Digital Dystopia

Jean Tirole

ONLINE APPENDIX

Assortative matching: Reputation as a positional good (Section I)

Consider a future relationship, with partners potentially exercising externality e_2 on the other. Let c_2 denote the date-2 cost of providing this externality, drawn from the uniform distribution on $[0, 1]$. So the probability that agent i provides the externality when her type is v is $\Pr(ve_2 \geq c_2) = ve_2$. So, if $\hat{F}_i(v)$ is the posterior distribution on v_i, the expected externality created by agent i is $[\int_0^1 ve_2 d\hat{F}_i(v)]e_2 = \hat{v}_i e_2^2$.

Agents optimally match with agents of the same reputation (they do not have access to agents with a better reputation). Anticipating a bit, those who have chosen $a_i \equiv 1$ choose as partners agents who have done so as well. Letting v^* denote the cutoff under which agents no longer contribute, the total externality enjoyed by all agents is independent of v^*:

$$\left[F(v^*) \left[\frac{\int_0^{v^*} v dF(v)}{F(v^*)} \right] + [1 - F(v^*)] \left[\frac{\int_{v^*}^1 v dF(v)}{1 - F(v^*)} \right] \right] e_2^2 = \bar{v} e_2^2.$$

Proof of observation on observable compliance (Section III (a))

Suppose, first, that $\operatorname{supp} G = \mathbb{R}^+$. In the case of unobservable b_i studied so far, there were only two equilibrium behaviors, $a_i = b_i = 1$ and $a_i = b_i = 0$. Therefore, observing b_i contained no information that was not already in the social score. The equilibrium characterized in Proposition 4 is still an equilibrium.

Suppose next that the support of G includes negative values of θ_i as well. Let $\hat{v}_{00}$ and $\hat{v}_{01}$ denote the reputation following $\{a_i = b_i = 0\}$ and $\{a_i = 0, b_i = 1\}$, respectively (both are associated with rating 0); and let $\hat{v}_1 = \hat{v}_{11}$ be the reputation following $\{a_1 = b_1 = 1\}$. Among those who choose $a_i = 0$, those with $\theta > \theta^*$ choose $b_i = 0$, where $\nu(\hat{v}_{00} - \hat{v}_{01}) + \theta^* = 0$. I claim that $\theta^* < 0$. Indeed, the corresponding cutoffs satisfy,[1] for $\theta \geq \theta^*$, $v_{00}^*(\theta) = v_{01}^* + (\theta - \theta^*)/e$, and so $\hat{v}_{01} < \hat{v}_{00}$. The intuition behind this result is again that dissenters have an excuse for not engaging in prosocial acts because they cannot obtain a good social rating anyway. The impact of bundling on $E[b_i]$ is less clear than when b_i is unobservable by future partners. As earlier, bundling induces some $\theta_i > 0$ types to choose $b_i = 1$. Types $\{\theta_i \in [\theta^*, 0], v_i < v_{00}^*(\theta)\}$ choose $b_1 = 0$ while they selected $b_i = 1$ in the absence of bundling: They are in search of an excuse. ∎

---

[1] Existence as earlier follows from Brouwer's fixed-point theorem.

Non image sanctions (Section III (b))

To fix ideas, suppose that the state can impose economic sanctions P on blacklisted agents, at deadweight loss $L(P)$ with $L(0) = 0$, $L' > 0$ and $L'' > 0$. Underconsumption underlies this deadweight loss. Consider a social rating in which agents who do not select $a = b = 1$ are blacklisted. Blacklisting implies both an image penalty $\hat{v}_1 - \hat{v}_0$ and an economic penalty P, as is the case in Chinese pilots.[2] The government's welfare becomes

$$V = W - E[L(P)] + \gamma E[b_i].$$

The cutoff $v^b(\theta)$, if interior, is now given by

$$v^b(\theta)e - c + \nu\hat{v}_1 - \max\{\theta, 0\} = \nu\hat{v}_0 - P.$$

Everything is as if the cost of engaging in prosocial behavior were $c - P$ instead of c. This implies that $v^b(\theta)$ has the same shape as in Figure 1 (with a kink at 0), just shifted down. [By contrast under unbundling, the cutoff is still v^u. And $b = 1$ if and only if $\theta \leq P$.]

A complete characterization of economic sanctions lies out of the scope of this paper (for one thing, there is no reason for P to be the same under bundling and unbundling).

---

[2]In these pilots, the penalty may take the form of non-access for blacklisted agents to "discounts" in some public enterprises. An alternative to this "double whammy" is a separate sanction P when $b = 0$, independently of the choice of a. This separation between the economic and image sanction is difficult to implement, though: As I already noted, agents have an incentive to disclose the punishment P and thereby prove that $b = 0$ to create an excuse for having selected $a = 0$ (or raise the glory for having selected $a = 1$). So, there is unravelling, and the two dimensions of behavior are necessarily intertwined. I consider here the double whammy designed in current pilots.