed upward jumps.

Case Study: School Fines

Gneezy and Rustichini (2000) showed that introducing modest incentives can lead to the discouragement of the activity it is designed to promote: they showed that introducing a modest fine for being late to pick up children at a day-care center resulted in more parents picking up their children late. A behavioral interpretation of the finding can be found in Benabou and Tirole (2006). The idea there is that in the absence of fines, picking up children on time serves a valuable signaling purpose. With fines, the signaling value of being on time is lowered enough that it becomes worthwhile to be a little late and pay the fine.

In our setting the intervention ω_t represents the level of the fine. In the field experiment initially there was no fine $\omega_1 = 0$, then one was imposed $\omega_2 > 0$. As there was no prior warning or discussion of the fine, it is reasonable to think it was unanticipated. From a social point of view it is likely both that parents at day-care know each other and that there are some mild social sanctions towards parents who are persistently late - in fact it is unlikely that prior to the fine parents simply picked up their children at the moment of the day they personally found most convenient. Moreover, as the fine was introduced suddenly and without explanation it might well have been anticipated to be of short duration (as in fact it was) so that it would not be worth renegotiating to identify the new optimal social norm. Moreover, the field experiment was conducted in 1998 well before the advent of social media (Facebook was founded in 2004). That is: it is plausible that F was relatively large in this case.

Given large F and an intermediate level of fine, the interpretation of the Gneezy-Rustichini finding according to Theorem 4(ii) is that while prior to the fine lateness was disciplined through a social norm among parents, after its imposition the old norm was abandoned and not replaced by a new one, and consequently lateness increased.[13]

Case Study: Minimum Wage

In the Cournot setting ω_2 represents a minimum wage for unskilled workers. In the absence of norms the classical result applies that an increase of the minimum wage results in a drop in production. When firms can be viewed as member of a norm-enforcing large cartel Theorem 4 raises the possibility of output jumping upwards in response to an increase in the minimum wage if the change is neither too small nor too large. The rationale according to our model is the same as in the day-care case: the increase in output is due to the old norm being abandoned and not replaced by a new one. Typically studies of the minimum wage examine employment rather than output. With multiple factors of production there is substitution that will tend to lower employment of unskilled labor so the latter may go down while output increases. In the extreme case where unskilled labor is needed in fixed proportion with an aggregate of other factors of production - that is the relationship between unskilled labor and the aggregate is Leontief - then the employment of unskilled labor will be proportional to output. More generally if other factors are poor substitutes for unskilled labor then the output effect will dominate the substitution effect and there can be an upward jump in employment of unskilled labor in response to a minimum wage increase. This has been observed at least once in the study of the impact of a 1992 minimum wage increase in New Jersey on employment of unskilled labor in the fast food industry by Card and Krueger (1994).

5. Redistribution, Group Welfare and Behavior

Welfare

We turn next to the welfare consequences of using a Pigouvian tax in place of the informal system of sanctions set up by the group. We first consider the case in which $\alpha = 1$, so the entire proceeds of the tax are returned to the group. In the day care experiment this may be a reasonable approximation: since the school is supported by fees from the parents and different schools compete with each other. Implicitly, the money from fines either reduces what parents have to pay, or increases the services they receive.

What are the welfare consequences of intervention? With small scale interventions output remains unchanged and the intervention has no welfare consequence at all. Behavior does not

[13] Our theory does not explain why when the fine was removed parents continued being late - however, the data after the fine was removed is very short in duration so we cannot say whether in a few more weeks or months lateness began to drop. In general we expect the frictions (and time) to agree to a non-trivial social norm to be greater than that needed to revert to the default.

change, taxes are collected from group members but returned lump sum so everybody is exactly as well off as in the first period. In the day-care case study we argued that F is likely to be large, so when the jump in output occurs at $\underline{\omega}_2$ the group is indifferent between the simple social norm and the default social norm, so welfare at the default social norm is also the same as in the first period. As the intervention increases output then declines: this increases welfare. The taxes are a wash, being collected and returned to the group, but since output was greater than the first best, the reduction makes everyone better off.

In the classical analysis the upward jump in output is regarded as a failure of policy. The goal of the policy is to reduce output in the face of an externality. But that analysis misses the mark. Here increased output is an indication that the policy has a desirable effect. While the increase in output has a negative consequence for welfare, overall welfare goes up because by switching to the default social norm the cost of monitoring is avoided and this more than makes up for the loss from increased output.

What about the case in which F is low and the switch is to a new social norm? Here we might ask a broader question. In this model a formal tax can be levied only by an outside agency; the group itself has access only to informal punishments. Never-the-less it is useful to ask what the group would do if it could control the tax as well as use informal sanctions - and indeed, there are circumstances where the group may be able to influence the outside agency setting the taxes. In the case where $\alpha = 1$ Theorem 11 in the Appendix shows that if the group can choose ω_t along with y_t, P_t then welfare is unambiguously increasing with the level of tax (bear in mind we have assumed it can be no higher than the optimal Pigouvian tax). The key element of a tax (with full rebate) is that it is simply a transfer payment within the group: there is no net loss of utility to the group in carrying out a punishment in the form of a fine. By contrast the informal punishments P_t^i represent a net loss to the group. Hence the fine technology is a superior technology and if it is available the group prefers to use that in place of informal enforcement.

In the case of intervention the implication is that once the jump (down) takes place to a new social norm, group welfare is increasing with the level of the tax. Overall an intervention with a Pigouvian tax is either neutral (if the intervention is too small) or an improvement.

The situation is different in the case in which $\alpha < 1$ so that only some of the proceeds of the tax are returned to the group with the remainder being taken by the outside agency. In classical theory lump sum transfers are not supposed to matter, and indeed it is sometimes argued that a policy is desirable if it is possible to pay compensation in the form of lump sum transfers and make everyone better off. As we shall see in the face of social mechanisms this can be misleading. We focus in particular on the difference in incentives between the group and the outside planner.

To see the difference in incentives recall that when $\alpha = 1$, so incentive are aligned, after the jump at $\underline{\omega}_2$ the welfare of the group is increasing in the ω_2. In the case $\alpha < 1$ locally the opposite is true: at the jump point the group is necessarily indifferent, so utility just after the jump is close to indifferent. However, while with $\alpha = 1$ group welfare did not change in the region $\omega_1 \leq \omega_2 \leq \underline{\omega}_2$ when $\alpha < 1$ welfare necessarily decreases due to the fact that taxes are collected and not all are returned. Hence after the jump the group remain worse off than at ω_1.

Behavioral Consequences

The difference in incentives has behavioral consequences as well. Often behavioral economists consider naive agents unable to discern their personal interest against sophisticated planners who are better able to determine what is best for individuals. We will take the opposite point of view, and consider a naive planner. A naive planner chooses a Pigouvian tax ignoring the fact that it may be $\alpha < 1$: it chooses the tax so that it is individually optimal to produce the $\overline{x}^*$ maximizing $U(\overline{x}) - L(\overline{x})$, that is it chooses ω^* satisfying $U'(\overline{x}^*) = L'(\overline{x}^*) = \omega^*$. In other words it computes the marginal cost of the externality and sets the tax equal to that (note that in case $\alpha = 1$ it is $\omega^* = \overline{\omega}$). In all the debates we have seen about, for example, a carbon tax, this seems to be the computation made by planners and would-be planners. Suppose this Pigouvian tax is in fact set but that $\alpha < 1$.[14]

It is shown in the Appendix Theorem 2 that if the group chooses a new social norm it maximizes

$$W_1(y_t^i) \equiv -(1-\alpha)\tau(\omega^*, y_t^i) + U(y_t^i) - L(y_t^i) - \theta\left[\left(U(x^b(\omega^*, y_t^i)) - \tau(\omega^*, x^b(\omega^*, y_t^i))\right) - \left(U(y_t^i) - \tau(\omega^*, y_t^i)\right)\right].$$

With the Pigouvian tax $x^b(\omega^*, \overline{x}^*) = \overline{x}^*$, hence the derivative of this expression is

$$W_1'(y_t^i) \equiv -(1-\alpha)\omega^* + U'(y_t^i) - L'(y_t^i) + \theta\left[\left(U'(y_t^i) - \omega^*\right)\right].$$

If in fact $\alpha = 1$ we see that the first order condition is satisfied at $\overline{x}^*$ as we expect. If $\alpha < 1$ we see that the derivative is strictly smaller, implying that the (unique) solution of the first order condition must be smaller: $\overline{y}^o < \overline{x}^*$.

In other words: behavior is not invariant to distribution, and a naive planner who sets a Pigouvian tax with $\alpha < 1$ will not induce optimal output, but will induce underproduction.

6. Partial Redistribution and Tax Repeal

Let us again address the broader issue of what the group would do if it could influence taxes as well as control informal sanctions. For any given α let $\Omega(\alpha)$ denote the set of optimal taxes for the group, that is the set of taxes that are part of a solution to the design problem in which they choose taxes, quotas, and punishments. This problem is not as straightforward as it seems. Consider for example the Pigouvian tax in a situation where θ is large and α small. To lower the tax creates a dilemma: true it creates a tax gain to the group. At the same time it creates an incentive to increase output. Because θ is large it is costly to keep prevent output from increasing a lot. However: if it is allowed to increase a lot since α is quite small that creates a large tax loss.

[14] For technical reasons we have not actually allowed the limit case $\omega_2 = \overline{\omega}$. The reason for this is that in the limit case the upward constraints do not strictly bind and our results about strict inequalities do not hold. To avoid a series of special cases in which they hold only weakly we have ruled out this case. Modulo the upward constraints not strictly binding, the case $\omega_2 = \overline{\omega}$ is otherwise well behaved.

Hence it may actually be best to raise the tax and use the improved incentive to reduce output thereby offsetting the tax loss from the increased tax.

In the Appendix Theorem 12 we show that if θ is small and α is large in fact the optimal tax must decline as α goes down. Specifically

Theorem 5. *There exists $\underline{\theta} > 0$, $\overline{\alpha}(\theta) < 1$ such that for $0 < \theta < \underline{\theta}$, $\overline{\alpha}(\theta) < \alpha \leq 1$ the optimal tax $\omega_2^*(\alpha)$ is unique, smooth and strictly differentiably increasing in α with $\omega_2^*(1)$ the Pigouvian tax.*

Consider then the case where this is true, that F is small and that a naive planner observes that current taxes are below the Pigouvian level (perhaps because the group was able to influence taxes in the first period). The planner then intervenes to raise taxes to the Pigouvian level. As F is small the group is willing to organize itself and find a new social norm. Once it is designing a new mechanism, if it can influence taxes, it might choose to redesign taxes as well. That is, rather than improving efficiency, the naive planner may instead create a political backlash that will lead to lower rather than higher taxes.

Case Study: Yellow Vests

Groups responding to the imposition of naive Pigouvian taxes by engaging in some form of tax repeal has been observed. Boyer et al [2019] in particular have documented that this happened in the case of the French "Yellow Vests."

In this case x_t^i represents driving speed, while the intervention ω_2 is the inverse of the speed limit. On July 1, 2018 the French Federal Government lowered the speed limit on secondary highways from 90 km/h to 80 km/h. The bulk of the impact fell on rural communities where there are no primary highways and secondary highways are widely used. Let us first observe that on rural secondary highways social norms regulating driving speed are likely to play an important role. Although driving is to an extent anonymous, drivers who are perceived to drive excessively fast are often punished, for example, by blocking their progress by intentionally slowing down, making it difficult to pass, or simply through obscene gestures. While fictional, the Damián Szifron film "Relatos Salvajes" illustrates the idea well. As drivers observe one another well, we may hypothesize that θ is relatively low. Two other facts are relevant. First, $\alpha < 1$: the speed camera revenue is not returned to rural drivers who receive only an indirect benefit. Second F was quite low due to the advent of social media. Indeed we know that Facebook played a key role in the organization of rural communities. Hence our theory says that if the group could do so at low cost it would organize not only a new driving speed norm, but also a lowering of taxes.

Although it is perhaps less well known than the more publicized riots in Paris, the group did indeed act to "repeal" the tax. The rate of traffic camera destruction jumped by 400% and in the year following the speed limit change about 75% of all traffic cameras in France were destroyed. We refer the interested reader to Boyer et al (2019) who document both the link between the change in speed limit and the yellow vest movement, as well as the systematic way in which that group organized itself.

There are several points to make on this case. First, it is not true in this model that taxes are generally rejected. As Theorem 5 shows even with $\alpha < 1$ the group will often prefer a positive tax rate to partially substitute even for relatively inexpensive monitoring. Second, setting the Pigouvian tax - the usual policy prescription - is a mistake when $\alpha < 1$ for two reasons. First, if a new social norm is introduced output will be too low. Second, it creates an incentive for the group to attempt to (partially) repeal the tax. All of this emphasizes that a credible commitment to $\alpha = 1$ is highly desirable for implementing Pigouvian taxes.

7. Conclusion

The main contention of this paper is that the role of social norms in enforcing pro-social behavior should not be neglected. To take one of many examples consider the field experiment of Gneezy and List (2006) in which they paid some solicitors a fixed bonus above the market wage and others not. They discovered that initially those with the bonus increased their effort, but over the entire course of the experiment did about the same amount of work per unit of pay as those without the bonus. This is consistent with a social norm in which the wages per unit of effort are part of a social norm: solicitors do the amount of work per pay as called for by the social norm regardless of whether the money is paid as a piece rate or a lump sum.

The social norms we have studied are optimally chosen by the group, that is they are endogenous. An instructive example of the endogeneity and adaptability of social norms is the custom of tipping service providers: this is commonplace in the US and UK, but rare, for example, in Italy. In Italy it works rather the other way around: not only is there no tipping but repeat customers get a discount - kind of a negative tip. In the US and UK there is a definite social sanction for not tipping. Other people at your table as well as the waiters may sneer at you - indeed you may be explicitly told not to return. We would argue that these are not just arbitrary customs, but rather are based on the need for incentives. With low waiter turn-over both within restaurants and within communities social norms among waiters can support good service and tipping is not needed - this is the situation in Italy. With high waiter turn-over and waiters not tied to the local community it is difficult for social norms to support good service, and so tipping is a needed incentive.

Changing social norms is costly: they adjust slowly when incentives are weak - for example, the twelve years debate leading up to the change from driving on the left to the right in Sweden - and rapidly when incentives are strong - for example, only few hours after the attack on the World Trade Center on 9/11 the social norm of never resist hijackers changed permanently to always resist hijackers.[15] This is strongly suggestive of adjustment costs. Here we have examined how endogenous social norms modeled through mechanism design subject to adjustment costs interact with external policies.

Our analysis includes a number of findings that should be cautionary. Increases in taxes designed to decrease an externality may increase it - but never-the-less increase welfare. The impact of these

[15] See Levine (2012) for a discussion of both.

taxes on behavior can depend upon how the revenue is spent. The ability of groups to self-organize through social media can play a significant role in determining the consequences of interventions. These observations and analysis need to be considered in a number of economic areas. In blood donations the type of trade-off between voluntary and paid contributions, studied for example by Meyer and Tripodi (2017), should be viewed in light of our findings. Much the same can be said about environmental issues. Peer pressure within certain groups to be "green" is significant - the sorting of garbage, for example, cannot easily be enforced otherwise. To develop useful policies it is necessary to know how external incentives complement and substitute with internal incentives. This paper is a step in that direction.

References

Bénabou, Roland, and Jean Tirole (2006): "Incentives and prosocial behavior," *The American Economic Review* 96(5): 1652-1678.

Bigoni, M., S. Bortolotti, M. Casari., D. Gambetta and F. Pancotto (2016): "Amoral familism, social capital, or trust? The behavioural foundations of the Italian North–South divide," *The Economic Journal* 126:1318-1341.

Block, J. I., and Levine, D. K. (2016): Codes of conduct, private information and repeated games," *International journal of game theory*, 45: 971-984.

Boyer, Pierre C., Thomas Delemotte, Germain Gauthier, Vincent Rollet and Benoît Schmutz (2019): "Les déterminants de la mobilisation des "gilets jaunes", Working Papers 2019-06, Center for Research in Economics and Statistics.

Calvo, G. A. (1983): "Staggered prices in a utility-maximizing framework," *Journal of monetary Economics* 12(3): 383-398.

Card, D., and Krueger, A. B. (1994): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *The American Economic Review* 84(4): 772-793.

Dell, Melissa, Nathan Lane, and Pablo Querubin (2918): "The historical state, local collective action, and economic development in Vietnam," *Econometrica* 86: 2083-2121.

Dye, R. A. (1985): "Costly contract contingencies," *International Economic Review* 26(1): 233-250.

Ergin, H., and T. Sarver (2010): "A unique costly contemplation representation," *Econometrica* 78: 1285-1339.

Feddersen, T. , and A. Sandroni (2006): "A theory of participation in elections," A*merican Economic Review* 96: 1271-1282.

Fehr, E., and S. Gächter (2000): "Fairness and retaliation: The economics of reciprocity," *Journal of Economic Perspectives* 14: 159-181.

Fudenberg, Drew, David Levine and Eric Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica* 62(5): 997-1039.

Fudenberg, D., D. K. Levine and W. Pesendorfer (1998): "When are Non-Anonymous Players Negligible," *Journal of Economic Theory* 79: 46-71

Gale, D and Sabourian, H. (2005): "Complexity and competition," *Econometrica*, 73: 739-769.

Gneezy, U., and List, J. A. (2006): "Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments," *Econometrica* 74(5): 1365-1384.

Gneezy, U., and Rustichini, A. (2000): "A fine is a price," *The Journal of Legal Studies* 29(1): 1-17.

Hart, O., and Moore, J. (1988). "Incomplete contracts and renegotiation," *Econometrica* 56(4): 755-785.

Levine, David K. (2012): *Is behavioral economics doomed?: The ordinary versus the extraordinary* Open Book Publishers.

Levine, David and Salvatore Modica (2016): "Peer Discipline and Incentives within Groups", *Journal of Economic Behavior and Organization* 123: 19-30.

Levine, David and Salvatore Modica (2017): "Size, Fungibility, and the Strength of Lobbying Organizations", *European Journal of Political Economy* 49: 71-83.

Meyer, Christian Johannes and Tripodi, Egon, Sorting into Incentives for Prosocial Behavior (October 24, 2017). Available at SSRN: https://ssrn.com/abstract=3058195

Modica, S. and A. Rustichini (1994): "Awareness and partitional information structures," *Theory and Decision* 37: 107-124.

Olson Jr., Mancur (1965): *The Logic of collective action: public goods and the theory of groups*, Harvard Economic Studies.

Ostrom, Elinor (1990): *Governing the commons: The evolution of institutions for collective action*, Cambridge university press.

Sims, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics* 50: 665-690.

Tirole, J. (2009): "Cognition and incomplete contracts." *American Economic Review* 99(1): 265-94.

Townsend, Robert M. (1994): "Risk and Insurance in Village India," *Econometrica* 62: 539-539.

Appendix: Proofs

Outline of the ideas

There are four key ideas about the utility function that we need

1. concavity in own action: so that best responses are unique
2. concavity overall: so that the optimization problem without monitoring costs is well-behaved
3. direction in which the externality effects utility and marginal utility
4. direction in which the type effects utility and marginal utility
5. a technical assumption (the one on third derivatives): this is to make sure that the monitoring cost does not cause a non-convexity

There are several preliminary steps needed

1. establish the uniqueness and properties of best responses
2. establish that the downward constraint does not bind
3. establish a formula for the least monitoring cost

The basic idea in all parts is to establish with respect to "stronger" interventions that

1. for the simple norm initially nothing changes, then output declines
2. for the default and contingent social norm output declines: the former is "easy" the latter is "hard"

The basic result is: as intervention increases initially nothing changes then depending on whether F is large or small output jumps either up or down, then eventually decreases to a level lower than without intervention.

Summary of Assumptions

We start by listing the assumptions used in the text. Then a proof is given of the basic characterization of the mechanism design problem. Next the central results of the paper are proved. Finally we check that the assumptions are satisfied in the two applications (Pigou and Cournot) we cover in the text.

We recall that the individual utility $u(\omega^i, \overline{x}_t, x_t^i)$ is assumed to be smooth throughout, and that there is an upper bound on the possible interventions: $\omega_1 \leq \omega^i < \overline{\omega}$. The assumptions are listed in the following table:

	List of Assumptions
	Basic Results:
1	Strict differentiable concavity in x_t^i: $D_{33}u(\omega^i, \overline{x}_t, x_t^i) < 0$
2	Lower boundary: $D_3u(\omega^i, \overline{x}_t, 0) \geq 0$
3	Upper boundary: $D_3u(\omega^i, \overline{x}_t, X) \leq 0$
4	Externality is negative: $D_2u(\omega^i, \overline{x}_t, x_t^i) \geq 0$ and $x_t^i > 0 \Rightarrow D_3u(\omega^k, \overline{x}_t, \overline{x}_t) > 0$
4a	Externality is negative: $D_2u(\omega^i, \overline{x}_t, x_t^i) < 0$ for $x_t^i > 0$
5	Externality is negative: $D_{32}u(\omega^i, \overline{x}_t, x_t^i) \leq 0$
6	Higher types have less incentive to deviate: $D_{31}u(\omega^i, \overline{x}_t, x_t^i) < 0$
	One type:
7	$u(\omega^i, \overline{x}_t, \overline{x}_t)$ strictly differentiably concave in $\overline{x}_t$
8	Lower boundary: $D_2u(\omega^i, 0, 0) + D_3u(\omega^i, 0, 0) > 0$
9	Higher types have lower benefit from production: $D_{21}u(\omega^i, \overline{x}_t, \overline{x}_t) + D_{31}u(\omega^i, \overline{x}_t, \overline{x}_t) \leq 0$
10	$D_{223}u(\omega_j, \overline{x}_t, x_t^i) \geq 0$
11	$D_{123}u(\omega_j, \overline{x}_t, x_t^i) \geq 0$
12	Possibility of large interventions: $D_2u(\omega_1, \overline{x}_t, \overline{x}_t) + D_3u(\omega_1, \overline{x}_t, \overline{x}_t) < 0 \Rightarrow D_3u(\overline{\omega}, \overline{x}_t, \overline{x}_t) < 0$

For each part the assumptions pertaining the earlier parts are maintained. As we mentioned in the text, by strict differentiable concavity we mean that the Hessian matrix is negative definite (which is sufficient but not necessary for strict concavity).

Basic Results

For our basic results we allow types ω^i to vary by individual. Recall that for any type ω^i and average output $\overline{x}$ the maximizer of $u(\omega^i, \overline{x}, x^i)$ is denoted by $x^b(\omega^i, \overline{x})$.

Proposition 1. *There is a unique best response $x^b(\omega^i, \overline{x})$, it is weakly interior, smooth and decreasing, and strictly decreasing in ω^i with, for $h = 1, 2$*

$$D_h x^b(\omega^i, \overline{x}) = -\Big(D_{33}u(\omega^i, \overline{x}, x^b(\omega^i, \overline{x}))\Big)^{-1} D_{3h}u(\omega^i, \overline{x}, x^b(\omega^i, \overline{x})).$$

Proof. The derivative of own utility with respect to own action is

$$D_3u(\omega^i, \overline{x}, x^i).$$

From the concavity #1 and boundary conditions #2, #3 this has a unique zero at $x^b(\omega^i, \overline{x})$:

$$D_3u(\omega^i, \overline{x}, x^b(\omega^i, \overline{x})) = 0.$$

The inverse function theorem for $h = 1, 2$ then gives $D_h x^b(\omega^i, \overline{x})$ as in the statement. For $h = 1$ that expression is negative from #1 and #6, for $h = 2$ it is non-positive from #1 and #5. □

Theorem 6. *The default social norm $y^d(\omega_2)$ is well-defined, weakly interior, smooth and $\overline{y}^d(\omega_2)$ is strictly decreasing.*

Proof. From Proposition 1 $f(\overline{x}_t) = x^b(\omega_2, \overline{x}_t)$ is continuous and decreasing in $\overline{x}_t \in [0, X]$. This implies the existence of a unique fixed point in $\overline{x}_t$ denoted $\overline{y}^d(\omega_2)$. As $f(\overline{x}_t) - \overline{x}_t$ is smooth with strictly negative derivatives we may apply the implicit function theorem to conclude that $\overline{y}^d(\omega_2)$ is smooth and strictly decreasing. From Proposition 1, this implies in turn that $y^d_j(\omega_2) = x^b(\omega_j, \overline{y}^d(\omega_2))$ is smooth and weakly interior. □

The General Problem

The remainder of the section will consider the *general problem* of choosing a *scheme* of individual quotas y^i, individual output levels x^i and individual punishments P^i to maximize the objective function

$$\int \left[u(\omega^i, \int x^i di, x^i) - \mathbf{1}[x^i > y^i]\pi_1 P^i - \mathbf{1}[x^i \leq y^i]\pi P^i \right] di$$

subject to individuals making optimal choices of x^i given the quotas and punishments

$$x^i \in \arg\max u(\omega^i, \int x^i di, x^i) - \mathbf{1}[x^i > y^i](\pi_1 - \pi)P^i.$$

Our goal is to prove

Proposition 2. *A scheme y^i, x^i, P^i is optimal if and only if for $\overline{y} = \int y^i di$*

(i) for almost all i we have $x^i = y^i = y(\omega^i)$

(ii) $y(\omega^i)$ is a solution to

$$\max \int \left(u(\omega^i, \overline{y}, y(\omega^i)) - \theta \left[u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y})) - u(\omega^i, \overline{y}, y(\omega^i)) \right] \right) di$$

(iii) $P^i = (u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y}) - u(\omega^i, \overline{y}, y(\omega^i))/(\pi_1 - \pi)$.

(iv) The solution satisfies

(a) $y(\omega^i) < x^b(\omega^i, \overline{y})$ if $y(\omega^i) > 0$

(b) $0 < \overline{y} < \overline{y}^d$.

We will do this in two steps. First we will show that the result holds if it is possible to costlessly impose downward quotas; then we will show that quotas y^i, individual output levels x^i and punishments P^i solve that problem if and only if they solve the problem without downward quotas - that is, we will show that at the solution to the problem with costless downward quotas the downward quotas do not bind. The argument is articulated in Lemmas 1 to 6.

Lemma 1. *With costless downward quotas the general problem is equivalent to the problem of choosing quotas y^i and punishments P^i to maximize*

$$\int \left[u(\omega^i, \overline{y}, y^i) - \pi P^i\right] di$$

subject to

$$\mathbf{1}[x^b(\omega^i, \overline{y}) \geq y(\omega^i)] \left(u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y})) - u(\omega^i, \overline{y}, y^i)\right) \leq (\pi_1 - \pi)P^i.$$

Proof. From the revelation principle we can assume that $x^i = y^i$ subject to incentive compatibility. Deviating downwards is not feasible, hence incentive compatibility requires that the maximum gain from deviating upwards be less than or equal to the cost of punishment times the increased chance of punishment and this is what the constraint states. As the quota is not violated the punishment P^i has an expected cost of πP^i. □

Lemma 2. *With costless downward quotas the general problem is equivalent to the problem of choosing quotas y^i to solve*

$$\max \int \left(u(\omega^i, \overline{y}, y^i) - \mathbf{1}[y^i \leq x^b(\omega^i, \overline{y})]\theta \left[u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y})) - u(\omega^i, \overline{y}, y^i)\right]\right) di$$

where $\overline{y} = \int y^i di$.

Proof. For given quotas y^i punishment cost P^i must be minimized, so the incentive constraint must hold with equality. Plugging in from Lemma 1 gives the result. □

To prove Proposition 2 with costless downward quotas, it suffices to show that optimal quotas depend only on type.

Lemma 3. *With costless downward quotas an optimal quota scheme has $y^i = y(\omega^i)$ for almost all i.*

Proof. Consider the problem defined in Lemma 2 of maximizing

$$\int \left(u(\omega^i, \overline{y}, y^i) - \mathbf{1}[y^i \leq x^b(\omega^i, \overline{y})]\theta \left[u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y})) - u(\omega^i, \overline{y}, y^i)\right]\right) di.$$

subject to average output being fixed, $\int y^i di = \overline{y}$. The objective function is additively separable in y^i. Each component to the left of $x^b(\omega^i, \overline{y})$ is strictly concave by #1 and similarly to the right. Moreover, the function is maximized at $x^b(\omega^i, \overline{y})$ so in fact it is continuously differentiable and strictly concave. As we have a strictly concave objective function subject to a linear constraint, the solution is characterized by a Lagrange multiplier λ (of indeterminate sign) and the independent problems of maximizing

$$u(\omega^i, \overline{y}, y^i) - \mathbf{1}[y^i \leq x^b(\omega^i, \overline{y})]\theta \left[u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y})) - u(\omega^i, \overline{y}, y^i)\right] - \lambda y^i.$$

Since the problem is strictly concave it has a unique solution - and that depends only on ω^i, λ and $\overline{y}$. □

To prove the full version of Proposition 2 we need to show that in the solution to the costless downward quota problem the downward quotas do not bind. We next show that it is also the case that the upwards constraints, on the contrary, do bind. This latter fact proves part (a) of the final assertion of the theorem.

Lemma 4. *Under #4a with costless downward quotas at the optimum the downward constraints do not bind and the upwards constraints strictly bind for $y(\omega^i) > 0$. Under #4 this is also true provided there is only one type.*

Proof. The downward constraints do bind when $x^b(\omega^i, \overline{y}) < y(\omega^i)$ and $y(\omega^i) > 0$. The upward constraints strictly bind when $x^b(\omega^i, \overline{y}) > y(\omega^i)$ or $y(\omega^i) = X$ and $D_3 u(\omega^i, \overline{y}, X) > 0$. The latter, however, violates the boundary constraint #3. We wish to rule out the cases where downward constraints bind or the upward ones do not, that is we have to rule out the case $x^b(\omega^i, \overline{y}) \le y(\omega^i)$ with $y(\omega^i) > 0$. Let I_D be the set of such i and I_U the complement of that set (both measurable by general continuity considerations). If $\int_{I_D} di = 0$ we are done. If $\int_{I_D} di = 1$ then we may lower the quotas for all $i \in I_D$ slightly. For a particular i the objective function is

$$u(\omega^i, \overline{y}, y^i) - \mathbf{1}[y^i \le x^b(\omega^i, \overline{y})]\theta \left[u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y})) - u(\omega^i, \overline{y}, y^i)\right].$$

If $x^b(\omega^i, \overline{y}) < y(\omega^i)$ with $y(\omega^i) > 0$ then for a small enough decrease in the quota, we still get $\mathbf{1}[y^i \le x^b(\omega^i, \overline{y})] = 0$, and so utility changes by

$$D_2 u(\omega^i, \overline{y}, y^i) d\overline{y} + D_3 u(\omega^i, \overline{y}, y^i) dy^i.$$

If instead we had $y^i = x^b(\omega^i, \overline{y})$ then lowering leads to

$$\begin{aligned} &D_2 u(\omega^i, \overline{y}, y^i) d\overline{y} + D_3 u(\omega^i, \overline{y}, y^i) dy^i \\ &- \theta \left[D_2 u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y})) - D_2 u(\omega^i, \overline{y}, y^i)\right] d\overline{y} + \theta D_3 u(\omega^i, \overline{y}, y^i) dy^i \\ =&D_2 u(\omega^i, \overline{y}, y^i) d\overline{y} + D_3 u(\omega^i, \overline{y}, y^i) dy^i \\ &- \theta \left[D_2 u(\omega^i, \overline{y}, y^i) - D_2 u(\omega^i, \overline{y}, y^i)\right] d\overline{y} + \theta D_3 u(\omega^i, \overline{y}, y^i) dy^i. \end{aligned}$$

Finally $y^i = x^b(\omega^i, \overline{y})$ implies $D_3 u(\omega^i, \overline{y}, y^i) = 0$ (by interiority) and so we again obtain $D_2 u(\omega^i, \overline{y}, y^i) d\overline{y} + D_3 u(\omega^i, \overline{y}, y^i) dy^i$. In the multiple type case since $d\overline{y} < 0$, $D_2 u(\omega^i, \overline{y}, y^i) < 0$ by #4a, and $D_3 u(\omega^i, \overline{y}, y^i) \le 0$ from $x^b(\omega^i, \overline{y}) \le y(\omega^i)$ and #1 we see that this is strictly positive, contradicting the supposed optimality of the quotas. In the single type case by Lemma 3 we may write $D_2 u(\omega^i, \overline{y}, \overline{y}) d\overline{y} + D_3 u(\omega^i, \overline{y}, \overline{y}) dy^i$. By #4 if $D_2 u(\omega^i, \overline{y}, \overline{y}) \ge 0$ then $D_3 u(\omega^i, \overline{y}, \overline{y}) > 0$ contradicting the hypothesis that $x^b(\omega^i, \overline{y}) \le y(\omega^i) = \overline{y}$. Hence $D_2 u(\omega^i, \overline{y}, \overline{y}) d\overline{y} + D_3 u(\omega^i, \overline{y}, \overline{y}) dy^i$ is again strictly positive.

Finally, if $0 < \int_{I_D} di < 1$ we again lower the quotas for $i \in I_D$. As noted in I_U we cannot have $y^i = X$ ($D_3 u(\omega^i, \overline{y}, X) \le 0$ so upward constraint cannot bind at X). Hence we may increase the quotas for $i \in I_U$ slightly and we can arrange the increase and decrease so that the average quota $\overline{y}$ does not change. Since the average quota did not change (and by the argument above)utility for a particular i changes by

$$(1 + \mathbf{1}[i \in I_U]\theta)\, D_3u(\omega^i, \overline{y}, y^i)dy^i$$

which is non-negative for $i \in I_D$ and strictly positive for $i \in I_U$. This increase in utility contradicts the supposed optimality of the quotas. □

Lemma 5. *At any optimum* $0 < \overline{y} < X$.

Proof. If $\overline{y} = X$ then $y(\omega^i) = X$ for almost all ω^i. From Lemma 4 this implies $D_3u(\omega^i, X, X) > 0$ contradicting the boundary condition #3. If $\overline{y} = 0$ then $y(\omega^i) = 0$ for almost all ω^i. From Lemma 2 and 4 we must be maximizing $\int \left(u(\omega^i, \overline{y}, y(\omega^i)) - \theta\left[u(\omega^i, \overline{y}, x^b(\omega^i, \overline{y})) - u(\omega^i, \overline{y}, y(\omega^i))\right]\right) di$ and cannot improve utility by choosing $y(\omega^i) = \overline{y} > 0$. Hence the derivative

$$\int \Big(\big[D_2u(\omega^i, 0, 0) + D_3u(\omega^i, 0, 0)\big] + \theta D_3u(\omega^i, 0, 0) + \\ + \theta \Big[D_2u(\omega^i, 0, 0) - D_2u(\omega^i, 0, x^b(\omega^i, 0))\Big]\Big) di$$

must be non-positive. By #5 $D_{23}u(\omega_2, 0, x_t^i) \leq 0$ the final term is non-negative; the first term is strictly positive by #8 and the middle term is non-negative by #2. Hence the optimum does not lie on the lower boundary. □

Lemma 6. *At any optimum* $\overline{y} < \overline{y}^d$.

Proof. Suppose instead that $\overline{y} \geq \overline{y}^d$. By Proposition 1 $x^b(\omega^i, \overline{y}) \leq x^b(\omega^i, \overline{y}^d)$. By Lemma 5 $\overline{y} > 0$ so by Lemma 4 $y^i < x^b(\omega^i, \overline{y})$ for a positive measure set of i and $y^i \leq x^b(\omega^i, \overline{y})$ for all i. Hence $\overline{y} < \int x^b(\omega^i, \overline{y})di \leq \int x^b(\omega^i, \overline{y}^d)di = \overline{y}^d$ a contradiction. □

One Type: the Central Results

From now on we concentrate on the case where in each period all members have a common type ω_t (as in the text). Recall that if the mechanism designer chooses $\left(y_t^i, P_t^i\right)$ we denote by $X\left(y_t^i, P_t^i\right)$ the set of pure strategy Nash equilibria of the induced game.

Theorem 7. $X\left(y_t^i, P_t^i\right)$ *is non-empty and closed.*

Proof. Individual utility is $u(\omega_t, \overline{x}_t, x_t^i) - \mathbf{1}(x_t^i \leq y_t^i)\pi P_t^i - \mathbf{1}(x_t^i > y_t^i)\pi_1 P_t^i$. This is not continuous in x_t^i but as $\pi_1 > \pi$ it can jump upwards and not downwards so it is upper semi-continuous in x_t^i as well as y_t^i. This implies that the best response function is upper hemi-continuous as a function of $\overline{x}_t, y_t^i, P_t^i$, which has two implications:

1. $X\left(y_t^i, P_t^i\right)$ is closed.
2. the best response correspondence for each member $B_t^i(\overline{x}_t)$ is upper hemi-continuous.

To prove existence, that is, $X\left(y_t^i, P_t^i\right)$ non-empty, let $B(\overline{x}_t) = \int B_t^i(\overline{x}_t)di$. Since everything is bounded this is upper hemi-continuous from #2 and the dominated convergence theorem. Hence we need only prove $B(\overline{x}_t)$ is convex valued to apply Kakutani.

Observe that $B_t^i(\overline{x}_t)$ takes on one of the two values $y_t^i, x^b(\omega_t, \overline{x}_t^i)$ and is either single valued or there is indifference with $y_t^i < x^b(\omega_t, \overline{x}_t^i)$ and

$$u(\omega_t, \overline{x}_t, y_t^i) - \pi P_t^i = u(\omega_t, \overline{x}_t, x^b(\omega_t, \overline{x}_t^i)) - \pi_1 P_t^i.$$

Let J denote the set of i which are indifferent and consider a threshold ι in which agents in J with $i > \iota$ play y_t^i and those with $i \le \iota$ play $x^b(\omega_t, \overline{x}_t)$, while those not in J play their unique best-response. Let $\overline{x}^\iota$ be the corresponding average output: clearly $\overline{x}^\iota \in B(\overline{x}_t)$. Moreover $\overline{x} \in B(\overline{x}_t)$ implies $\overline{x}^0 \le \overline{x} \le \overline{x}^1$. Hence $B(\overline{x}_t) = \{\overline{x}^\iota\}_{\iota \in [0,1]}$. Since $\overline{x}^\iota$ is an increasing continuous function it follows that $\{\overline{x}^\iota\}_{\iota \in [0,1]} = [\overline{x}^0, \overline{x}^1]$ which is convex. □

Recall that the monitoring cost for the simple social norm is defined as

$$M \equiv \theta\left[u(\omega_1, \overline{y}^s, x^b(\omega_1, \overline{y}^s)) - u(\omega_1, \overline{y}^s, \overline{y}^s)\right].$$

Since $\overline{y}^s > 0$ Proposition 2 also shows that $x^b(\omega_1, \overline{y}^s) > \overline{y}^s$, and by Proposition 1 this implies that the utility gain $u(\overline{y}^s, x^b(\omega_1, \overline{y}^s)) - u(\omega_1, \overline{y}^s, \overline{y}^s)$ is strictly positive and by assumption $\theta > 0$, so the monitoring difficulty is strictly positive, that is $M > 0$.

Our assumption that f is "small" can now be stated. We assume specifically that $f < M$, that is, the fixed cost of switching to a default social norm in the second period is less than the monitoring cost from maintaining the simple social norm.[16]

Proposition 3. *There is a unique contingent social norm $\overline{y}^c(\omega_2)$ strictly decreasing. Moreover,* $D_2u(\omega_2, \overline{y}^c(\omega_2), \overline{y}^c(\omega_2)) + D_3u(\omega_2, \overline{y}^c(\omega_2), \overline{y}^c(\omega_2)) < 0$.

Proof. From monitoring cost minimization Proposition 2 the objective function with a single type ω_2 is

$$(1+\theta)u(\omega_2, y^c, y^c) - \theta u(\omega_2, y^c, x^b(\omega_2, y^c)).$$

We are interested in the region where $x^b(\omega_2, y^c) \ge y^c$. Since by Proposition 1 $x^b(\omega_2, y^c)$ is decreasing in y^c and strictly positive there is a unique value $Y \le X$ such that $x^b(\omega_2, y^c) \ge y^c$ if and only if $y^c \in [0, Y]$. Hence we restrict our analysis to this interval.

From the envelope theorem and Proposition 1 the derivative is

$$(1+\theta)\left[D_2u(\omega_2, y^c, y^c) + D_3u(\omega_2, y^c, y^c)\right] - \theta D_2u(\omega_2, y^c, x^b(\omega_2, y^c)).$$

[16] Note that if the simple social norm is maintained in the second period in the face of an intervention then the monitoring cost cannot be lower than M but it could be higher if some members violate the quota.

The second derivative is then

$$\begin{aligned}[D_{22}u(\omega_2, y^c, y^c) + 2D_{32}u(\omega_2, y^c, y^c) + D_{33}u(\omega_2, y^c, y^c)] + \\ +2\theta D_{32}u(\omega_2, y^c, y^c) + \theta D_{33}u(\omega_2, y^c, y^c) \\ +\theta \left[D_{22}u(\omega_2, y^c, y^c) - D_{22}u(\omega_j, y^c, x^b(\omega_2, y^c))\right] + \\ -\theta D_{23}u(\omega_2, y^c, x^b(\omega_2, y^c))D_2x^b(\omega_2, y^c)\end{aligned}$$

Everything on the first (split) line is non-positive by #7, #5 and #1. In $[0, Y]$ we have $x^b(\omega_2, y^c) \geq y^c$ and this together with the technical assumption #10 that $D_{223}u(\omega_2, \overline{x}_t, x_t^i) \geq 0$ gives

$$\left[D_{22}u(\omega_2, y^c, y^c) - D_{22}u(\omega_2, y^c, x^b(\omega_2, y^c))\right] \leq 0.$$

From Proposition 1 $D_2x^b(\omega_2, y^c) \leq 0$ so $D_{23}u(\omega_2, y^c, x^b(\omega_2, y^c))D_2x^b(\omega_2, y^c) \geq 0$. Hence we have a strictly concave optimization problem on the relevant domain $[0, Y]$. As for given ω_2 the problem is concave on the relevant domain it is characterized by first order conditions there. Hence we need only check that the derivative with respect to y^c is strictly decreasing in ω_2, implying that $\overline{y}^c(\omega_2)$ is strictly decreasing. Differentiating

$$\begin{aligned}&(1 + \theta)\left[D_{21}u(\omega_2, y^c, y^c) + D_{31}u(\omega_2, y^c, y^c)\right] - \theta D_{21}u(\omega_2, y^c, x^b(\omega_2, y^c)) + \\ &- \theta D_{23}u(\omega_2, y^c, x^b(\omega_2, y^c))D_1x^b(\omega_2, y^c)\end{aligned}$$

and rewriting

$$\begin{aligned}&[D_{21}u(\omega_2, y^c, y^c) + D_{31}u(\omega_2, y^c, y^c)] + \theta D_{31}u(\omega_2, y^c, y^c) + \\ &+ \theta \left[D_{21}u(\omega_2, y^c, y^c) - D_{21}u(\omega_2, y^c, x^b(\omega_2, y^c))\right] + \\ &- \theta D_{23}u(\omega_2, y^c, x^b(\omega_2, y^c))D_1x^b(\omega_2, y^c).\end{aligned}$$

The first terms is non-positive by #9, the second strictly negative by #6, the third non-positive by #11 that$D_{123}u(\omega_2, \overline{x}_t, x_t^i) \geq 0$ and the final one non-positive because $D_{23}u(\omega^i, \overline{x}_t, x_t^i) \leq 0$ (#5) and from Proposition 1.

As we have showed interiority the first order condition holds with equality

$$(1 + \theta)\left[D_2u(\omega_2, y^c, y^c) + D_3u(\omega_2, y^c, y^c)\right] - \theta D_2u(\omega_2, y^c, x^b(\omega_2, y^c)) = 0.$$

If $D_2u(\omega_2, y^c, x^b(\omega_2, y^c)) < 0$ it follows directly that $D_2u(\omega_2, y^c, y^c) + D_3u(\omega_2, y^c, y^c) < 0$. Otherwise rewrite the first order condition as

$$D_2u(\omega_2, y^c, y^c) + (1 + \theta)D_3u(\omega_2, y^c, y^c) + \theta\left[D_2u(\omega_2, y^c, y^c) - D_2u(\omega_2, y^c, x^b(\omega_2, y^c))\right] = 0.$$

If $D_2u(\omega_2, y^c, x^b(\omega_2, y^c)) \geq 0$ then by #5 ($D_{23}u(\omega_2, y^c, x_t^i) \leq 0$) and $y^c < x^b(\omega_2, y^c)$ we have also $D_2u(\omega_2, y^c, y^c) \geq 0$ and that the expression in square brackets is non-negative; by #4 also $D_3u(\omega_2, y^c, y^c) > 0$. Therefore the entire LHS is strictly positive, a contradiction. □

Proof of the Main Result

The idea of the main result is this. As ω_2 increases the simple social norm initially is strictly best and nothing changes. Output at the default social norm is initially higher than the simple social norm and declines. As ω_2 increases there is a critical value at which output from the default social norm equals the quota at the simple social norm (this is the $\hat{\omega}_2$ below). This is where the incentive constraints bind with equality at the simple social norm and for higher ω_2 output at the simple social norm will be exactly the same as the default social norm as the constraint no longer binds. However: at the critical value of ω_2 the default social norm is strictly better than the simple social norm because the simple social norm has positive monitoring cost M and the default social norm has a fixed cost $f < M$. Hence there is another and lower critical cutoff for ω_2 where there is indifference between the simple social norm and the default social norm (this will be $\underline{\omega}_2$). For large enough F so that the contingent social norm is not used this means that at this second lower critical cutoff we switch from simple to default and output jumps up.

For any positive F and small enough ω_2 the simple social norm is always better than the contingent social norm. However, as F declines the range of ω_2 for which this is true declines. In other words, there is also a critical cutoff $\omega_2(F)$ below which the simple social norm is better than contingent and at least locally above which the contingent social norm is better. For large F the critical cutoff is "infinite." As $F \to 0$ this critical cutoff approaches ω_1. Hence there is a critical value of F for which the contingent norm cutoff matches the default social norm cutoff. For F above this we switch from simple to default and output jumps up. Below this we switch from simple to contingent and output jumps down (this is why we want output in the contingent norm decreasing).

If we can also assure that ω_2 can be made large enough that the critical cutoff with respect to the default exists (is not "infinite") then output eventually has to be less than at ω_1. A sufficient condition is #12: $D_2u(\omega_1, \overline{x}_t, \overline{x}_t) + D_3u(\omega_1, \overline{x}_t, \overline{x}_t) < 0 \Rightarrow D_3u(\overline{\omega}, \overline{x}_t, \overline{x}_t) < 0$. It says that to the right of the social optimum (with no monitoring cost) there is an ω_2 big enough that it is individually optimal to move to the left. This means the default must for high enough ω_2 hit the simple social norm (since this has higher output than the social optimum).

Lemma 7. *There is a unique value $\hat{\omega}_2$ for which $\overline{y}^d(\omega_2) = \overline{y}^s$, with $\overline{\omega} > \hat{\omega}_2 > \omega_1$. For $\omega_2 < \hat{\omega}_2$ we have $D_3u(\omega_2, \overline{y}^s, \overline{y}^s) > 0$ and for $\omega_2 > \hat{\omega}_2$ we have $D_3u(\omega_2, \overline{y}^s, \overline{y}^s) < 0$.*

Proof. From Proposition 3 $D_2(\omega_1, \overline{y}^s, \overline{y}^s) + D_3(\omega_1, \overline{y}^s, \overline{y}^s) < 0$ hence by #12 we have $D_3u(\overline{\omega}, \overline{y}^s, \overline{y}^s) < 0$. Since (Lemma 4) the constraint strictly binds $D_3u(\omega_1, \overline{y}^s, \overline{y}^s) > 0$. From #6 we have $D_{31}u(\omega_2, \overline{y}^s, \overline{y}^s) < 0$ so the result follows from the intermediate value theorem. □

For a given simple social norm, $\overline{y}^s$, let $\overline{x}^s(\omega_2, \overline{y}^s)$ denote average output in a Nash equilibrium in the second period corresponding to the new intervention ω_2.

Corollary 1. *Given the $\hat{\omega}_2$ identified in Lemma 7, for $\omega_2 < \hat{\omega}_2$ we have $\overline{x}^s(\omega_2, \overline{y}^s) = \overline{y}^s$, while for $\omega_2 > \hat{\omega}_2$ we have $\overline{x}^s(\omega_2, \overline{y}^s) = \overline{y}^d(\omega_2) < \overline{y}^s$.*

Proof. Observe first that $\overline{x}^s(\omega_2, \overline{y}^s) \leq \overline{y}^s$. To see this recall that the expected punishment in the simple social norm is such that the expected benefit from the best deviation upward is precisely 0. So if we instead had $\overline{x}^s(\omega_2, \overline{y}^s) > \overline{y}^s$ then by #5 and #6, any agent would strictly prefer to produce the quota $\overline{y}^s$ instead of a greater amount, a contradiction. At this point the result follows directly from Lemma 7, Theorem 6 and the fact that if $\overline{x}^s(\omega_2, \overline{y}^s) < \overline{y}^s$ then it must be that $\overline{x}^s(\omega_2, \overline{y}^s) = \overline{y}^d(\omega_2)$. □

Let $\overline{y}^o(\omega_2)$ denote the social optimum output corresponding to ω_2 (which may be the simple norm, the default or the contingent norm). We recall that the contingent and default norms have fixed costs respectively of F and $0 \leq f < F$, and that by assumption $f < M$.

Theorem 8. *For $\omega_2 > \hat{\omega}_2$ we have $\overline{y}^o(\omega_2) < \overline{y}^s$.*

Proof. For $\omega_2 > \hat{\omega}_2$ by Corollary 1 we have $\overline{x}^s(\omega_2, \overline{y}^s) = \overline{y}^d(\omega_2)$. This means it is better to play $\overline{y}^d(\omega_2)$ and pay $M > 0$ than sticking to $\overline{y}^s$; hence the default social norm is strictly better than the simple social norm. Hence candidate $\overline{y}^o(\omega_2)$ are default and contingent. But $\overline{y}^c(\omega_2) < \overline{y}^d(\omega_2) < \overline{y}^s$, the first by Proposition 2(iv)(b) and the second by Corollary 1. The result follows. □

At this point we want to define the threshold $\underline{\omega}_2$ below which the simple norm is better than the default (as we shall show later it is in fact optimal). To this end denote by $V^k(\omega_2)$ utility excluding fixed cost for the social norm $k \in \{d, s, c\}$ when there is an intervention. Then define $G^{jk}(\omega_2) \equiv V^j(\omega_2) - V^k(\omega_2)$ - which is the gain of j over k excluding fixed cost. Thus for example since the simple norm has no fixed cost the default norm yields a higher group payoff than simple when $G^{ds} \geq f$. Notice that this inequality is strict at the previously identified $\hat{\omega}_2$: there $u(\hat{\omega}_2, \overline{y}^s, \overline{y}^s) = V^d(\hat{\omega}_2)$ so that from $V^s(\hat{\omega}_2) = u(\hat{\omega}_2, \overline{y}^s, \overline{y}^s) - M$ we get $G^{ds}(\hat{\omega}_2) = M > f$.

From the parametric transversality theorem we may now assume that f is chosen generically so that on $[0, \overline{\omega}]$ the function $G^{ds}(\omega_2)$ has non-vanishing derivative at the point(s) ω_2 where $G^{ds}(\omega_2) = f$. Hence as transversality implies finitely many crossings we may let $\underline{\omega}_2$ be the smallest solution of $G^{ds}(\omega_2) \geq f$. Thus for $\omega_2 < \underline{\omega}_2$ the simple norm is preferred to the default, and for ω_2 slightly larger than $\underline{\omega}_2$ the opposite holds.

Lemma 8. $0 < \underline{\omega}_2 < \hat{\omega}_2$.

Proof. At ω_1 we have by construction $u(\omega_1, \overline{y}^s, \overline{y}^s) - M > u(\omega_1, \overline{y}^d(\omega_1), \overline{y}^d(\omega_1))$ that is $V^s(\omega_1) > V^d(\omega_1)$ so that $G^{ds}(\omega_1) < 0$. Also $G^{ds}(\hat{\omega}_2) = M > 0$, so the result follows from the mean value theorem. □

The next lemma implies that below $\underline{\omega}_2$ the simple norm is preferred also to the quota $\overline{y}^c(\omega_2)$.

Lemma 9. *For $\omega_2 < \hat{\omega}_2$ we have $V^c(\omega_2) - V^s(\omega_2)$ strictly increasing.*

Proof. Note that as we change ω_2 keeping the simple social norm the punishments remain fixed despite the reduced incentive to deviate, so the monitoring cost does not change. We compute

$$V^c(\omega_2) - V^s(\omega_2) = \left[(1+\theta)u(\omega_2, y^c, y^c) - \theta u(\omega_2, y^c, x^b(\omega_2, y^c))\right] - [u(\omega_2, y^s, y^s) - M].$$

From the envelope theorem the derivative is

$$(1+\theta)D_1u(\omega_2, y^c, y^c) - \theta D_1u(\omega_2, y^c, x^b(\omega_2, y^c)) - D_1u(\omega_2, y^s, y^s).$$

This may be written as

$$[D_1u(\omega_2, y^c, y^c) - D_1u(\omega_2, y^s, y^s)] + \theta\left[D_1u(\omega_2, y^c, y^c) - D_1u(\omega_2, y^c, x^b(\omega_2, y^c))\right].$$

Since $D_{31}u(\omega^i, \overline{x}_t, x_t^i) < 0$ by #6 and $x^b(\omega_2, y^c)) > y^c$ by 4 the final term is positive. For the first term, write

$$D_1u(\omega_2, y^c, y^c) - D_1u(\omega_2, y^s, y^s) = \int_{y^s}^{y^c} [D_{12}u(\omega_2, y, y) + D_{13}u(\omega_2, y, y)]\, dy.$$

Moreover $\overline{y}^c(\omega_2)$ strictly decreasing from Proposition 3 implies $y^c < y^s$ so from #9 this is non-negative. □

Next define $\underline{F} = G^{cs}(\underline{\omega}_2)$. Thus $\underline{F}$ makes simple and contingent indifferent where simple is also indifferent to the default norm; for higher F the simple and default norm are preferred to the contingent norm (at $\underline{\omega}_2$), for lower F it is the reverse.

Theorem 9 (Pigou, Upward Jump). *For $F > \underline{F}$ there exists $\epsilon > 0$ such that for*

(i) $\omega_2 < \underline{\omega}_2$ we have $\overline{y}^o(\omega_2) = \overline{y}^s$

(ii) $\underline{\omega}_2 < \omega_2 < \underline{\omega}_2 + \epsilon$ we have $\overline{y}^o(\omega_2) = \overline{y}^d(\omega_2) > \overline{y}^s + \epsilon$ and decreasing in ω_2

In addition there is an $F^+ > \underline{F}$ such that

(iii) $F > F^+$ and $\omega_2 > \underline{\omega}_2$ imply $\overline{y}^o(\omega_2)$ is decreasing

Proof. For $F > \underline{F}$ the simple and default norm are preferred to the contingent norm at $\underline{\omega}_2$; below that simple is better than default by definition and better than contingent by Lemma 9. This proves (i).

Next we claim that there is an ϵ so that for $\underline{\omega}_2 < \omega_2 < \underline{\omega}_2 + \epsilon$ it is optimal to use the default social norm. For small enough ϵ it is strictly sub-optimal to use the contingent social norm. At $\underline{\omega}_2$ we have $G^{ds}(\omega_2) = f$ so by the generic choice of f we know $DG^{ds}(\omega_2) \neq 0$. By construction for $0 \leq \omega_2 < \underline{\omega}_2$ we have $G^{ds}(\omega_2) < f$ so in fact $DG^{ds}(\omega_2) > 0$. It follows that for ϵ sufficiently small it is optimal to use the default social norm. Moreover, since $\underline{\omega}_2 < \hat{\omega}_2$, $\overline{y}^d(\hat{\omega}_2) = \overline{y}^s$ (Lemma 7) and $\overline{y}^d(\omega_2)$ strictly decreasing (by Theorem 6) it follows that $\overline{y}^d(\underline{\omega}_2) > \overline{y}^s$and consequently this remains true for ϵ small enough, proving (ii).

For (iii) observe that for F sufficiently large it cannot be optimal ever to use $\overline{y}^c$. Hence the jump from $\overline{y}^s$ to $\overline{y}^d$ once it occurs remains so for larger ω_2. □

For $f < F < \underline{F}$ let $\underline{\omega}_2^F$ be the unique solution in $[0, \hat{\omega}_2]$ of $G^{cs}(\omega_2) = F$. Observe that by definition $G^{cs}(\underline{\omega}_2) = \underline{F} > f$, and by Lemma 9 $G^{cs}(\omega_2)$ is strictly increasing; so $0 < \underline{\omega}_2^F < \underline{\omega}_2$, and in fact to the left of $\underline{\omega}_2^F$ the simple norm is better than contingent and to the right the opposite is true. Also recall that for $\omega_2 < \underline{\omega}_2$ the simple norm is better than the default norm.

Theorem 10 (Pigou, Downward Jump). *For $0 < F < \underline{F}$ there exists $\epsilon > 0$ such that for*

(i) $0 \leq \omega_2 < \underline{\omega}_2^F$ we have $\overline{y}^o(\omega_2) = \overline{y}^s$

(ii) $\underline{\omega}_2^F < \omega_2 < \underline{\omega}_2^F + \epsilon$ we have $\overline{y}^o(\omega_2) = \overline{y}^c(\omega_2) < \overline{y}^s - \epsilon$ and decreasing in ω_2

In addition there is an $F^- < \underline{F}$ such that

(iii) $F < F^-$ implies $\overline{y}^o(\omega_2) \leq \overline{y}^s$ and for $\omega_2 > \underline{\omega}_2$ we have $\overline{y}^o(\omega_2) < \overline{y}^s$

Proof. For (i) and (ii) we may assume that $\omega_2 < \underline{\omega}_2$ so it is strictly sub-optimal to use the default social norm. But as we just observed to the left of $\underline{\omega}_2^F$ the simple norm is better than contingent norm and to the right the opposite is true. So (i) follows directly, and (ii) follows from $\overline{y}^c(\omega_1) = \overline{y}^s$ and the fact that $\overline{y}^c(\omega_2)$ is strictly decreasing (Proposition 3).

For part (iii) observe that at $F = f$ the new social norm $\overline{y}^c(\omega_2)$ is strictly better than the default social norm so if we take $F^- < \min_{\omega_2 \leq \underline{\omega}_2} G^{cd}(\omega_2)$ to the left of $\underline{\omega}_2^F$ the simple social norm is best; in $(\underline{\omega}_2^F, \underline{\omega}_2]$ the new social norm is best, and in either case we have $\overline{y}^o(\omega_2) \leq \overline{y}^s$. To the right of $\underline{\omega}_2$ the simple social norm cannot be best, and both the default and new social norm have less output than $\overline{y}^s$ so indeed $\overline{y}^o(\omega_2) < \overline{y}^s$. □

Output Limits

The tax is given by $\tau = \Phi\Omega(x_t^i, \Lambda)$ where Ω is the probability of being caught when minimum intended output is x_t^i and the limit is Λ. We assume the lowest possible speed limit $\underline{\Lambda} \geq X$ the highest minimum speed to ensure that $\Omega \leq 1$.

Suppose that $\tilde{a}$ is a standard Pareto and that actual output is $\tilde{a}x_t^i$. Let $0 < \kappa \leq 1$ be the probability of getting caught if in fact the limit is exceeded. Using the fact that the inverse of a standard Pareto is a standard uniform we may compute the probability of being caught when intended output is x_t^i as

$$\Omega(x_t^i, \Lambda) = \kappa \Pr(\tilde{a}x_t^i > \Lambda) = \kappa \Pr(\Lambda^{-1}x_t^i > 1/\tilde{a}) = \kappa\Lambda^{-1}x_t^i.$$

Hence $\tau = \kappa\Phi\Lambda^{-1}x_t^i$.

Optimal Taxes

In the one-type Pigouvian case define

$$W^P(\omega_2, \alpha) = (1+\theta)u(\omega_2, y^c(\omega_2), y^c(\omega_2)) - \theta u(\omega_2, y^c(\omega_2), x^b(\omega_2, y^c(\omega_2))).$$

Theorem 11. *$W^P(\omega_2, 1)$ is strictly increasing.*

Proof. From the envelope theorem the derivative with respect to ω_2 is

$$\frac{\partial W^P(\omega_2, \alpha)}{\partial \omega_2} = D_1 u(\omega_2, y^c, y^c) + \theta \left[D_1 u(\omega_2, y^c, y^c) - D_1 u(\omega_2, y^c, x^b(\omega_2, y^c)) \right],$$

and applying the Pigouvian functional form this is

$$\frac{\partial W^P(\omega_2, \alpha)}{\partial \omega_2} = -(1-\alpha) y^c + \theta \left[x^b(\omega_2, y^c) - y^c \right].$$

Observe that since $x^b(\omega_2, y^c) > y^c$ the final term is strictly positive, so if $\alpha = 1$ indeed $W^P(\omega_2, 1)$ is strictly increasing. □

Theorem 12. *For given θ there exists $\overline{\alpha} < 1$ such that for $\overline{\alpha} < \alpha \leq 1$ the optimal tax $\omega_2^*(\alpha)$ is unique and smooth with $\omega_2^*(1)$ the Pigouvian tax. If in addition $((1+\theta)U''(x^*) - L''(x^*))\, x^* + \theta \omega^* < 0$ and in particular for θ sufficiently small $\omega_2^*(\alpha)$ is strictly differentiably increasing in α and conversely.*

Proof. With the Pigouvian functional form we may write

$$W^P(\omega_2, y^c, \alpha) = (1+\theta) \left(U(y^c) - \omega_2 y^c - L(y^c) + \alpha \omega_2 y^c \right) - \theta \left(U(x^b(\omega_2, y^c)) - \omega_2 x^b(\omega_2, y^c) - L(y^c) + \alpha \omega_2 y^c \right),$$

and noting that for Pigou x^b does not depend on $\overline{x}_t$, we can further simplify this to

$$W^P(\omega_2, y^c, \alpha) = -(1+\theta-\alpha)\omega_2 y^c + (1+\theta)U(y^c) - L(y^c) - \theta \left(U(x^b(\omega_2)) - \omega_2 x^b(\omega_2) \right).$$

We then compute the derivatives. With respect to ω_2 we have

$$D_1 W^P = -(1+\theta-\alpha) y^c + \theta x^b(\omega_2) - \theta \left(U'(x^b(\omega_2)) - \omega_2 \right) \partial x^b(\omega_2)/\partial \omega_2$$

where the last expression is zero by the envelope theorem, so this simplifies to

$$D_1 W^P = -(1+\theta-\alpha) y^c + \theta x^b(\omega_2).$$

From this we may compute second derivatives

$$D_{11} W^P = \theta \frac{\partial x^b(\omega_2)}{\partial \omega_2} = \theta \frac{1}{U''(x^b(\omega_2))} < 0$$

the second equality from Proposition 1. We also have, with respect to α,

$$D_{13} W^P = y^c.$$

Next, with respect to y^cwe have

$$\begin{gathered} D_2W^P = -(1+\theta-\alpha)\omega_2 + (1+\theta)U'(y^c) - L'(y^c) \\ D_{22}W^P = (1+\theta)U''(y^c) - L''(y^c) \\ D_{23}W^P = \omega_2 \qquad D_{21}W^P = -(1+\theta-\alpha). \end{gathered}$$

Next we consider $\alpha = 1$ in which case we know that $x^b = y^c = x^*$ and $\omega_2 = \omega^*$ to find

$$\begin{gathered} D_{11}W^P = \theta/U''(x^*) \\ D_{22}W^P = (1+\theta)U''(x^*) - L''(x^*) < 0 \\ D_{21}W^P = -\theta \qquad D_{13}W^P = y^c \qquad D_{23}W^P = \omega^*. \end{gathered}$$

From the implicit function theorem we have

$$\begin{bmatrix} d\omega_2 \\ dy^c \end{bmatrix} = -\begin{bmatrix} D_{11}W^P & D_{12}W^P \\ D_{12}W^P & D_{22}W^P \end{bmatrix}^{-1} \begin{bmatrix} D_{13}W^P \\ D_{23}W^P \end{bmatrix} = -\frac{1}{\Delta}\begin{bmatrix} D_{22}W^P & -D_{12}W^P \\ -D_{12}W^P & D_{11}W^P \end{bmatrix} \begin{bmatrix} D_{13}W^P \\ D_{23}W^P \end{bmatrix}$$

where Δ is the determinant of the matrix. First the determinant:

$$\Delta/\theta = \frac{1}{U''(x^*)}\left[(1+\theta)U''(x^*) - L''(x^*)\right] - \theta = 1 - \frac{L''(x^*)}{U''(x^*)}$$

which is positive. In particular the matrix is negative definite, so there is a unique optimum ω_2, y^c.

Finally

$$d\omega_2 = -\frac{1}{\Delta}\left[\left((1+\theta)U''(x^*) - L''(x^*)\right)x^* + \theta\omega^*\right]$$

which is positive if and only if

$$\left((1+\theta)U''(x^*) - L''(x^*)\right)x^* + \theta\omega^* < 0.$$

□

Application Assumptions

To check that assumptions #1 to #12 are verified in the applications it is convenient to list them by the order of derivative to which they apply. We also give equivalent versions which are easier to work with, justifying equivalence case by case.

First derivative assumptions

		equivalent
2	$D_3u(\omega^i, \overline{x}_t, 0) \geq 0$	$D_3u(\overline{\omega}, \overline{x}_t, 0) \geq 0$ (by #6)
3	$D_3u(\omega^i, \overline{x}_t, X) \leq 0$	$D_3u(\omega_1, \overline{x}_t, X) \leq 0$ (by #6)
4	$D_2u(\omega^i, \overline{x}_t, x_t^i) \geq 0$ and $x_t^i > 0 \Rightarrow D_3u(\omega^k, \overline{x}_t, \overline{x}_t) > 0$	
8	$D_2u(\omega^i, 0, 0) + D_3u(\omega^i, 0, 0) > 0$	$D_2u(\overline{\omega}, 0, 0) + D_3u(\overline{\omega}, 0, 0) > 0$ (by #9)
12	$D_2u(\omega_1, \overline{x}_t, \overline{x}_t) + D_3u(\omega_1, \overline{x}_t, \overline{x}_t) < 0 \Rightarrow D_3u(\overline{\omega}, \overline{x}_t, \overline{x}_t) < 0$	

Second derivative assumptions

$u(\omega^i, \overline{x}_t, \overline{x}_t)$ strictly differentiably concave in $\overline{x}_t$ (which is #7), since first derivative is $D_2u(\omega^i, \overline{x}_t, \overline{x}_t) + D_3u(\omega^i, \overline{x}_t, \overline{x}_t)$ implies that the following second derivative is negative:

$$D_{22}u(\omega^i, \overline{x}_t, \overline{x}_t) + 2D_{23}u(\omega^i, \overline{x}_t, \overline{x}_t) + D_{33}u(\omega^i, \overline{x}_t, \overline{x}_t) < 0.$$

Therefore the second derivatives assumptions are:

1	$D_{33}u(\omega^i, \overline{x}_t, x_t^i) < 0$
5	$D_{32}u(\omega^i, \overline{x}_t, x_t^i) \leq 0$
6	$D_{31}u(\omega^i, \overline{x}_t, x_t^i) < 0$
7	$D_{22}u(\omega^i, \overline{x}_t, \overline{x}_t) + 2D_{23}u(\omega^i, \overline{x}_t, \overline{x}_t) + D_{33}u(\omega^i, \overline{x}_t, \overline{x}_t) < 0$
9	$D_{21}u(\omega^i, \overline{x}_t, x_t^i) + D_{31}u(\omega^i, \overline{x}_t, x_t^i) \leq 0$

Third derivative assumptions

11	$D_{231}u(\omega_j, \overline{x}_t, x_t^i) \geq 0$
10	$D_{232}u(\omega_j, \overline{x}_t, x_t^i) \geq 0$

Pigou

Recall that in our first application we consider a simple negative externality where the intervention is a Pigouvian tax. Effort x_t^i brings an individual benefit $U(x_t^i)$ which is strictly concave $U''(x_t^i) < 0$ and a social cost $L(\overline{x}_t^i)$ strictly increasing $L'(\overline{x}_t) > 0$ and weakly convex $L''(\overline{x}_t) \geq 0$. The Pigouvian tax is $\omega^i x_t^i$. A portion of the tax $\alpha \in [0, 1]$ is returned to the group as an equally distributed lump sum, with the remainder going to the outside agency imposing the tax. Therefore

$$u(\omega^i, \overline{x}_t, x_t^i) = U(x_t^i) - \omega^i x_t^i - L(\overline{x}_t) + \alpha\omega^i\overline{x}_t.$$

We impose two boundary conditions, the first that the individual marginal benefit is large at the lower bound in the sense that $U'(0) > L'(0) + \overline{\omega}$ and the second that the upper bound is sufficiently large that individual benefit is no longer strictly increasing $U'(X) \leq 0$.

To ensure that there is an adequate range of policy interventions we assume that the initial marginal Pigouvian tax is not too large in the sense that $L'(X) > \alpha\omega_1$. In the case where the initial $\omega_1 = 0$ this is certainly the case, but we allow the possibility that the initial tax is positive, just not too large.

Under these assumptions Lemma 11 that there is a unique solution x^* to the group optimization problem of maximizing $u(\omega_1, \overline{x}_t, \overline{x}_t)$ that lies in the interior, and that $U'(x^*) > \overline{\omega}$ meaning that the initial tax rate ω_1 is in fact sufficiently low that at the social optimum individuals would like to increase effort. Our final assumption is that the highest tax rate is high enough that it becomes individually optimal to implement x^*, that is, $U'(x^*) - \overline{\omega} = 0$. In the case where ω_1 corresponds to no tax and there is a full rebate $\alpha = 1$, this says that $\overline{\omega}$ is "the" Pigouvian tax. We do not examine the consequences of setting tax rates higher than this.

Computation of derivatives.

$D_1 u(\omega^i, \overline{x}_t, x_t^i) = -x_t^i + \alpha \overline{x}_t$
$D_2 u(\omega^i, \overline{x}_t, x_t^i) = -L'(\overline{x}_t) + \alpha \omega^i$
$D_3 u(\omega^i, \overline{x}_t, x_t^i) = U'(x_t^i) - \omega^i)$

$D_{11} u(\omega^i, \overline{x}_t, x_t^i) = -D_{11}\tau(\omega^i, x_t^i)$	$D_{12} u(\omega^i, \overline{x}_t, x_t^i) = \alpha$	$D_{13} u(\omega^i, \overline{x}_t, x_t^i) = -1$
$D_{21} u(\omega^i, \overline{x}_t, x_t^i) = \alpha$	$D_{22} u(\omega^i, \overline{x}_t, x_t^i) = -L''(\overline{x}_t)$	$D_{23} u(\omega^i, \overline{x}_t, x_t^i) = 0$
$D_{31} u(\omega^i, \overline{x}_t, x_t^i) = -1$	$D_{32} u(\omega^i, \overline{x}_t, x_t^i) = 0$	$D_{33} u(\omega^i, \overline{x}_t, x_t^i) = U''(x_t^i)$

The maintained assumptions for the Pigou application (except for the last one which is stated after Lemma 11) are the following:

$L'(\overline{x}_t) > 0$

$U'(0) > L'(0) + \overline{\omega}$

$U'(X) \leq 0$

$L'(X) > \omega_1$

$U''(x_t^i) < 0$, $L''(\overline{x}_t) \geq 0$

Lemma 10. *To verify #8 we must show that* $-L'(0) + U'(0) - (1-\alpha)\overline{\omega} > 0$

Proof. Follows from $U'(0) > L'(0) + \overline{\omega}$. □

Lemma 11. *There is a unique maximizer* x^* *of* $U(\overline{x}_t) - L(\overline{x}_t) - (1-\alpha)\omega_1 \overline{x}_t$*, and it is strictly interior.*

Proof. The objective function is concave by assumption with interiority from 10 and the fact that $L'(\overline{x}_t) > 0$, $U'(X) \leq 0$ a imply that on the upper boundary $-L'(X) + U'(X) - (1-\alpha)\omega_1 < 0$. □

At this point we can state our final assumption: $U'(x^*) - \overline{\omega} = 0$.

In the following tables for each assumption we stack the original version and the form it takes in the application. The last column contains a proof that the assumption is verified (or reference thereof). First derivative assumptions (using the equivalent formulations given above):

	requirement	reason
2	$D_3u(\overline{\omega}, \overline{x}_t, 0) \geq 0$: $U'(0) - \overline{\omega}$	$U'(0) > L'(0) + \overline{\omega}$, $L'(0) > 0$
3	$D_3u(\omega_1, \overline{x}_t, X) \leq 0$: $U'(X) - \omega_1 \leq 0$	$U'(X) \leq 0, \omega_1 \geq 0$
4	$D_2u(\omega^i, \overline{x}_t, x_t^i) \geq 0$ and $x_t^i > 0 \Rightarrow D_3u(\omega^k, \overline{x}_t, \overline{x}_t) > 0$: $-L'(\overline{x}_t) + \alpha\omega^i \geq 0 \Rightarrow U'(\overline{x}_t) - \omega^i > 0$	Lemma 12 below
8	$D_2u(\overline{\omega}, 0, 0) + D_3u(\overline{\omega}, 0, 0) > 0$: $-L'(0) + U'(0) - (1 - \alpha)\overline{\omega} > 0$	Lemma 10 above
12	$D_2u(\omega_1, \overline{x}_t, \overline{x}_t) + D_3u(\omega_1, \overline{x}_t, \overline{x}_t) < 0 \Rightarrow D_3u(\overline{\omega}, \overline{x}_t, \overline{x}_t) < 0$: $-L'(\overline{x}_t) + \alpha\omega_1 + U'(\overline{x}_t) - \omega_1 < 0 \Rightarrow U'(\overline{x}_t) - \overline{\omega} < 0$	Lemma 13 below

Second derivative assumptions:

	requirement	reason
1	$D_{33}u(\omega^i, \overline{x}_t, x_t^i) < 0$: $U''(x_t^i) < 0$	assumption
5	$D_{32}u(\omega^i, \overline{x}_t, x_t^i) \leq 0$: $0 \leq 0$	true
6	$D_{31}u(\omega^i, \overline{x}_t, x_t^i) < 0$: $-1 < 0$	true
7	$D_{22}u(\omega^i, \overline{x}_t, \overline{x}_t) + 2D_{23}u(\omega^i, \overline{x}_t, \overline{x}_t) + D_{33}u(\omega^i, \overline{x}_t, \overline{x}_t) < 0$: $-L''(\overline{x}_t) + U''(\overline{x}_t) < 0$	assumption
9	$D_{21}u(\omega^i, \overline{x}_t, \overline{x}_t) + D_{31}u(\omega^i, \overline{x}_t, \overline{x}_t) \leq 0$: $\alpha - 1 \leq 0$	$\alpha \in [0, 1]$

Third derivative assumptions:

	requirement	reason
11	$D_{231}u(\omega_j, \overline{x}_t, x_t^i) \geq 0$: $0 \geq 0$	true
10	$D_{232}u(\omega_j, \overline{x}_t, x_t^i) \geq 0$: $0 \geq 0$	true

Lemma 12. *To verify #4 we must show that* $-L'(\overline{x}_t) + \alpha\omega^i \geq 0 \Rightarrow U'(\overline{x}_t) - \omega^i > 0$.

Proof. First we show that the hypothesis - which we re-write as $L'(\overline{x}_t) - \alpha\omega^i \leq 0$ - implies that $\overline{x}_t < x^*$. Recall that $U'(x^*) - \omega_1 - L'(x^*) + \alpha\omega_1 = 0$; then $\omega_1 < \overline{\omega}$ we get $L'(x^*) - \alpha\omega_1) = U'(x^*) - \omega_1 > U'(x^*) - \overline{\omega} = 0$. Since by assumption $L''(\overline{x}_t)) \geq 0$, the hypothesis $L'(\overline{x}_t) - \alpha\omega^i) \leq 0$ implies $\overline{x}_t < x^*$.

Assume then that $\overline{x}_t < x^*$. From $U'(x^*) - \omega^i \geq 0$ and $U''(\overline{x}_t) < 0$ we draw the desired conclusion that $U'(\overline{x}_t) - \omega^i > 0$. □

Lemma 13. *To verify #12 we must show that*

$$-L'(\overline{x}_t) + \alpha D_2\tau(\omega_1, \overline{x}_t) + U'(\overline{x}_t) - \omega_1 < 0 \Rightarrow U'(\overline{x}_t) - \overline{\omega} < 0$$

Proof. The contrapositive is $U'(\overline{x}_t) - \overline{\omega} \geq 0 \Rightarrow -L'(\overline{x}_t) + \alpha\omega_1 + U'(\overline{x}_t) - \omega_1 \geq 0$. First the hypothesis implies $\overline{x}_t \leq x^*$. Indeed, from $U'(x^*) - \overline{\omega} = 0$ and $U''(\overline{x}_t) < 0$ the hypothesis $U'(\overline{x}_t) - \overline{\omega} \geq 0$ implies that $\overline{x}_t \leq x^*$.

On the other hand $-L'(x^*) + \alpha\omega_1 + U'(x^*) - \omega_1 = 0$ $(-L''(\overline{x}_t) + U''(\overline{x}_t) < 0)$ show that for $\overline{x}_t \leq x^*$ we have $-L'(\overline{x}_t) + \alpha\omega_1 + U'(\overline{x}_t) - \omega_1 \geq 0$. □

Cournot

Recall that in the cartel application we have $u(\omega^i, \overline{x}_t, x_t^i) = p(\overline{x}_t)x_t^i - c(\omega^i, x_t^i)$, and that revenue is $r(\overline{x}_t) \equiv p(\overline{x}_t)\overline{x}_t$. Price is positive, strictly downwards sloping, convex and marginal revenue is strictly downwards sloping. Marginal cost D_2c is positive, strictly upwards sloping and strictly increasing in ω_2. The capacity constraint is $x_t^i \leq X$. The highest type $\overline{\omega}$ is willing to enter at the lowest price $p(X)$ and the lowest type ω_1 is not willing to produce to capacity at the highest possible price $p(0)$. Under these assumptions at ω_1 there is a unique level of monopoly output x^m - maximizing $r(\overline{x}_t) - c(\omega_1, \overline{x}_t)$ - which is interior (Lemma 14 below), and we assume that at the monopoly level of output corresponding to ω_1 the marginal cost for highest type $\overline{\omega}$ is higher than the monopoly price. In summary the list of assumptions for this case is:

$p(\overline{x}_t) > 0$
$p'(\overline{x}_t) < 0$
$p''(\overline{x}_t) \geq 0$
$r''(\overline{x}_t) < 0$
$D_2c(\omega^i, x_t^i) > 0$
$D_{22}c(\omega^i, x_t^i) > 0$
$D_{21}c(\omega^i, x_t^i) > 0$
$p(X) - D_2c(\overline{\omega}, 0) \geq 0$
$p(0) - D_2c(\omega_1, X) \leq 0$
$p(x^m) - D_2c(\overline{\omega}, x^m) < 0$

We compute as before first and second derivatives:

$D_1u(\omega^i, \overline{x}_t, x_t^i) = -D_1c(\omega^i, x_t^i)$
$D_2u(\omega^i, \overline{x}_t, x_t^i) = p'(\overline{x}_t)x_t^i$
$D_3u(\omega^i, \overline{x}_t, x_t^i) = p(\overline{x}_t) - D_2c(\omega^i, x_t^i)$

$D_{11}u(\omega^i, \overline{x}_t, x_t^i) = -D_{12}c(\omega^i, x_t^i)$	$D_{12}u(\omega^i, \overline{x}_t, x_t^i) = 0$	$D_{13}u(\omega^i, \overline{x}_t, x_t^i) = -D_{12}c(\omega^i, x_t^i)$
$D_{21}u(\omega^i, \overline{x}_t, x_t^i) = 0$	$D_{22}u(\omega^i, \overline{x}_t, x_t^i) = p''(\overline{x}_t)x_t^i$	$D_{23}u(\omega^i, \overline{x}_t, x_t^i) = p'(\overline{x}_t)$
$D_{31}u(\omega^i, \overline{x}_t, x_t^i) = -D_{21}c(\omega^i, x_t^i)$	$D_{32}u(\omega^i, \overline{x}_t, x_t^i) = p'(\overline{x}_t)$	$D_{33}u(\omega^i, \overline{x}_t, x_t^i) = -D_{22}c(\omega^i, x_t^i)$

First derivative assumptions

	requirement	reason
2	$D_3u(\omega^i, \overline{x}_t, 0) \geq 0$: $p(\overline{x}_t) - D_2c(\omega^i, 0) \geq 0$	assumption plus $p'(\overline{x}_t) < 0$ and $D_{21}c(\omega^i, x^i_t) > 0$
3	$D_3u(\omega^i, \overline{x}_t, X) \leq 0$: $p(\overline{x}_t) - D_2c(\omega^i, X) \leq 0$	assumption plus $p'(\overline{x}_t) < 0$ and $D_{22}c(\omega^i, x^i_t) > 0$
4a	$D_2u(\omega^i, \overline{x}_t, x^i_t) < 0$ for $x^i_t > 0$: $p'(\overline{x}_t)x^i_t < 0$ for $x^i_t > 0$	$p'(\overline{x}_t) < 0$
8	$D_2u(\omega^i, 0, 0) + D_3u(\omega^i, 0, 0) > 0$: $p(0) - D_2c(\omega^i, 0) > 0$	from $p(\overline{x}_t) - D_2c(\overline{\omega}, 0) \geq 0$ for $\overline{x}_t = X$ and $p'(\overline{x}_t) < 0$ and $D_{21}c(\omega^i, x^i_t) > 0$
12	$D_2u(\omega_1, \overline{x}_t, \overline{x}_t) + D_3u(\omega_1, \overline{x}_t, \overline{x}_t) < 0 \Rightarrow D_3u(\overline{\omega}, \overline{x}_t, \overline{x}_t) < 0$: $p'(\overline{x}_t)\overline{x}_t + p(\overline{x}_t) - D_2c(\omega_1, \overline{x}_t) < 0$ $\Rightarrow p(\overline{x}_t) - D_2c(\overline{\omega}, \overline{x}_t) < 0$	Lemma 15 below

Second derivative assumptions

	requirement	reason
1	$D_{33}u(\omega^i, \overline{x}_t, x^i_t) < 0$: $-D_{22}c(\omega^i, x^i_t) < 0$	assumption
5	$D_{32}u(\omega^i, \overline{x}_t, x^i_t) \leq 0$: $p'(\overline{x}_t) \leq 0$	assumption
6	$D_{31}u(\omega^i, \overline{x}_t, x^i_t) < 0$: $-D_{21}c(\omega^i, x^i_t) < 0$	assumption
7	$D_{22}u(\omega^i, \overline{x}_t, \overline{x}_t) + 2D_{23}u(\omega^i, \overline{x}_t, \overline{x}_t) + D_{33}u(\omega^i, \overline{x}_t, \overline{x}_t) < 0$: $p''(\overline{x}_t)\overline{x}_t + 2p'(\overline{x}_t) - D_{22}c(\omega^i, \overline{x}_t) < 0$	equivalent to $r''(\overline{x}_t) < 0$
9	$D_{21}u(\omega^i, \overline{x}_t, \overline{x}_t) + D_{31}u(\omega^i, \overline{x}_t, \overline{x}_t) \leq 0$: $-D_{21}c(\omega^i, \overline{x}_t) \leq 0$	assumption

Third derivative assumptions

	requirement	reason
11	$D_{231}u(\omega_j, \overline{x}_t, x^i_t) \geq 0$: $0 \geq 0$	true
10	$D_{232}u(\omega_j, \overline{x}_t, x^i_t) \geq 0$: $p''(\overline{x}_t) \geq 0$	assumption

Lemma 14. *x^m, the solution to maximizing $r(\overline{x}_t) - c(\omega_1, \overline{x}_t)$, is unique and interior*

Proof. The second derivative is $r''(\overline{x}_t) - D_{22}c(\omega_1, \overline{x}_t)$ which is negative from $r''(\overline{x}_t) < 0$ and $D_{22}c(\omega^i, x^i_t) > 0$ so the solution is unique. The derivative is

$$p'(\overline{x}_t)\overline{x}_t + p(\overline{x}_t) - D_2c(\omega_1, \overline{x}_t).$$

From $p(X) - D_2c(\overline{\omega}, 0) \geq 0$, $p'(\overline{x}_t) < 0$ and $D_{21}c(\omega^i, x^i_t) > 0$ we have $p(0) - D_2c(\omega_1, 0) > 0$ so the solution does not lie on the lower boundary. One the upper boundary $p(0) - D_2c(\omega_1, X) \leq 0$ and

$p'(\overline{x}_t) < 0$ imply $p(X) - D_2c(\omega_1, X) < 0$, and $p'(\overline{x}_t) < 0$ implies

$$p'(X)X + p(X) - D_2c(\omega_1, X) < p(X) - D_2c(\omega_1, X).$$

□

Lemma 15. $p'(\overline{x}_t)\overline{x}_t + p(\overline{x}_t) - D_2c(\omega_1, \overline{x}_t) < 0 \Rightarrow p(\overline{x}_t) - D_2c(\overline{\omega}, \overline{x}_t) < 0$

Proof. We may write this as $r'(\overline{x}_t) - D_2c(\omega_1, \overline{x}_t) < 0 \Rightarrow p(\overline{x}_t) - D_2c(\overline{\omega}, \overline{x}_t) < 0$. Since $r'(x^m) - D_2c(\omega_1, x^m) = 0$ and $r''(\overline{x}_t) < 0$, $D_{22}c(\omega^i, x_t^i) > 0$ the condition is satisfied if and only if $\overline{x}_t > x^m$. By assumption $p(x^m) - D_2c(\overline{\omega}, x^m) < 0$ and from $p'(\overline{x}_t) < 0$, $D_{22}c(\omega^i, x_t^i) > 0$ the result follows. □