Self-Enforcing Cooperation with Graduated Punishments*

Dilip Abreu, Princeton University
B. Douglas Bernheim, Stanford University
Avinash Dixit, Princeton University

Preliminary version, July 1, 2005

Abstract

Case studies of self-enforcing cooperation in repeated interactions usually find that the punishments inflicted on deviators are mild to start with, and increase only if there is evidence of persistent deviation. We model this using a combination of imperfect monitoring and asymmetric information in a one-sided prisoner's dilemma. Call the player with the temptation to cheat player B, and the other player A. Player B can be one of several types, distinguished by the size of this temptation; B's type is his private information. A's payoff may be high or low, and the probability of a low payoff is higher if B has cheated. If player A receives the low payoff, he can inflict an immediate punishment on B, at a cost to himself. We find that A's optimal choice of punishment is generally a non-concave problem, and the comparative statics can be discontinuous and asymmetric. If the temptation of each type of B is non-stochastic, A's optimal choice of punishment is not lowered after a high payoff, and may be raised or lowered after a bad payoff depending on the prior distribution of B's types. With stochastic temptations, as the prior probability of B being a bad type increases, the optimal level of punishment chosen by A is initially low and decreasing, but then jumps to a higher level and increases further thereafter, and for very high levels of this probability the relationship may be terminated. Non-extreme levels of punishment are most informative about B's type, and the value of this information to A is highest when the prior probabilities are also non-extreme.

*This is an extremely preliminary statement of work still in progress. It was prepared by Dixit for an informal seminar presentation to elicit feedback, and has not yet been read or approved by the coauthors. Comments very welcome, addressed to `dixitak@princeton.edu` .

1 Motivation

Over the past several decades, a large literature has emerged concerning cooperation and enforcement of implicit agreements in ongoing relationships. This includes theoretical modeling as well as case studies and empirical research. The theory and the practice of sustaining cooperation in repeated interaction agree on most points, for example the effects of patience, the number of players, the importance of information, and so on. However, a large gap remains when it comes to the harshness of punishments.

The case study literature (for example Ostrom 1990, pp. 97–99; Ellickson 1991, pp. 53–64) emphasizes that many successful cooperative arrangements involve graduated punishments. The immediate response to a deviation is usually a mild step such as a quiet complaint and a request to make restitution. Only if this fails are successively more serious steps, such as retaliation, employed. Recourse to courts, or breaking off of the relationship, are usually the last steps. Casual observation suggests that, to some extent, this process operates in both directions, in the sense that lighter punishments are sometimes applied after good behavior, but there appears to be less sensitivity in this direction. Instances where bad behavior ratchets up punishments quickly are numerous, while examples where good behavior ratchets punishments down are hard to find.

In dynamic models with perfect monitoring the harshest punishment may be invoked to deal with *all* deviations, repeated and otherwise. Moreover, deterrence is perfect and players do not cheat in equilibrium. The theory of such games is exposited in Fudenberg and Tirole (1991, chapter 5). In models with imperfect monitoring, e.g. Green and Porter (1984), Abreu, Pearce, and Stacchetti (1990), one can get moderate rather than extreme punishments because there is a positive probability that costly punishments will have to be carried out on the equilibrium path even though no one is cheating. However, in these models there is no natural tendency for punishments to increase in response to repeated "bad" signals. For instance in the symmetric environment investigated in Abreu, Pearce, and Stachetti (1986) the optimal equilibrium entails only two states, a collusive state and a punishment state, with transitions between these two states following a stationary Markovian rule.

In searching for theoretical explanations of graduated punishments one is led naturally to consider models with asymmetric information. However, with perfect monitoring it is still possible to achieve perfect deterrence with sufficiently harsh punishments that are costless in that they do not arise in equilibrium.[1] Thus we are led to models with both asymmetric information and imperfect monitoring. To our knowledge, the literature contains only a few examples of models with this combination of elements, and that only at the abstract level of folk theorems for highly patient players; e.g. Fudenberg, Levine and Maskin (1994, section 8). Here we seek to improve our understanding of more concrete applications in more

[1] In models with asymmetric information and perfect monitoring, for example Athey and Bagwell (2001, 2004), Athey, Bagwell and Sanchirico (2002), there are equilibria in which actions interpretable as bad behavior (increased market share) are followed by punishment (equilibrium price war), but these appear to be less efficient than equilibria with rigid equilibrium behavior, and in any case analysis of these models has, as yet, shed no light on the treatment of repeat offenders.

realistic situations with significant or even extreme impatience.

Our analysis thus far has covered only a few simple examples and models, but it has already improved the fit between theoretical and empirical work, and has yielded some surprising and seemingly counterintuitive results. As an example of improved fit, our results support the observed asymmetry between ratcheting up and ratcheting down. Another such result is our finding that termination of relationship is indeed best used as a last resort, not as the immediate threat of the trigger strategy kind. More surprisingly, we find a robust set of circumstances where bad experiences lead to milder, not harsher, punishments. We are currently building on and generalizing these preliminary findings.

2 The basic idea

Consider an ongoing relationship where: (1) players have private information concerning their incentives to act selfishly, and these incentives are reasonably stable over time, (2) selfish behavior is never jointly efficient, (3) choices are not directly observable even with a time lag, (4) choices probabilistically affect observable outcomes, and (5) the scope for (credible) punishment is large. By setting a sufficiently large punishment, it is possible to discourage a player from cheating on an agreement, irrespective of his private information. However, since it is necessary to condition punishments on bad outcomes rather than on opportunistic choices, there is some likelihood (as in the Green-Porter model) that the punishment will nevertheless be triggered; if inflicting the punishment is sufficiently costly, this may make the equilibrium unattractive.

Consider using a milder punishment. This change has four effects on payoffs. (1) Less cooperation. A player deviates when he has a high incentive to act opportunistically. (2) More frequent punishment. With less cooperative behavior, bad outcomes are more likely, so punishments are triggered more often. (3) Milder punishments. When punishment occurs, the players don't sacrifice as much utility. (4) Learning. If punishment is insufficient to deter cheating in all cases, then learning occurs, allowing players to use more appropriate punishments in subsequent periods.

Equilibrium payoffs decline as a result of the first two effects and rise as a consequence of the last two. When the likelihood of having a high incentive to act opportunistically is sufficiently small, the first two effects become relatively unimportant, while the third remains significant (recall that the punishment is sometimes triggered regardless of private information and behavior). In that case, it is advantageous to set the punishment at a level insufficient to deter all opportunistic behavior. Players with the greatest incentives to act opportunistically do so in equilibrium, raising the probability of a bad outcome that triggers punishment. Thus punishments sometimes occurs because people act badly, and not merely because the state of the world is unfavorable. When players enter a punishment phase, they know they are punishing opportunistic behavior, rather than nature, with some probability.

In this setting, experience also reveals information about one's fellow player. If the outcome is good, it is more likely that the other player has a relatively small incentive to act opportunistically; if the outcome is bad, this is less likely. As the probability distribution of the other player's type changes, so does the balance between the effects discussed in the last two paragraphs, and this means the optimal punishment also changes. If the optimal punishment becomes more severe as probability shifts to types with greater incentives to act opportunistically, then a bad outcome leads to the use of a more severe punishment in the next round, and a good outcome has the opposite effect; cheaters are punished in equilibrium, repeat offenders are punished more severely, and those with good records are treated with leniency. However, the optimization problem for the punishment is generally non-concave and has complex and discontinuous comparative statics. Therefore we find that there are robust circumstances in which optimal punishments either do not change, or become less severe as probability shifts to more opportunistic types, and this leads to some interesting insights (e.g. concerning the relative insensitivity of punishments to good experience, as discussed below).

Note that both asymmetric information and imperfect monitoring are critical for this argument. Without asymmetric information concerning incentives to act opportunistically, we would be in the Green-Porter world; the optimal punishment would assure the absence of cheating, and would remain unchanged over time, regardless of bad or good experience. Without imperfect monitoring, it would be optimal to use a punishment level that would deter cheating irrespective of private information. It would never be used in equilibrium, and there would be no reason to change it with experience.[2]

3 The model

We consider a one-sided prisoner's dilemma game in which player A decides whether to initiate each stage game. If he does so, then player B chooses one of two actions, which we call "generous" and "selfish". These result in different probabilities of A getting one of two payoffs, "high" and "low". The low payoff to A is more likely if B takes the selfish action than if he takes the generous action, but both probabilities are strictly positive and strictly less than 1 so B's action cannot be inferred for sure from A's payoffs. We define B's "temptation" as his expected excess payoff in a stage game if he takes the selfish as opposed to the generous action. Player B can be one of several "types". The type is invariant across periods. The temptation depends on the type; in one variant of our model the temptation is a random variable whose distribution depends on the type but is IID across periods. B's type as well as the actual temptation in any period are B's private information; A can only ever observe his own payoffs. If A gets the low payoff, he can inflict an immediate punishment in the stage game on B, with an attendant cost to himself. Such punishment within the stage game is a realistic feature in many case studies, so we assume it.

In fact, for most of this paper, we consider just a one-shot game where A can commit to inflicting this punishment in the stage game if he received the low payoff. As usual,

[2] The matter becomes a bit more complicated if, contrary to our fifth assumption, arbitrarily severe credible punishments are not available, since then it may not be possible to keep players in line irrespective of private information. It is also more complicated when, contrary to our second assumption, selfish actions are sometimes jointly efficient, as in Athey and Bagwell (2001, 2004) and Athey, Bagwell, and Sanchirco (2002).

with infinite repetition and sufficient patience, this commitment can become endogenously credible in the sense of subgame perfection. Since we have nothing new to add to the usual theory in this respect, we suppress the explicit analysis of subgame perfection with repetition. This enables us to focus attention on the new aspect of what changes from one period to the next, namely A's beliefs about the type of the B player.

Now we describe the stage game and set up the notation.

1. Player B comes in types characterized by a parameter θ. A's prior beliefs about the probabilities of B being of these different types are common knowledge.

2. Player A decides whether to initiate the interaction with B. If he does not, then he gets an outside payoff a. If he chooses to initiate the interaction, he does it by committing himself to a punishment rule and chooses its level p. The punishment will go into effect if A gets zero payoff in the game that follows. If the punishment is invoked, it will reduce B's payoff by p and A's by $m\,p$, where $m > 0$.

3. If A has initiated the game, B gets the choice between two actions, "generous" and "selfish". B's expected payoff from the selfish action exceeds that from the generous action by the amount y; call this B's temptation level. We consider two cases, labeled fixed and stochastic temptation.

3 f. In the case of fixed temptation, we take $y \equiv \theta$ and let $F(\theta)$ denote A's prior cumulative distribution function of B's types.

3 s. In the case of stochastic temptation, we assume that the cumulative distribution function of y is contingent on B's type θ, and denoted by $F(y, \theta)$ over a type-invariant support $[0, \overline{y}]$. The density is $F_y > 0$. Also, $F_\theta < 0$ for all $y \in (0, \overline{y})$, so the larger is θ, the further to the right is the distribution of y in the sense of first-order stochastic dominance. So larger θ means worse types. (Of course $F(0, \theta) = 0$ and $F(\overline{y}, \theta) = 1$ for all θ.)

4. A does not observe the actions, but only the resulting payoff, which can be either "high" c or "low" 0, where $c > a > 0$. The probabilities of the two outcomes are as follows:

$$\begin{aligned} \mu &= \text{Prob[A gets low payoff} \mid \text{B takes selfish action]}, \\ \lambda &= \text{Prob[A gets low payoff} \mid \text{B takes generous action]}, \end{aligned} \tag{1}$$

with $1 > \mu > \lambda > 0$.

5. We simplify the analysis still further by assuming throughout that player B is infinitely impatient, so in each stage game he is concerned with his payoff solely in that stage game. Then there is a very simple relationship between A's choice of punishment p and the highest level of B's temptation which it would deter. B will take the generous action if the expected payoff gain from taking the selfish action, namely the temptation level y, is less than the expected payoff loss due to the higher probability of invoking the punishment, namely $(\mu - \lambda)\,p$. So the highest temptation deterred and the punishment are linked by $y = (\mu - \lambda)\,p$. We can then regard A's choice of the punishment p as equivalent to the choice of the highest level of temptation y he chooses to deter, and the cost to A of choosing y is $m\,y/(\mu - \lambda)$, which we abbreviate as $\gamma\,y$. An increase in the precision with which A's payoff reveals B's action corresponds to an increase in μ and/or a decrease in λ. Both of these decrease γ for given m. Therefore we can interpret the new parameter γ as A's effective unit cost of deterrence.

6. Moreover, in the next two sections, we assume that player A is also infinitely impatient, concerned only with the current period's payoff in each period. Then the only link across periods is that of the revision of A's beliefs about B's type, which is our main new concern. In Section 6 we consider the additional features introduced when A is finitely impatient.

4 Fixed Temptation Case

If A has set the punishment at a level to deter temptations up to y, then

$$\text{Prob[B chooses generous]} = \text{Prob}[\theta \le y] = F(y)\,.$$

Define

$$L(y) = \text{Prob[A gets low payoff]} = \lambda\; F(y) + \mu\; [1 - F(y)]\,.$$

A chooses y to maximize his expected payoff,

$$\begin{aligned} V(y) &= c\;[1 - L(y)] + (-\gamma\, y)\; L(y) \\ &= c - (c + \gamma\, y)\;\{\;\lambda\; F(y) + \mu\;[1 - F(y)]\;\} \end{aligned} \qquad (2)$$

In the light of his payoff realization this period, A will revise his prior and set the optimal y next period. The intuition based on the case studies would be that y is revised upward after a bad experience, and may be revised downward after a good experience. The formal result is different in both respects. A good experience does not change the optimal y at all; after a bad experience the optimal y may be revised upward or downward.

Here is the intuition for this apparently counterintuitive result; the formal proof is in Appendix A.

Suppose the punishment for the current period is set at y^*. If A's outcome is good, his posterior distribution changes as follows: the likelihood of all types with temptation $y < y^*$ shifts up by the same proportion, and the likelihood of all types with temptation $y > y^*$ shifts down by another constant proportion.

Consider changing the punishment from some $y' < y^*$ to y^*. There are three effects: (1) higher cost of punishments imposed on those with $y < y'$ in the event of "accidental" bad outcomes; (2) deterrence of those with $y' < y < y^*$, along with higher cost of punishment in the event of "accidental" bad outcomes; (3) higher cost of certain punishment imposed on those with $y > y^*$. (1) and (3) are costs, but (2) can be beneficial. If y^* is optimal initially then, for the prior distribution, (2) is a net benefit and is weakly greater (in absolute value) than the sum of (1) and (3). Now switch to the posterior distribution. Effects (1) and (2) increase proportionately, and (3) decreases, so the change is strictly beneficial.

Consider changing the punishment from y^* to some $y'' > y^*$. There are again three effects: (1) higher cost of punishments imposed on those with $y < y^*$ in the event of "accidental" bad outcomes, (2) deterrence of those with $y^* < y < y''$, along with higher cost of punishment in the "accidental" bad outcomes, and (3) higher cost of certain punishment imposed on those with $y > y''$. (1) and (3) are again costs, but (2) can be beneficial. If y^* is optimal initially then, for the prior distribution, (2) does not outweigh the sum of (1) and (3). Now switch to the posterior distribution. Effects (2) and (3) decrease proportionately, and (1) increases, so the change is strictly detrimental.

Reversing this argument, one can see that bad news increases the desirability of both increasing and decreasing the punishment, which is why it can cause punishment to rise or fall. We now present a numerical example showing how the seemingly perverse pattern is possible for non-exceptional values of the parameters.

Example where punishment goes down after a bad experience:

The parameters are $c = 12$, $\gamma = 1$, $\mu = 1$, and $\lambda = 0.5$. The types (equal to temptations) of the B player are $\theta = 1$, 6, 12, and A's prior probabilities on these types are respectively 0.25, 0.50, 0.25. Now A need consider only three discrete deterrence levels corresponding to B's temptation levels. Table 1 shows the consequences. The optimum is $y = 6$, which deters the two lower types of B but lets the worst type behave selfishly.

Table 1: A's First-Period Decision

A's choices of y	1	6	12
$1 - L$ = Prob[high payoff to A]	0.125	0.375	0.500
L = Prob[low payoff to A]	0.875	0.625	0.500
$V = c - (c + \gamma y) L$	0.625	0.750	0.000

Suppose A makes this choice and experiences the low payoff. Ex ante, this would happen with probabilities $0.5 \times 0.25 = 0.125$, $0.5 \times 0.50 = 0.25$, and $1 \times 0.25 = 0.25$ respectively from the three types of B. The total probability of A getting the low payoff is then $0.125 + 0.25 + 0.25 = 0.625$. Therefore A's posterior probabilities on B's type are respectively 0.125/0.625 = 0.2, 0.250/0.625 = 0.4, 0.250/0.625 = 0.4

Table 2 shows A's decision for the second period. The optimum is $y = 1$, which deters only the lowest type of B. In fact $y = 6$, which was the first-period optumum, has become the worst choice.

Table 2: A's Second-Period Decision

A's choices of y	1	6	12
$1 - L$ = Prob[high payoff to A]	0.100	0.300	0.500
L = Prob[low payoff to A]	0.900	0.700	0.500
$V = c - (c + \gamma y) L$	0.300	−0.600	0.000

What has happened is that the experience of low payoff has made A think that the worst type of B is now more likely than was the case before, and deterring that type is too costly.

Therefore A finds it better to ease up on the deterrence. If A's outside payoff a is less than 0.300, he finds it optimal to continue the relationship with this milder punishment; but if $a > 0.300$ he will terminate the relationship and take the outside opportunity.

5 Stochastic Temptation Case

First we set up some notation and obtain some useful properties of the underlying functions for a given type θ of the player B. The CDF of the temptation is $F(y, \theta)$. The probability that A gets the low payoff when he has chosen the deterrence level y and B's type is θ is denoted by $B(y, \theta)$; thus

$$B(y, \theta) = \lambda\, F(y, \theta) + \mu\, [1 - F(y, \theta)]\,. \tag{3}$$

Then

$$\begin{aligned} B_y &= -(\mu - \lambda)\; F_y < 0 \qquad (4)\\ B_\theta &= -(\mu - \lambda)\; F_\theta > 0\,. \qquad (5) \end{aligned}$$

Also

$$B(0, \theta) = \mu > B(\overline{y}, \theta) = \lambda \qquad \text{for all } \theta\,. \tag{6}$$

Figure 1 shows the typical shape of B as a function of y for two different values of θ, with $\theta_1 < \theta_2$. Since $B_\theta > 0$, the $B(y, \theta_2)$ function lies entirely above the $B(y, \theta_1)$ function, except that the two coincide at the end-points.[3]

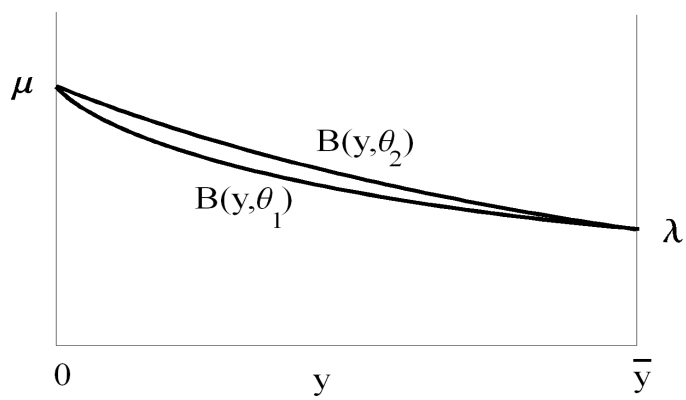

Figure 1: Probability of A receiving low payoff

Now suppose the θ_1 and θ_2 introduced above are actually the two possible types of the player B, with π being the probability of the "bad" type θ_2. Denote by $L(y, \pi)$ is the probability that A gets a bad payoff when he has chosen the deterrence level y, so

$$L(y, \pi) = (1 - \pi)\; B(y, \theta_1) + \pi\; B(y, \theta_2) \tag{7}$$

[3]The curvature shown is for the case where the density F_y is decreasing in y which is broadly what is needed for the local second-order conditions below.

Then

$$
\begin{aligned}
L_y &= (1-\pi)\ B_y(y,\theta_1) + \pi\ B_y(y,\theta_2) < 0 && (8)\\
L_\pi &= B(y,\theta_2) - B(y,\theta_1) > 0. && (9)
\end{aligned}
$$

Remember that we are assuming both players to be infinitely impatient. However, we are also assuming that both players have perfect memories. Therefore the only link across periods is the updating of A's probabilities of the various types for B. In this section we consider the case of two types.

A's objective function in the current period is his expected payoff

$$
\begin{aligned}
V(y,\theta) &= c\,[1 - L(y,\pi)] - \gamma\, y\ L(y,\pi)\\
&= c - L(y,\pi)\ (c + \gamma\, y)\,. && (10)
\end{aligned}
$$

The first-order condition for a local interior maximum is

$$V_y = -L_y\ (c+\gamma\, y) - L\,\gamma = 0\,. \tag{11}$$

The issues of the local second-order condition and end-point optima are discussed in the Appendix. The major issue that arises is the possibility of multiple local optima. But first we set up some local comparative statics.

Let $y(\pi)$ denote a local maximizer y regarded as a function of the probability π of the bad type. We want to know when this is an increasing function. To do this comparative statics, differentiate the first-order condition totally:

$$V_{yy}\ \frac{dy}{d\pi} + V_{y\pi} = 0\,,$$

where the partial derivatives of V are evaluated at the optimum. Using the second-order condition, then

$$\text{Sign}\left(\frac{dy}{d\pi}\right) = \text{Sign}(V_{y\pi})\,.$$

The details of the algebra that yields an interpretation of this sign are in Appendix C. The key concept is the local sensitivity of the different types of B to the deterrence level. Recall that $B(y,\theta)$ is the probability that player A gets the bad outcome when he chooses to deter temptations up to y, and player B is of type θ. As y increases, B decreases, and the proportional rate of decrease is $-B_y/B$ (remember that B_y is negative). Call this the "marginal sensitivity to deterrence" of a type-θ player B. Define

$$S(y,\theta) = -B_y(y,\theta)\,/\,B(y,\theta)\,. \tag{12}$$

Then the algebra yields

$$\text{Sign}\left[\frac{dy}{d\pi}\right] = \text{Sign}[\,S(y,\theta_2) - S(y,\theta_1)\,] \tag{13}$$

To help fix when the right hand sign is positive and when it is negative, look at Figure 1. The $B(y, \theta_2)$ function lies entirely above the $B(y, \theta_1)$ function, except that the two coincide at the end-points. Therefore the slopes at the end-points must satisfy

$$-B_y(0, \theta_1) > -B_y(0, \theta_2), \qquad -B_y(\overline{y}, \theta_1) < -B_y(\overline{y}, \theta_2)\,.$$

Dividing each inequality by the common value of the functions at that point,

$$S(0, \theta_1) > S(0, \theta_2), \qquad S(\overline{y}, \theta_1) < S(\overline{y}, \theta_2)\,.$$

By continuity, S will be increasing in θ for y sufficiently close to $\overline{y}$ and S will be a decreasing function of θ for y sufficiently close to zero. Thus the criterion in (13) simply *cannot* give a globally unambiguous sign to $dy/d\pi$.

Consider the simplest possibility compatible with this, namely a "single crossing" when there is be a $\widehat{y}(\theta_1, \theta_2)$ such that

$$S(y, \theta_2) - S(y, \theta_1) \begin{cases} < 0 & \text{for } 0 < y < \widehat{y}(\theta_1, \theta_2) \\ > 0 & \text{for } \widehat{y}(\theta_1, \theta_2) < y < \overline{y} \end{cases}$$

Therefore the comparative static result is

$$\frac{dy}{d\pi} \begin{cases} < 0 & \text{for } 0 < y < \widehat{y}(\theta_1, \theta_2) \\ > 0 & \text{for } \widehat{y}(\theta_1, \theta_2) < y < \overline{y} \end{cases}$$

Of course $\widehat{y}(\theta_1, \theta_2)$ is itself endogenous and depends on all the parameters of the problem, but both cases can arise in non-exceptional circumstances. To see this, first observe that the behavior of S depends only on the form of the distribution function F and the two parameters μ and λ, so these things and the two $\theta's$ fix the $\widehat{y}$. In particular, $\widehat{y}$ is independent of π. Next, the optimum depends on two other parameters, c and m (with $\gamma = m/(\mu - \lambda)$ which can be varied from 0 to ∞ by varying m). For small c and/or large m, the optimum y is small (and may even be at a corner solution $y = 0$), whereas for large c and/or small m, the optimum y is large (and may even be at a corner solution $y = \overline{y}$.) This depending on these two parameters, the optimum can be less than or greater than the critical $\widehat{y}(\theta)$ above which our condition is satisfied.

Figure 2 illustrates the possibilities. If the optimal punishment is less than $\widehat{y}(\theta_1, \theta_2)$, an increase in π will make the optimal punishment even smaller, keeping it in the range below $\widehat{y}(\theta_1, \theta_2)$ and therefore decreasing further as π increases. The optimal y may even hit zero for some $\pi < 1$ and then will stay at zero for any higher values of π. If the consequences of B's cheating are not too large or punishing B is too costly to A, then A gradually gives up when things get even worse.[4] Conversely, if the optimal punishment is above $\widehat{y}(\theta_1, \theta_2)$, then an increase in π makes the optimal punishment size even higher. It may hit $\overline{y}$ for some $\pi < 1$ and then stay there. There seems to be an interesting bifurcation in the choice of punishments.

[4]This is similar to the apparently counterintuitive finding in the fixed temptation case.

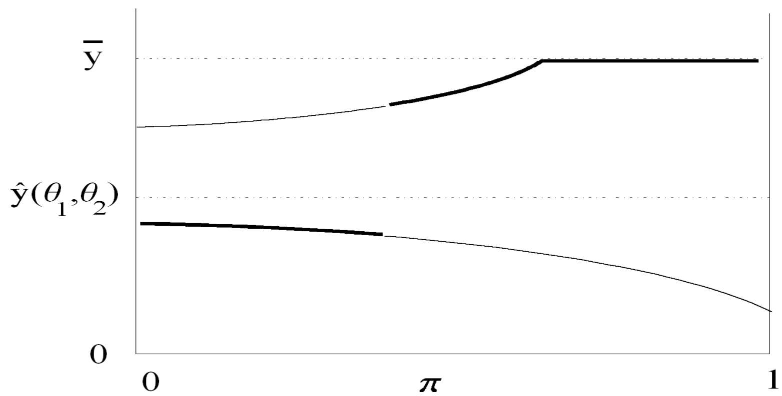

Figure 2: Bifurcation of Optimal Punishments and Local Optima

The bifurcation goes naturally with the possibility of multiple local optima. Consider the special situation where the distribution of temptation for player B of type i is highly concentrated around y_i, with $y_1 < y_2$. Start with a low probability π of the bad type 2. If A wanted to deter the bad type, he would have to set y above y_2. But with low π, the punishment would be invoked most of the time when B was of type 1 and behaved generously but the outcome turned out to be low. This would be very costly. Therefore it is not optimal for A to try to deter the few bad types; y^* is close to y_1. If π increases a little while staying in this range, the punishment still does not deter type 2's, but its cost must be borne more often. This is not optimal; therefore the optimal punishment goes down. But if π rises sufficiently, it becomes desirable to deter type 2's by ratcheting up the punishment discontinuously, and thereafter keep on raising it. In Figure 2, this is shown by the two thicker portions of the bifurcating curves.

If the distribution of B's temptation is not so concentrated, for example for the negative exponential case, $V(y)$ can have a unique regular interior maximum global maximum. It will be located along just one of the two branches shown in Figure 2, and move continuously along this branch as π changes; which branch depends on the other parameters.

The possible irregularities of the optima notwithstanding, A's expected payoff monotonically decreases as the probability of the bad type increases, regardless of whether the punishment increases or decreases. To see this, use the envelope theorem. The optimized value V^* changes with θ according to

$$\frac{dV^*}{d\pi} = \frac{\partial V}{\partial \pi} = -L_\pi(y, \theta)\ (c + \gamma\, y)$$

evaluated at the optimum $y(\theta)$. We know from (9) that $L_\pi > 0$. Therefore $dV^*/d\pi$ is negative, so V^* decreases as π increases. It is possible that V^* drops to A's outside opportunity level a for some π. Then for any larger π, A will not initiate the interaction at all.

All of this was done as comparative statics in a stage game, but it can immediately be interpreted in the context of A's Bayesian updating of the probability of B's types from one period to the next. If in any period A receives the low (zero) payoff, he revises the probability of B being the bad type to

$$\Pi(0) = \frac{\pi \ B(y,\theta_2)}{\pi \ B(y,\theta_2) + (1-\pi) \ B(y,\theta_1)} > \pi \,. \tag{14}$$

If instead A receives the high payoff c, he revises the probability of B being the bad type to

$$\Pi(c) = \frac{\pi \ [1 - B(y,\theta_2)]}{\pi \ [1 - B(y,\theta_2)] + (1-\pi) \ [1 - B(y,\theta_1)]} < \pi \,. \tag{15}$$

Therefore, if in the first period the parameter values are such that the optimal y is above the threshold $\widehat{y}$, then bad experiences lead A to revise π upward and use higher punishments, perhaps culminating in a breakdown of the relationship, whereas good experiences lead A to revise π downward and use smaller punishments. However, if in the first period the parameter values are such that the optimal y is less than the threshold $\widehat{y}$, then bad experiences lead A to revise π upward and use smaller punishments, whereas good experiences lead A to revise π downward and use larger punishments. It is still true that bad experiences reduce A's expected payoff, perhaps leading to an eventual termination of the relationship.

Example of Multiple Local Optima

In conformity with the intuition explained above, this example assumes that the distribution of temptation y conditional on θ is normal with a somewhat small standard deviation. Specifically, the two types have normally distributed temptations, with means 4 for type 1 and 6 for type 2, and the same standard deviations 0.14 for each. The other parameters are $c = 10$, $m = 0.5$, $\mu = 1$, $\lambda = 0.5$, so $\gamma = 1$.

Table 3: Local Maxima and Discontinuous Jumps

π	Lower y^*	$V(y^*)$	Upper y^*	$V(y^*)$
0.050	4.380	2.426		
0.100	4.374	2.068		
0.130	4.371	1.853	(6.267)	(1.836)
0.135	(4.370)	(1.818)	6.270	1.834
0.150			6.278	1.831
0.200			6.298	1.822

Table 3 shows the global maximizers y^* and maximized values $V(y^*)$, and also shows one further local maximizer that is not global on each side of the switchover to clarify the comparisons. Figure 3 shows the graph of $V(y)$ for the π value when the two local optima are almost equal.

From the table we see in the numerical solutions the features that arose as possibilities in the previous discussion of comparative statics and the schematic Figure 2, namely discontinuity and bifurcation. The "Lower y^*" column corresponds to the lower curve in Figure 2;

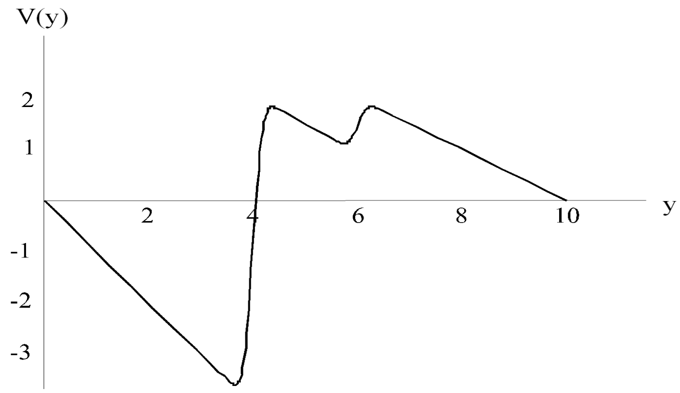

Figure 3: Multiple Local Maxima

it is decreasing as π increases. Between $\pi = 0.130$ and 0.135, the global optimum switches to "Upper y^*" column, and thereafter it increases as π increases.

Some other noteworthy features are: [1] Both optima are nearly two to 2.5 standard deviations above the corresponding mean: for each type, deterrence is almost complete when it is attempted at all. [2] A's optimized payoff $V(y^*)$ is a decreasing convex function of π. This can be proved rigorously and with much greater generality; see Propositions 1 and 2 in Appendix F. [3] Therefore the outside opportunity is taken when π goes above a critical value: breakdown of relationship is last resort, not the first threat as in a trigger strategy.

Appendixes D and E develop the theory for the case with many types of B; we omit it in the text.

6 Repeated Game with Finitely Impatient A

Now suppose A has finite impatience with discount factor δ, but B is infinitely impatient. Let B be of two possible types as before. The Bellman function for A's recursive dynamic programming problem can be written as

$$W(\pi) = \max_y \left\{ V(y, \pi) + \delta \ \{ L(y, \pi) \ W(\Pi(0)) + [1 - L(y, \pi)] \ W(\Pi(c)) \ \} \right\}, \qquad (16)$$

where $\Pi(0)$ and $\Pi(c)$ denote the vectors of the posterior probabilities of types after getting payoffs zero and c respectively, with the components defined in (14) and (15) respectively. Note that both $\Pi(0)$ and $\Pi(c)$ depend on the vector of prior probabilities π and the choice of the maximum temptation to deter y.

The contraction mapping procedure that establishes the basic properties of the solution are standard but tedious, and are in Appendix F. Here we consider only the comparative statics of a local optimum, and the new considerations that arise because of A's concern for the future.

When the Bellman value function W is the appropriate fixed point W^*, write the function that is maximized by choice of y as

$$SW^*(y, \pi) = V(y, \pi) + \delta \ \{ \ L(y, \pi) \ W^*(\Pi(0)) + [1 - L(y, \pi)] \ W^*(\Pi(c)) \ \} \,. \qquad (17)$$

Recognize that

$$L(y,\pi)\ W^*(\Pi(0)) + [1 - L(y,\pi)]\ W^*(\Pi(c))$$

is just A's expected continuation payoff; write it as $EW^*(y,\pi)$. Using this abbreviation, but expanding out the V function in SW^*, we have

$$SW^*(y,\pi) = c - L(y,\pi)\ (c + \gamma\, y) + \delta\ EW^*(y,\pi)\,.$$

The first-order condition for maximization is

$$SW^*_y = -\gamma\ L - (c + \gamma\, y)\ L_y + \delta\ EW^*_y = 0\,.$$

As usual, the sign of $dy/d\pi$ is the same as the sign of $SW^*_{y\pi} > 0$ at the optimal point.

$$SW^*_{y\pi} = -\gamma\ L_\pi - (c + \gamma\, y)\ L_{y\pi} + \delta\ EW^*_{y\pi}\,.$$

Multiplying the equation for $SW^*_{y\pi}$ by L, that for SW^*_y by $-L_\pi$, and adding the two, we get

$$L\ SW^*_{y\pi} - L_\pi\ SW^*_y = (c + \gamma\, y)\ (L_y\, L_\pi - L\ L_{y\pi}) + \delta\ (L\ EW^*_{y\pi} - L_\pi\ EW^*_y)\,.$$

So evaluated at the optimal y where $SW^*_y = 0$, we have

$$L\ SW^*_{y\pi} = (c + \gamma\, y)\ (L_y\, L_\pi - L\ L_{y\pi}) + \delta\ (L\ EW^*_{y\pi} - L_\pi\ EW^*_y)\,. \tag{18}$$

Thus the overall effect we are looking for is a combination of two terms. The first, $(L_y\, L_\pi - L\ L_{y\pi})$, appeared in the static problem with infinite impatience. The second, $(L\ EW^*_{y\pi} - L_\pi\ EW^*_y)$, is essentially dynamic, and arises when A is finitely patient and concerned with the continuation payoff. The relative weights on the two depend on δ. When δ goes to zero we get the static case; when δ goes to 1, only the dynamic term matters because EW^* and its derivatives contain a factor $1/(1-\delta)$. To consider all cases of δ, the signs of the two must be examined separately. The static term was extensively discussed in the previous sections, so let us consider the dynamic term here.

Appendix F contains the argument showing that the dynamic aspect contributes a positive term to $dy/d\pi$ in two situations, (1) when π is low and y is also low, and (2) when π is high and y is also high. In the other two situations, (3) low π, high y, and (4) high π, low y, the dynamic aspect contributes a negative term to $dy/d\pi$. Recall that parameters c and m affect y separately from π and can trace the whole range $[0, \overline{y}]$, so all four cases are logical possibilities. And note that the "high" and "low" in the above statements are rough and not expressed in terms of precise cutoffs; that would not be possible because for example the value of π for which the EW^*y/G curve attains its maximum or minimum depends on the exact value of y.

The taxonomy may look complicated, but it can be expressed more simply and interpreted. The general idea is: if π moves away from its extreme values toward an intermediate range, so does y. In cases (1) and (3), when π increases starting from near zero, y increases starting from a low level in case (1) and decreases from a high level in case (3). In cases (2) and (4), if π decreases starting from near 1, then y decreases starting from a high level in

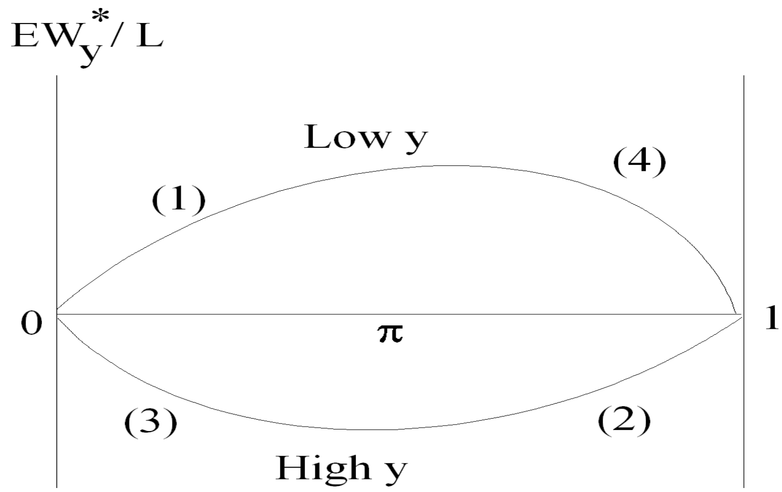

Figure 4: Behavior of Dynamic Effect on Punishments

case (2) and increases starting from a low level in case (4). The idea that moving π toward an intermediate value also moves y toward an intermediate value can be interpreted in terms of the information value of punishments.

Write the Bayes' Rule formula (14) for the posterior probability of B being the bad type after A experiences a low payoff as

$$\Pi(0) = \frac{\pi}{(1-\pi)\ B(y,\theta_1)/B(y,\theta_2) + \pi}$$

This regarded as a function of y for fixed π equals π when $y = 0$ (where both the B functions equal μ) and also when $y = \overline{y}$ (when both the B functions equal λ). For all intermediate y, we have $B(y,\theta_1)/B(y,\theta_2) < 1$ so $\Pi(0) > \pi$. Then $\Pi(0)$ will be maximum at an intermediate y, call it y_0. Intuitively, both the extreme punishments – zero (which deters neither type for any temptation) and $\overline{y}$ (which deters both types for all temptations) are uninformative. Information is maximized for an intermediate punishment.

The value of information also depends on π. When $\pi = 0$, both $\Pi(0)$ and $\Pi(c)$ are also zero; when $\pi = 1$, both posteriors are also 1. Therefore no new information accrues when the prior is at an extreme, no matter what punishment is being used. It is only for intermediate priors that significant information can be obtained by using punishments.

The dynamic effect is essentially that A will be able to start the next period with updated probabilities, and his current calculation takes into account the discounted continuation value. Therefore when information is more likely to accrue (intermediate π), the dynamic consideration argues in favor of using a more informative punishment (intermediate y). We should again emphasize that the statements are deliberately vague in their use of "high" and "low"; there is no clear cutoff and there may be some ambiguities in an intermediate range, for example because of the two different values of y that maximize information upon A's receiving high versus low payoffs.

Further details of this are discussed in Appendix F. The general ideas concerning the value of information and informativeness of actions conform to the classic analysis of Howard

(1966, 1967).

7 Concluding Remarks

Our analysis this far consists of some simple and special cases, with a mixture of theory and numerical calculations. But it has already produced some surprising results and a rich set of questions for further research. We are currently working on several of these. The most important is the case of multiple periods when both players have finite impatience. Then the simple one-to-one correspondence between choice of punishment p and the maximum temptation deterred y is lost, and the problem gets even more complicated. We are also examining a more general mechanism design approach.

References

Abreu, Dilip. 1986. “Extremal Equilibria of Oligopolistic Supergames.” *Journal of Economic Theory,* 39(1), June, 191–225.

——— , David Pearce, and Ennio Stacchetti. 1986. “Optimal Cartel Equilibria with Imperfect Monitoring.” *Journal of Economic Theory,* 39(1), June, 251–269.

——— , ——— , and ——— . 1990. “Toward a Theory of Discounted Repeated Games with Imperfect Monitoring.” *Econometrica,* 58(5), September, 1041–1063.

Athey, Susan and Kyle Bagwell. 2001. “Optimal Collusion with Private Information.” *RAND Journal of Economics,* 32(3), Autumn, 428–465.

———- , ——— , and Chris Sanchirico. 2004. “Collusion and Price Rigidity.” *Review of Economic Studies,* 71(2), April, 317–349.

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes.* Cambridge, MA: Harvard University Press.

Fudenberg, Drew and Jean Tirole. 1991. *Game Theory.* Cambridge, MA: MIT Press.

Fudenberg, Drew, David Levine, and Eric Maskin. 1994. “The Folk Theorem with Imperfect Public Information.” *Econometrica,* 62(5), September, 997–1039.

Green, Edward J. and Robert H. Porter. 1984. “Noncooperative Collusion under Imperfect Price Information.” *Econometrica,* 52(1), January, 87–100.

Howard, Ronald A. 1966. “Information Value Theory.” *IEEE Transactions on Systems Science and Cybernetics,* Vol. SCC-2, No. 1, 22–26.

Howard, Ronald A. 1967. “Value of Information Lotteries.” *IEEE Transactions on Systems Science and Cybernetics,* Vol. SCC-3, No. 1, 24–60.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, UK and New York, NY: Cambridge University Press.

Appendix: Proofs

A. Fixed Temptation Case

Proposition: y is not revised after A experiences a high payoff; after A experiences a low payoff, y may be revised either upward or downward.

Proof:

Suppose that in the first period the prior CDF of the temptation θ (now same thing as the type) is $F(\theta)$. The optimal y^* maximizes

$$V(y) = c - (c + \gamma y)\ \{\, \lambda\, F(y) + \mu\ [1 - F(y)]\,\}\,,$$

or minimizes

$$M(y) = (c + \gamma y)\ \{\, \lambda\, F(y) + \mu\ [1 - F(y)]\,\}\,.$$

Therefore

$$M(y) - M(y^*) > 0 \quad \text{both when } y < y^* \text{ and when } y > y^*.$$

Now suppose player A experiences a good outcome. The posterior CDF is

$$F^P(\theta) = \begin{cases} \dfrac{(1-\lambda)\ F(\theta)}{(1-\lambda)\ F(y^*) + (1-\mu)\ [1 - F(y^*)]} & \text{if } \theta < y^* \\[2ex] \dfrac{(1-\lambda)\ F(y^*) + (1-\mu)\ [F(\theta) - F(y^*)]}{(1-\lambda)\ F(y^*) + (1-\mu)\ [1 - F(y^*)]} & \text{if } \theta > y^* \end{cases}$$

Observe that in the above expressions, the denominator (abbreviated as Δ), can be written as

$$\Delta = \begin{cases} (1-\lambda) - (\mu - \lambda)\ [1 - F(y^*)] & \text{this form is used below when } y < y^* \\ (1-\mu) + (\mu - \lambda)\ F(y^*) & \text{this form is used below when } y > y^* \end{cases}$$

The next period, the function to be minimized must be written using two separate formulas. If $y < y^*$, we have

$$\begin{aligned} M^P(y) &= (c + \gamma y)\ \{\, \lambda\, F^P(y) + \mu\ [1 - F^P(y)]\,\} \\ &= \Delta^{-1}\ (c + \gamma y)\ \{\ \lambda\, (1-\lambda)\, F(y) + \mu\, [\Delta - (1-\lambda)\, F(y)]\ \} \\ &= \Delta^{-1}\ (c + \gamma y)\ \{\ \lambda\, (1-\lambda)\, F(y) + \mu\, [(1-\lambda)[1 - F(y)] - (\mu - \lambda)[1 - F(y^*)]\,]\ \} \\ &= \Delta^{-1}\ (c + \gamma y)\ \{\ (1-\lambda)\, [\lambda\, F(y) + \mu\, [1 - F(y)]\,] - \mu\, (\mu - \lambda)\, [1 - F(y^*)]\ \} \\ &= \Delta^{-1}\ \{\ (1-\lambda)\ M(y) - \mu\, (\mu - \lambda)\, [1 - F(y^*)]\ (c + \gamma y)\ \} \end{aligned}$$

Then

$$M^P(y) - M^P(y^*) = \Delta^{-1}\ \{\ (1-\lambda)\ [M(y) - M(y^*)] + \mu\, (\mu - \lambda)\, [1 - F(y^*)]\, \gamma\, (y^* - y) > 0$$

And if $y > y^*$, note that

$$1 - F^P(y) = \frac{(1-\mu)\, [1 - F(y)]}{\Delta}\,.$$

Therefore

$$\begin{aligned} M^P(y) &= (c+\gamma y)\ \{\ \lambda\, F^P(y) + \mu\ [1 - F^P(y)]\ \} \\ &= \Delta^{-1}\ (c+\gamma y)\ \{\ \lambda\,[\Delta - (1-\mu)[1-F(y)]\,] + \mu\,(1-\mu)[1-F(y)]\ \} \\ &= \Delta^{-1}\ (c+\gamma y)\ \{\ \lambda\,[\Delta - (1-\mu)] + (1-\mu)\,[\lambda\, F(y) + \mu\,[1-F(y)]\,]\ \} \\ &= \Delta^{-1}\ (c+\gamma y)\ \{\ \lambda\,(\mu-\lambda)\, F(y^*) + (1-\mu)\,[\lambda\, F(y) + \mu\,[1-F(y)]\,]\ \} \\ &= \Delta^{-1}\ \{\ \lambda\,(\mu-\lambda)\, F(y^*)\,(c+\gamma\, y) + (1-\mu)\, M(y)\ \} \end{aligned}$$

Then

$$M^P(y) - M^P(y^*) = \Delta^{-1}\ \{\ \lambda\,(\mu-\lambda)\, F(y^*)\,\gamma\,(y-y^*) + (1-\mu)\,[M(y) - M(y^*)]\ \} > 0$$

Thus for any $y \neq y^*$, we have $M^P(y) > M^P(y^*)$, so y^* is again optimal.

In particular, if the prior $F(y)$ is differentiable so the previous period had a regular (smooth) minimum at $y = y^*$ with $M'(y) = 0$, we now have

$$\begin{aligned} \left.\frac{dM^P(y)}{dy}\right|_{y\uparrow y^*} &= \Delta^{-1}\ \{\ (1-\lambda)\ M'(y^*) - \mu\,(\mu-\lambda)\,[1-F(y^*)]\,\gamma\ \} \\ &= -\ \frac{\mu\,(\mu-\lambda)\,[1-F(y^*)]\,\gamma}{\Delta} < 0\,. \end{aligned}$$

and similarly

$$\left.\frac{dM^P(y)}{dy}\right|_{y\downarrow y^*} = \Delta^{-1}\ \{\ \lambda(\mu-\lambda)\ F(y^*)\,\gamma + (1-\mu)\, M'(y^*)\} > 0$$

Thus y^* becomes an optimum with a kink.

The analysis of what happens after A experiences a low payoff follows the same steps and is left to the reader. The result in the case of a differentiable prior $F(y)$ is as follows. If the previous period had a regular (smooth) minimum at $y = y^*$ with $M'(y) = 0$, we now have

$$\left.\frac{dM^P(y)}{dy}\right|_{y\uparrow y^*} > 0\,, \qquad \left.\frac{dM^P(y)}{dy}\right|_{y\downarrow y^*} < 0$$

Thus y^* becomes a local maximum instead of a minimum of $M^P(y)$, and depending on the global conditions, the new optimal deterrence level y, namely the minimizer of $M^P(y)$, shifts discontinuously either down or up.

B. Stochastic Temptation Case: Preliminaries

B-1. Rules of the Stage Game

For convenience we repeat the description of the stage game and set up the notation.

1. Player B comes in types characterized by a parameter θ.

2. Player A decides whether to initiate the interaction with B. If he does not, then he gets an outside payoff a. If he chooses to initiate the interaction, he does it by committing himself to a punishment rule and chooses its level p. The punishment will go into effect if A gets zero payoff in the game that follows. If the punishment is invoked, it will reduce B's payoff by p and A's by $m\,p$, where $m > 0$.

3. If A has initiated the game, B gets the choice between two actions, "generous" and "selfish". B's expected payoff from the selfish action exceeds that from the generous action by a random y; call this B's temptation level. The support of y is $[0, \overline{y}]$. The cumulative distribution function of y is contingent on B's type θ, and denoted by $F(y, \theta)$. The density is $F_y > 0$. Also, $F_\theta < 0$ for all $y \in (0, \overline{y})$, so the larger is θ, the further to the right is the distribution of y in the sense of first-order stochastic dominance. So larger θ means worse types. (Of course $F(0, \theta) = 0$ and $F(\overline{y}, \theta) = 1$ for all θ.)

4. A does not observe the actions, but only the resulting payoff, which can be either "high" c or "low" 0, where $c > a > 0$. The probabilities of the two outcomes are as follows:

$$\begin{aligned} \mu &= \text{Prob[A gets low payoff} \mid \text{B takes selfish action]}, \\ \lambda &= \text{Prob[A gets low payoff} \mid \text{B takes generous action]}, \end{aligned} \tag{19}$$

with $1 > \mu > \lambda > 0$.

B-2. One-Shot Game with Just One Type of Player B

As a preliminary exercise, suppose B's type is known, so the distribution of B's types is concentrated on just one θ. We ask as a comparative static question whether the punishment level p chosen by A is an increasing function of θ. This should remain an essential property in all the subsequent dynamic modeling that we do.

Equivalently, we can think of the highest temptation level y that A chooses to deter. B will take the generous action if the expected payoff gain from taking the selfish action, namely the temptation level y, is less than the expected payoff loss due to the higher probability of invoking the punishment, namely $(\mu - \lambda)\,p$. So the highest temptation deterred and the punishment are linked by $y = (\mu - \lambda)\,p$.

A chooses y to maximize

$$\begin{aligned} V(y, \theta) &= F(y, \theta)\,[(1 - \lambda)\,c + \lambda\,(-m\,p)] + [1 - F(y, \theta)]\,[(1 - \mu)\,c + \mu\,(-m\,p)] \\ &= c - \{\,\lambda\,F(y, \theta) + \mu\,[1 - F(y, \theta)]\,\}\,(c + mp) \end{aligned}$$

Now define

$$B(y, \theta) = \lambda\,F(y, \theta) + \mu\,[1 - F(y, \theta)]\,. \tag{20}$$

This is just the probability that A gets the low payoff when he has chosen y and B's type is θ. Then

$$B_y = -(\mu - \lambda)\;F_y < 0 \tag{21}$$

$$B_\theta = -(\mu - \lambda)\;F_\theta > 0\,. \tag{22}$$

Also define $\gamma = m/(\mu - \lambda)$ to simplify the notation. Then

$$V(y,\theta) = c - B(y,\theta)\ (c + \gamma\, y)\,. \tag{23}$$

The first-order condition is

$$V_y = -B_y\ (c + \gamma\, y) - B\,\gamma = 0\,, \tag{24}$$

and the second-order condition is

$$V_{yy} = -B_{yy}\ (c + \gamma\, y) - 2\, B_y\, \gamma < 0\,.$$

The second-order condition need only hold at all stationary points defined by the first-order condition. Write that condition as

$$-B_y\ (c + \gamma\, y) = B\,\gamma$$

and each of these is positive. Therefore dividing the terms in the expression for V_{yy} by the appropriate one of these, the second-order condition can be written as

$$\frac{-B_{yy}}{-B_y} - \frac{2\, B_y}{B} < 0\,,$$

or

$$\frac{-B\ B_{yy} + 2\ (B_y)^2}{(-B_y)\ B} < 0\,.$$

But

$$\frac{\partial}{\partial y}\left(\frac{-B_y}{B^2}\right) = \frac{-B^2\ B_{yy} - (-B_y)\ 2\, B\, B_y}{B^4} = \frac{-B\ B_{yy} + 2\ (B_y)^2}{B^3}$$

so the SOC holds if and only if $-B_y/B^2$ decreases as y increases for fixed θ, evaluated at the optimum.

There will be a local optimum at $y = 0$ if $V_y(0,\theta) < 0$. Using $B(0,\theta) = \mu$ and $-B_y(0,\theta) = (\mu - \lambda)\ F_y(0,\theta)$, the condition becomes

$$(\mu - \lambda)\, c\, F_y(0,\theta) - \mu\gamma < 0.$$

Similarly, there is a local optimum at $y = \overline{y}$ if

$$(\mu - \lambda)\, (c + \gamma\, \overline{y})\, F_y(\overline{y},\theta) - \lambda\gamma > 0\,.$$

Only interior optima exist if

$$F_y(0,\theta) > \frac{\mu\gamma}{(\mu - \lambda)\, c} > \frac{\lambda\gamma}{(\mu - \lambda)\, (c + \gamma\, \overline{y})} > F_y(\overline{y},\theta)\,. \tag{25}$$

Roughly speaking, this requires that the density of y at the low end (small temptations) should be sufficiently higher than that at the high end (large temptations). So we should not expect the kind of result we are looking for to hold if large temptations to cheat are more likely. However, more rigorously speaking, we do not need the density F_y to be monotonically decreasing throughout $(0, \overline{y})$.

For comparative statics of the maximizer $y(\theta)$, differentiate the first-order condition totally:

$$V_{yy} \, \frac{dy}{d\theta} + V_{y\theta} = 0 \, ,$$

where the partial derivaties of V are evaluated at the optimum. Using the second-order condition, then

$$\text{Sign} \left(\frac{dy}{d\theta} \right) = \text{Sign}(V_{y\theta}) \, .$$

Now

$$V_{y\theta} = -B_{y\theta} \, (c + \gamma \, y) - B_\theta \, \gamma \, .$$

And the first-order condition gives

$$-B_y \, (c + \gamma \, y) = B \, \gamma > 0 \, .$$

Dividing the two terms on the right-hand side of the expression for $V_{y\theta}$ by the appropriate one of these two expressions, we have

$$\begin{aligned} \text{Sign}(V_{y\theta}) &= \text{Sign} \left[\frac{-B_{y\theta}}{-B_y} - \frac{B_\theta}{B} \right] \\ &= \text{Sign} \left[\frac{B_y \, B_\theta - B \; B_{y\theta}}{(-B_y) \; B} \right] . \end{aligned}$$

The denominator is positive, so the sign of the numerator is the same as the sign of $dy/d\theta$. We want this to be positive.

To interpret this condition, recall that $B(y, \theta)$ is the probability that player A gets the bad outcome when he chooses to deter temptations up to y, and player B is of type θ. As y increases, B decreases, and the proportional rate of decrease is $-B_y/B$ (remember that B_y is negative). Call this player B's "marginal sensitivity to deterrence". Now

$$\begin{aligned} \frac{\partial}{\partial \theta} \, \frac{-B_y}{B} &= \frac{B \; (-B_{y\theta}) - (-B_y) \; B_\theta}{B^2} \\ &= \frac{B_y \, B_\theta - B \; B_{y\theta}}{B^2} \end{aligned}$$

Therefore our condition amounts to

$$\frac{\partial}{\partial \theta} \, \frac{-B_y}{B} > 0 \, . \tag{26}$$

This says that the worse types have higher marginal sensitivity to deterrence. This condition will appear repeatedly, so we will abbreviate

$$S(y, \theta) = -B_y(y, \theta) \, / \, B(y, \theta) \, . \tag{27}$$

Therefore $dy/d\theta > 0$ if and only if S is an increasing function of θ for any given y. So the property we want – higher punishment for worse types – is not universally valid, and does not follow from pure first-order stochastic dominance ($F_\theta < 0$), but needs this additional condition.

In fact the condition can never be fulfilled globally, as was seen in the text with the argument accompanying Figure 1.

Finally, by the envelope theorem, the optimized value V^* changes with θ according to

$$\frac{dV^*}{d\theta} = V_\theta = -B_\theta(y,\theta)\ (c+\gamma\, y)$$

evaluated at the optimum $y(\theta)$. This is negative, so V^* decreases as θ increases, and it is possible that it drops to A's outside opportunity level a for some θ. Then for any larger θ, A will not initiate the interaction at all.

C. Stochastic Temptation Case: Two Types of Player B

Next suppose the distribution of θ is concentrated on two points, θ_1 and θ_2, with $\theta_1 < \theta_2$, and the prior probability of the worse type θ_2 is equal to π. Then we want the punishment level p or the deterrence level y should be an increasing function of π.

Continuing the same notation as above with obvious slight modifications,

$$V(y,\pi) = c - L(y,\pi)\ (c+\gamma y) \tag{28}$$

where $L(y,\pi)$ is the ex ante probability that A gets the bad payoff when he deters temptations up to y:

$$L(y,\pi) = (1-\pi)\ B(y,\theta_1) + \pi\ B(y,\theta_2)\,. \tag{29}$$

Exactly the same steps as in Section B-2 will then yield

$$\text{Sign}\left(\frac{dy}{d\pi}\right) = \text{Sign}(V_{y\pi}) = \text{Sign}(L_y\ L_\pi - L\ L_{y\pi})\,.$$

Differentiating the expression (29) for G, we get

$$\begin{aligned} L_y\ L_\pi - L\ L_{y\pi} \;\; = \;\; & [\,(1-\pi)\ B_y(y,\theta_1) + \pi\ B_y(y,\theta_2)\,]\ [\,B(y,\theta_2) - B(y,\theta_1)\,] \\ & -[\,(1-\pi)\ B(y,\theta_1) + \pi\ B(y,\theta_2)\,]\ [\,B_y(y,\theta_2) - B_y(y,\theta_1)\,]\,. \end{aligned}$$

Using the definition (27) of the function S, this simplifies to

$$L_y\ L_\pi - L\ L_{y\pi} = B(y,\theta_1)\ B(y,\theta_2)\ \ [\ S(y,\theta_2) - S(y,\theta_1)\ ]\ .$$

As was discussed in the transparencies handout, near $y = 0$, $S(y,\theta)$ is a decreasing function of θ, and therefore the optimal y decreases when π increases, but near $y = \overline{y}$, $S(y,\theta)$ is increasing in θ, and therefore the optimal y increases when π increases.

Once again, the optimized value V^* decreases as π increases, so if the probability of the bad type is sufficiently high, A may do better to take his outside opportunity a than to initiate the stage game with B.

D. Stochastic Temptation: Several Types of Player B

In this section we focus on local comparative statics of an interior optimum, leaving aside the multiplicity issues that were discussed in the context of two types.

Suppose that θ can take on any of n values

$$\theta_1 < \theta_2 < \ldots < \theta_n .$$

A's prior probabilities for these are π_1, π_2, ... π_n respectively. Here we will do comparative statics by letting all these probabilities be functions of a parameter ζ, written $\pi_i(\zeta)$. An increase in ζ shifts the distribution of θ to the right, and we want such a shift to increase A's optimal choice of y, that is, $dy/d\zeta > 0$. The precise condition that is needed in this context will emerge during the calculation.

Steps similar to those for the two-type case give

$$V(y, \zeta) = c - L(y, \zeta)\ (c + \gamma y)\,,$$

where L is the ex ante probability of A getting a bad payoff, obtained by averaging the B function over the distribution of θ:

$$L(y, \zeta) = \mathrm{E}_\theta[B] = \sum_i\ B(y, \theta_i)\ \pi_i(\zeta)\ , \tag{30}$$

and

$$\mathrm{Sign}\left(\frac{dy}{d\zeta}\right) = \mathrm{Sign}(V_{y\zeta}) = \mathrm{Sign}(L_y\ L_\zeta - L\ L_{y\zeta})\,.$$

Now

$$L_y(y, \zeta) = \sum_i\ B_y(y, \theta)\ \pi_i(\zeta) = \mathrm{E}_\theta[B_y] = -\mathrm{E}_\theta[S\,B]$$

recalling the definition $S = -B_y/B$. Therefore

$$L_y\ L_\zeta - L\ L_{y\zeta} = -\mathrm{E}_\theta[S\,B]\ \frac{\partial \mathrm{E}_\theta[B]}{\partial \zeta} + \mathrm{E}_\theta[B]\ \frac{\partial \mathrm{E}_\theta[S\,B]}{\partial \zeta}\,.$$

This is positive if

$$\frac{\partial \mathrm{E}_\theta[S\,B]\,/\,\partial \zeta}{\mathrm{E}_\theta[S\,B]} > \frac{\partial \mathrm{E}_\theta[B]\,/\,\partial \zeta}{\mathrm{E}_\theta[B]}\,.$$

Note that all this is being done for fixed y. Now B is an increasing function of θ, so a first-order stochastic dominant shift to the right of the distribution of θ increase the expectation of B. If S is also an increasing function of θ, then we may expect the expectation of the product SB to increase proportionately faster than does S. That will fulfill the above inequality; then the optimal y will increase as the shift parameter ζ increases.

While this is easier to lay out in terms of comparative static derivatives, the analysis is easier to comprehend, and to use later in the Bayesian updating calculations, when expressed in terms of a discrete parameter shift. Let the probabilities change from π_1, π_2, ... π_n, to Π_1, Π_2, ... Π_n. Write $B_i = B(y, \theta_i)$ and $S_i = S(y, \theta_i)$. We want

$$\frac{\sum_i \; S_i \, B_i \, \Pi_i}{\sum_i \; S_i \, B_i \, \pi_i} > \frac{\sum_i \; B_i \, \Pi_i}{\sum_i \; B_i \, \pi_i}$$

Write this as

$$\sum_i \; \frac{S_i \, B_i \, \pi_i}{\sum_j \; S_j \, B_j \, \pi_j} \, \frac{\Pi_i}{\pi_i} > \sum_i \; \frac{B_i \, \pi_i}{\sum_j \; B_j \, \pi_j} \, \frac{\Pi_i}{\pi_i} \, .$$

Or, defining the weights

$$\omega_i = \frac{S_i \, B_i \, \pi_i}{\sum_j \; S_j \, B_j \, \pi_j}, \qquad \nu_i = \frac{B_i \, \pi_i}{\sum_j \; B_j \, \pi_j} \, .$$

Now the inequality we seek corresponds to the following inequality between two weighted averages of the new to old probabilities Π_i / π_i:

$$\sum_i \; \omega_i \, \frac{\Pi_i}{\pi_i} > \sum_i \; \nu_i \, \frac{\Pi_i}{\pi_i} \, .$$

If the function S is increasing in θ, the ratio of the weights $\omega_i / \nu_i = S_i$ is increasing in i. Then we will have the desired inequality if the ratio Π_i / π_i is itself increasing in i.

A first-order stochastic dominant shift to the right in the distribution of θ is not enough for this. We need something stronger, namely that the ratio of probability densities is increasing; this is called a shift to the right in the "likelihood ratio order". For this, it is necessary and sufficient that if we look at the distribution of θ restricted to an arbitrarily given interval of its support, it shifts to the right over that interval; see Shaked and Shanthikumar (1994, Theorem 1.C.2 on p.29). Thus the requirement that two distributions are comparable in the likelihood ratio order is quite demanding. However, we will see that a prior distribution of types, and its Bayesian posterior calculated after A experiences a high or a low payoff, are fortuitously comparable in this strong sense.

Two further remarks: [1] With just two types, $\pi_1 + \pi_2 = \Pi_1 + \Pi_2 = 1$, so a first-order stochastic dominant shift to the right simply means $\Pi_1 < \pi_1$ and therefore $\Pi_2 > \pi_2$. Then $\Pi_2 / \pi_2 > \Pi_1 / \pi_1$, so a shift in the likelihood ratio order is equivalent to the first-order stochastic dominant shift. That is why we did not need this additional condition with just two types. [2] The stronger condition of a shift to the right in the likelihood ratio order is only sufficient for the proof, but it is almost necessary in the sense that if it is violated, we can construct an example with appropriate S and H functions etc. that will contradict the result we want.

E. Repeated Game with Infinitely Impatient Players

So far we were considering a one-shot game. Even with infinitely impatient players who do not take into account future payoffs, however, successive plays of the game are linked by the evolution of information about types. Therefore next suppose that the game is repeated over a sequence of periods. Both players are infinitely impatient, that is, in each stage game they are concerned only with their payoffs in that period. But they have perfect memories, so A does the correct Bayesian revision of the π_i each period in the light of his payoff the previous period. Continue the set-up of the last section, where B comes in n discrete types θ_i arranged in increasing order, and the prior probability of B's type being i is π_i. If A chooses to deter temptations up to y, and gets the bad payoff 0, his posterior distribution over types will be given by

$$\Pi_i(0) = \frac{\pi_i \; B(y, \theta_i)}{\sum_j \; \pi_j \; B(y, \theta_j)} \, . \tag{31}$$

Then

$$\frac{\Pi_i(0)}{\pi_i} = \frac{B(y, \theta_i)}{\sum_j \; \pi_j \; B(y, \theta_j)} \, .$$

Because B is increasing in θ, this ratio is increasing in i. Next period A will use the $\Pi_i(0)$ as the prior probabilities. Compared to the original prior, this satisfies the condition of a shift to the right in the likelihood ratio order. Therefore, if the other condition (S increasing in θ) is met, a bad experience will lead A to use a higher punishment. It is very fortuitous and interesting that the seemingly demanding comparative static condition of a shift to the right in the likelihood ratio order is automatically satisfied for A's Bayesian updating upon getting a bad payoff.

In the opposite case, if in the first period A gets the good payoff c, his posterior probabilities will be

$$\Pi_i(c) = \frac{\pi_i \; [1 - B(y, \theta_i)]}{\sum_j \; \pi_j \; [1 - B(y, \theta_j)} \, . \tag{32}$$

Then

$$\frac{\Pi_i(c)}{\pi_i} = \frac{1 - B(y, \theta_i)}{\sum_j \; \pi_j \; [1 - B(y, \theta_j)]} \, .$$

This ratio is a decreasing function of i. Therefore the $\Pi(c)$ distribution is a shift of the π distribution to the left in the likelihood ratio order; therefore, if the other condition (S increasing in θ) is met, a good experience will lead A to use a smaller punishment.

F. Repeated Game with Finitely Impatient A

Now suppose A has finite patience with discount factor δ, but B remains infinitely impatient. Let B have n types θ_i as before, and let A's prior be denoted by the vector

$$\pi = (\pi_1, \pi_2, \; \ldots \; \pi_n) \, .$$

Recall that A's ex ante probability of getting the bad payoff with this prior and the choice of deterring temptations up to y is

$$L(y, \pi) = \sum_i \; \pi_i \; B(y, \theta_i) \, ,$$

and the one-period payoff is

$$V(y,\pi) = c - L(y,\pi)\ (c+\gamma y)\,.$$

Player A's Dynamic Programming Problem

A's recursive dynamic programming problem can be written as

$$W(\pi) = \max_y \ \Big\{ V(y,\pi) + \delta\ \{ L(y,\pi)\ W(\Pi(0)) + [1 - L(y,\pi)]\ W(\Pi(c))\ \}\ \Big\}\,, \tag{33}$$

where $\Pi(0)$ and $\Pi(c)$ denote the vectors of the posterior probabilities of types after getting payoffs zero and c respectively, with the components defined in (31) and (32) respectively. Note that both $\Pi(0)$ and $\Pi(c)$ depend on the vector of prior probabilities π and the choice of the maximum temptation to deter y.

Now we establish some basic properties of the function W. For this we need some new notation

$$\begin{array}{rcl} S^n & = & \text{Unit simplex in } R^n = \{\ \pi = (\pi_1, \pi_2\ \ldots\ \pi_n) \mid \pi_i \geq 0, \Sigma_i\, \pi_i = 1\ \} \\ \mathcal{C} & = & \text{Class of continuous functions } S^n \mapsto [-\gamma\,\overline{y}/(1-\delta), c/(1-\delta)] \end{array}$$

For a function $W \in \mathcal{C}$, define another function $\mathcal{S}W : [0,\overline{y}] \times S^n \mapsto [-\gamma\,\overline{y}/(1-\delta), c/(1-\delta)]$ by

$$\mathcal{S}W(y,\pi) = V(y,\pi) + \delta\ \{\ L(y,\pi)\ W(\Pi(0)) + [1 - L(y,\pi)]\ W(\Pi(c))\ \}\,. \tag{34}$$

Using the fact that V is continuous and satisfies $-\gamma\,\overline{y} \leq V(y,\pi) \leq c$ for all (y,π), it is easy to check that $\mathcal{S}W$ is indeed continuous and that its values are in the stated range when W has these properties. Also define another function $\mathcal{T}W$ by:

$$\mathcal{T}W(\pi) = \max_{0 \leq y \leq \overline{y}}\ \mathcal{S}W(y,\pi)\,. \tag{35}$$

It is similarly easy to check that $\mathcal{T}W$ is in the class $\mathcal{C}$ when W is; thus $\mathcal{T}$ is an operator that maps $\mathcal{C}$ into itself. It is a contraction mapping and therefore has a unique fixed point W^*, which is the Bellman value function of player A's dynamic optimization problem. The proofs of these are standard as in Stokey, Lucas and Prescott (1989).

Definition: A function $W \in \mathcal{C}$ has the property DEC if, whenever the probability distribution given by π shifts to the right in the sense of the likelihood ratio order (LR), $W(\pi)$ (weakly) decreases.

Lemma 1: If W has the property DEC, so does $\mathcal{T}W$.

Proof: Consider two vectors π^a, $\pi^b \in S^n$ such that the probability distribution given by π^b is a shift to the right of that given by π^a in the LR sense. Using the Bayesian updating formulas, the ratios of the corresponding posterior probabilities are: if A gets zero payoff,

$$\frac{\Pi(0)^b_i}{\Pi(0)^a_i} = \frac{\pi^b_i}{\pi^a_i}\ \frac{\sum_j\ \pi^a_j\, B_j}{\sum_j\ \pi^b_j\, B_j}\,,$$

and if A gets positive payoff,

$$\frac{\Pi(c)_i^b}{\Pi(c)_i^a} = \frac{\pi_i^b}{\pi_i^a} \, \frac{\sum_j \, \pi_j^a \, (1 - B_j)}{\sum_j \, \pi_j^b \, (1 - B_j)} \, ,$$

where we have used the abbreviation $B_j = B(y, \theta_j)$. Since π_i^b / π_i^a increases as i increases, so does each of the posterior probability ratios. Therefore each of the 0 and c posteriors for the π^b distribution is a rightward shift in the LR sense of the corresponding one for the π^a distribution. Using the assumed DEC property of W, we have

$$W(\Pi(0)^b) \leq W(\Pi(0)^a), \qquad W(\Pi(c)^b) \leq W(\Pi(c)^a) \, .$$

Next, for $k = a$ or b,

$$\frac{\Pi(0)_i^k}{\Pi(c)_i^k} = \frac{B_i}{1 - B_i} \, \frac{\sum_j \, \pi_j^k \, (1 - B_j)}{\sum_j \, \pi_j^k \, B_j}$$

is increasing in i; therefore $\Pi(0)^k$ is a shift to the right of $\Pi(c)^k$ in the LR sense. Therefore by the assumption that W has the property DEC, $W(\Pi(0)^k) \leq W(\Pi(c)^k)$.

Finally, a shift to the right in the LR sense implies a first-order stochastic dominant shift to the right, and the H_i are increasing in i, therefore

$$L(y, \pi^b) = \sum_i \, \pi^b \, B_i > \sum_i \, \pi^a \, B_i = L(y, \pi^a) \, .$$

Then $V(y, \pi^b) < V(y, \pi^a)$.

Using all this information, for any fixed y, we have

$$\begin{aligned} \mathcal{S}W(y, \pi^a) &= V(y, \pi^a) + \delta \, \{ \, L(y, \pi^a) \, W(\Pi(0)^a) + [1 - L(y, \pi^a)] \, W(\Pi(c)^a) \, \} \\ &> V(y, \pi^b) + \delta \, \{ \, L(y, \pi^a) \, W(\Pi(0)^b) + [1 - L(y, \pi^a)] \, W(\Pi(c)^b) \, \} \\ &\geq V(y, \pi^b) + \delta \, \{ \, L(y, \pi^b) \, W(\Pi(0)^b) + [1 - L(y, \pi^b)] \, W(\Pi(c)^b) \, \} = \mathcal{S}W(y, \pi^b) \end{aligned}$$

In going from the first line to the second line, we have used the above inequalities on the values at π^a versus π^b for the V and the two Ws, leaving the L and $(1-L)$ weights unchanged; in going from the second line to the third line we have shifted the weights toward the smaller of the two W-values.

For $k = a$, b, let

$$y^k = \arg\max \; S(y, \pi^k) \, .$$

Then evaluating the above inequality at y^b, we have

$$\mathcal{S}W(y^b, \pi^a) \geq \mathcal{S}W(y^b, \pi^b) = \mathcal{T}W(\pi^b) \, ,$$

and of course

$$\mathcal{T}W(\pi^a) \geq \mathcal{S}W(y^b, \pi^a) \, .$$

This completes the proof.

Proposition 1: The fixed point W^* defined by $W = \mathcal{T}W$ has the property DEC.

Proof: Lemma 1 has shown that the operator $\mathcal{T}$ maps the closed subspace of $\mathcal{C}$ consisting of functions that have the property DEC into itself; therefore the unique fixed point of $\mathcal{T}$ belongs to this subspace.

Corollary 1: For any $\pi \in S^n$ and its posteriors $\Pi(0)$ and $\Pi(c)$ in the events of A receiving zero and positive payoffs, we have $W^*(\Pi(0)) < W^*(\pi) < W^*(\Pi(c))$.

Proof: $\Pi(0)$ is a shift to the right, and $\Pi(c)$ a shift to the left, both in the LR sense, of π. Then the property DEC of W^* yields the result.

Lemma 2: If $W \in \mathcal{C}$ is convex, so is $\mathcal{T}W$.

Proof: Take any two π^a, $\pi^b \in S^n$, and define

$$\pi^m = \alpha \ \pi^a + (1-\alpha) \ \pi^b \qquad \text{for } \alpha \in (0,1).$$

Begin by considering the relationships between the corresponding posteriors after a good or a bad payoff. For $k = a$, b and m, define

$$L^k = \sum_j \ \pi_j^k \, B_j \, .$$

Then

$$\begin{aligned} L^m &= \sum_j \ \pi_j^m \, B_j = \sum_j \ [\alpha \ \pi_j^a + (1-\alpha) \ \pi_j^b] \ B_j \\ &= \alpha \ \sum_j \ \pi_j^a \, B_j + (1-\alpha) \ \sum_j \ \pi_j^b \ B_j \\ &= \alpha \ L^a + (1-\alpha) \ L^b \, . \end{aligned}$$

Note that for $k = a$, b and m,

$$\Pi(0)_i^k = \frac{\pi_i^k \ B_i}{\sum_j \ \pi_j^k \, B_j} = \frac{\pi_i^k \ B_i}{L^k}$$

Using this notation and relationships, we have

$$\begin{aligned} \Pi(0)_i^m &= \frac{\pi_i^m \ B_i}{L^m} \\ &= \frac{[\alpha \, \pi_i^a + (1-\alpha) \, \pi_i^b] \ B_i}{L^m} \\ &= \frac{\alpha \, L^a}{L^m} \ \frac{\pi^a \, B_i}{L^a} + \frac{(1-\alpha) \, L^b}{L^m} \ \frac{\pi_i^b \, B_i}{L^b} \\ &= \alpha_0 \ \Pi(0)_i^a + (1-\alpha_0) \ \Pi(0)_i^b \, , \end{aligned}$$

where we have defined

$$\alpha_0 = \frac{\alpha \, L^a}{L^m}, \qquad \text{and then} \qquad 1 - \alpha_0 = \frac{(1-\alpha) \, L^b}{L^m} \, .$$

Combining these component-by-component relations into the whole vector of posterior probabilities,

$$\Pi(0)^m = \alpha_0\ \Pi(0)^a + (1-\alpha_0)\ \Pi(0)^b\,.$$

Similarly, for the posterior upon getting positive payoff, we have

$$\Pi(c)_i^m = \alpha_c\ \Pi(c)_i^a + (1-\alpha_c)\ \Pi(c)_i^b\,,$$

where

$$\alpha_c = \frac{\alpha\,(1-L^a)}{(1-L^m)}, \qquad 1-\alpha_c = \frac{(1-\alpha)\,(1-L^b)}{(1-L^m)}\,.$$

Then

$$\Pi(c)^m = \alpha_c\ \Pi(c)^a + (1-\alpha_c)\ \Pi(c)^b\,.$$

Using these results, we have for any fixed y

$$\begin{aligned}
\mathcal{S}W(y,\pi^m) &= V(y,\pi^m) + \delta\ \{\ L(y,\pi^m)\ W(\Pi(0)^m) + [1-L(y,\pi^m)]\ W(\Pi(c)^m)\ \} \\
&= c - L^m\ (c+\gamma\, y) + \delta\ \{\ L^m\ W(\Pi(0)^m) + [1-L(y,^m]\ W(\Pi(c)^m)\ \} \\
&= \alpha\ [c - L^a\,(c+\gamma\, y)] + (1-\alpha)\ [c - L^b\,(c+\gamma\, y)] \\
&\quad +\delta\ \{\ L^m\ W(\Pi(0)^m) + [1-L(y,\pi^m)]\ W(\Pi(c)^m)\ \} \\
&\le \alpha\ [c - L^a\,(c+\gamma\, y)] + (1-\alpha)\ [c - L^b\,(c+\gamma\, y)] \\
&\quad +\delta\ \{\ L^m\ [\alpha_0\, W(\Pi(0)^a) + (1-\alpha_0)\, W(\Pi(0)^b)] \\
&\qquad +[1-L^m]\ [\alpha_c\, W(\Pi(c)^a) + (1-\alpha_c)\, W(\Pi(c)^b)]\ \} \\
&= \alpha\ [c - L^a\,(c+\gamma\, y)] + (1-\alpha)\ [c - L^b\,(c+\gamma\, y)] \\
&\quad +\delta\ \{\ \alpha\, L^a\, W(\Pi(0)^a) + (1-\alpha)\, L^b\, W(\Pi(0)^b) \\
&\qquad +\alpha\,(1-L^a)\, W(\Pi(c)^a) + (1-\alpha)\,(1-L^b)\, W(\Pi(c)^b)]\ \} \\
&= \alpha\ \{\ [c - L^a\,(c+\gamma\, y)] + \delta\ [\,L^a\, W(\Pi(0)^a) + (1-L^a)\, W(\Pi(c)^a)\,]\ \} \\
&\quad +(1-\alpha)\ \{\ [c - L^b\,(c+\gamma\, y)] + \delta\ [\,L^b\, W(\Pi(0)^b) + (1-L^b)\, W(\Pi(c)^b\,]\ \} \\
&= \alpha\ \{\ V(y,\pi^a) + \delta\ [\,L(y,\pi^a)\, W(\Pi(0)^a) + [1-L(y,\pi^a)]\, W(\Pi(c)^a)\,]\ \} \\
&\quad +(1-\alpha)\ \{\ V(y,\pi^b) + \delta\ [\,L(y,\pi^b)\, W(\Pi(0)^b) + [1-L(y,\pi^b)]\, W(\Pi(c)^b\,]\ \} \\
&= \alpha\ \mathcal{S}W(y,\pi^a) + (1-\alpha)\ \mathcal{S}W(y,\pi^b)
\end{aligned}$$

Here the crucial $\le$ step uses the assumption that W is convex. The other steps use the previous results relating the values of the L function and of the posteriors $\Pi(0)$ and $\Pi(c)$, corresponding to the three values of π being considered.

Evaluate this for y^m, the maximizer of $\mathcal{S}W(y,\pi^m)$. Then

$$\begin{aligned}
\mathcal{T}W(\pi^m) &= \mathcal{S}W(y^m,\pi^m) \\
&\le \alpha\ \mathcal{S}W(y^m,\pi^a) + (1-\alpha)\ \mathcal{S}W(y^m,\pi^b) \\
&\le \alpha\ \mathcal{S}W(y^a,\pi^a) + (1-\alpha)\ \mathcal{S}W(y^b,\pi^b) \\
&= \alpha\ \mathcal{T}W(\pi^a) + (1-\alpha)\ \mathcal{T}W(\pi^b)
\end{aligned}$$

This completes the proof.

Proposition 2: The fixed point W^* defined by $W = \mathcal{T}W$ is convex.

Proof: Lemma 2 has shown that the operator $\mathcal{T}$ maps the closed subspace of $\mathcal{C}$ consisting of convex functions into itself; therefore the unique fixed point of $\mathcal{T}$ belongs to this subspace.

For any given π, A's (regular, interior) optimal choice of y is defined by the conditions[5]

$$(\mathcal{S}W^*)_y(y,\pi) = 0, \qquad (\mathcal{S}W^*)_{yy}(y,\pi) < 0\,.$$

Question of Monotonicity of Punishments with Two Types

Now we examine whether A's optimal punishment increases when the distribution of B's types shifts for the worse, considering the case of just two types. It turns out that this is not automatic, therefore we should not expect any less ambiguous results for the general case of several types.

With just two types, let the scalar π denotes the prior probability of B being the bad type 2. The corresponding posterior probabilities upon A receiving the bad and the good payoff are denoted by $\Pi(0)$ and $\Pi(c)$ respectively, and using the general formulas (31) and (32), can be written as

$$\Pi(0) \;=\; \frac{\pi\ B(y,\theta_2)}{L} > \pi \tag{36}$$

$$\Pi(c) \;=\; \frac{\pi\ [1 - B(y,\theta_2)]}{1-L} < \pi\,, \tag{37}$$

where $L(y,\pi)$ is as defined in (29),

$$L(y,\pi) = (1-\pi)\ B(y,\theta_1) + \pi\ B(y,\theta_2)$$

and is A's prior probability of getting the low payoff when his prior probability of B being the bad type is π and he is deterring temptations up to y.

When the Bellman value function is the appropriate fixed point W^*, write the function that is maximized by choice of y as

$$\mathcal{S}W^*(y,\pi) = V(y,\pi) + \delta\ \{\ L(y,\pi)\ W^*(\Pi(0)) + [1 - L(y,\pi)]\ W^*(\Pi(c))\ \}\,. \tag{38}$$

Recognize that

$$L(y,\pi)\ W^*(\Pi(0)) + [1 - L(y,\pi)]\ W^*(\Pi(c))$$

is just A's expected continuation payoff; write it as $EW^*(y,\pi)$. Using this abbreviation, but expanding out the V function in $\mathcal{S}W^*$, we have

$$\mathcal{S}W^*(y,\pi) = c - L(y,\pi)\ (c + \gamma\, y) + \delta\ EW^*(y,\pi)\,.$$

The first-order condition for maximization is

$$\mathcal{S}W^*_y = -\gamma\ L - (c + \gamma\, y)\ L_y + \delta\ EW^*_y = 0\,.$$

[5]The parentheses in $(\mathcal{S}W^*)$ indicate that the whole function $(\mathcal{S}W^*)$ is being differentiated. If we had said $\mathcal{S}W^*_y$, it might appear as if the operator $\mathcal{S}$ is applied to the function W^*_y.

As usual, to get $dy/d\pi > 0$ we want $SW^*_{y\pi} > 0$ at the optimal point. Now

$$SW^*_{y\pi} = -\gamma\ L_\pi - (c + \gamma\, y)\ L_{y\pi} + \delta\ EW^*_{y\pi}\,.$$

Multiplying the equation for $SW^*_{y\pi}$ by L, that for SW^*_y by $-L_\pi$, and adding the two, we get

$$L\ SW^*_{y\pi} - L_\pi\ SW^*_y = (c + \gamma\, y)\ (L_y\, L_\pi - L\ L_{y\pi}) + \delta\ (L\ EW^*_{y\pi} - L_\pi\ EW^*_y)\,.$$

So evaluated at the optimal y where $SW^*_y = 0$, we have

$$L\ SW^*_{y\pi} = (c + \gamma\, y)\ (L_y\, L_\pi - L\ L_{y\pi}) + \delta\ (L\ EW^*_{y\pi} - L_\pi\ EW^*_y)\,. \qquad (39)$$

We know from previous work that if $S(y,\theta) = -B_y(y,\theta)\,/\,B(y,\theta)$ is increasing in θ for fixed y, then the first term on the right hand side is positive. So let us examine the conditions under which the second term on the right hand side is positive. This is only a sufficient condition, but it is "almost necessary" if we want the result to hold for all δ, because W implicitly contains a factor $1/(1-\delta)$ so for δ close to 1 the sign of the second term is going to determine the sign of the whole expression.

Note that

$$\frac{\partial}{\partial \pi}\,\frac{EW^*_y}{L} = \frac{L\ EW^*_{y\pi} - EW^*_y\ L_\pi}{L^2}\,.$$

Therefore we want $EW^*_y\,/\,L$ to be an increasing function of π for fixed y.

The following two lemmas establish some useful properties of EW^*.

Lemma 3: When $\pi = 0$ or $\pi = 1$, $EW^*_y = 0$.

Proof: Differentiating the expression for EW^*,

$$EW^*_y = [W^*(\Pi(0)) - W^*(\Pi(c))]\ L_y + L\ W^{*'}(\Pi(0))\ \frac{\partial \Pi(0)}{\partial y} + (1-L)\ W^{*'}(\Pi(c))\ \frac{\partial \Pi(c)}{\partial y}.$$

When $\pi = 0$, both posteriors $\Pi(0) = \Pi(c) = 0$, so $W(\Pi(0)) - W(\Pi(c)) = 0$. Also, from the expressions obtained in Lemma 1,

$$\frac{\partial \Pi(0)}{\partial y} = 0 = \frac{\partial \Pi(c)}{\partial y} \qquad \text{when } \pi = 0\,.$$

Therefore $EW^*_y = 0$ at $\pi = 0$. A similar argument applies at $\pi = 1$.

We saw earlier that $S_\theta > 0$ when y is close to $\overline{y}$ and $S_\theta < 0$ when y is close to 0. Therefore there exist y_H and y_L such that

$$S_\theta \begin{cases} < 0 & \text{for } 0 < y < y_L \\ > 0 & \text{for } y_H < y < \overline{y} \end{cases}$$

When y lies in these ranges, we can say more about $dy/d\pi$.

Lemma 4: When $0 < \pi < 1$, we have

$$EW^*_y \begin{cases} > 0 & \text{if } 0 < y < y_L \\ < 0 & \text{if } y_H < y < \overline{y} \end{cases}$$

Proof: When y changes,

$$\begin{aligned}\frac{\partial\,\Pi(0)}{\partial y} &= \frac{1}{L^2}\,\{\ [(1-\pi)\ B(y,\theta_1)+\pi\ B(y,\theta_2)]\,\pi\,B_y(y,\theta_2)\\ &\qquad -\pi\,B(y,\theta_2)\,[(1-\pi)\ B_y(y,\theta_1)+\pi\ B_y(y,\theta_2)]\ \}\\ &= \frac{\pi(1-\pi)\ [B(y,\theta_1)\ B_y(y,\theta_2)-B_y(y,\theta_1)\ B(y,\theta_2)]}{L^2}\\ &= \frac{-\pi(1-\pi)\ B(y,\theta_1)\ B(y,\theta_2)\ [S(y,\theta_2)-S(y,\theta_1)]}{L^2}\,.\end{aligned}$$

Similarly we find

$$\frac{\partial\,\Pi(c)}{\partial y} = \frac{\pi(1-\pi)\ [1-B(y,\theta_1)]\,[1-B(y,\theta_2)]\ [S(y,\theta_2)-S(y,\theta_1)]}{(1-L)^2}\,.$$

From the definition of y_L and y_H, we have

$$S(y,\theta_2)-S(y,\theta_1)\begin{cases} <0 & \text{for } 0<y<y_L\\ >0 & \text{for } y_H<y<\overline{y}\end{cases}$$

Therefore

$$\begin{aligned}&\text{When } y<y_L, \qquad \frac{\partial\,\Pi(0)}{\partial y}>0 \quad\text{and}\quad \frac{\partial\,\Pi(c)}{\partial y}<0\,,\\ &\text{When } y>y_H, \qquad \frac{\partial\,\Pi(0)}{\partial y}<0 \quad\text{and}\quad \frac{\partial\,\Pi(c)}{\partial y}>0\,.\end{aligned}$$

Also note that $\Pi(c)<\pi<\Pi(0)$, and

$$L(y,\pi)\ \Pi(0)+[1-L(y,\pi)]\ \Pi(c)=\pi\,.$$

Therefore, when $y<y_L$, an increase in y causes a mean-preserving spread of A's posterior probabilities. Since $W^*(\pi)$ was proved to be a convex function (in Proposition 2 above), the expected value EW^* increases. Thus, when $y<y_L$, $EW_y^*>0$.

Conversely, when $y>y_H$, an increase in y causes a mean-preserving contraction of the spread of A's posterior probabilities. By convexity of $W^*(\pi)$, the expected value EW^* decreases, that is, $EW_y^*<0$.

The sign of EW_y^*/L is the same as that of EW_y^*, therefore Lemmas 3 and 4 immediately translate into implications for the ratio. Figure 4 shows the shape of EW_y^*/L regarded as a function of π for the two cases of high and low y. We know from the calculation just before Lemma 3 that the dynamic effect of π on the optimal punishment y, namely the second term in the expression (39), has the same sign as the slope of EW_y^*/L regarded as a function of π. Therefore the dynamic aspect contributes a positive term to $dy/d\pi$ in two situations, (1) when π is low and y is also low, and (2) when π is high and y is also high. In the other two situations, (3) low π, high y, and (4) high π, low y, the dynamic aspect contributes a negative term to $dy/d\pi$. Recall that parameters c and m affect y separately from π and can trace the whole range $[0, \overline{y}]$, so all four cases are logical possibilities. And note that the "high" and "low" in the above statements are rough and not expressed in terms of precise cutoffs; that would not be possible because for example the value of π for which the EW_y^*/L curve attains its maximum or minimum depends on the exact value of y. Figure 4 in the text illustrates this.

The taxonomy may look complicated, but it can be expressed more simply and interpreted. The general idea is: if π moves away from its extreme values toward an intermediate range, so does y. In cases (1) and (3), when π increases starting from near zero, y increases starting from a low level in case (1) and decreases from a high level in case (3). In cases (2) and (4), if π decreases starting from near 1, then y decreases starting from a high level in case (2) and increases starting from a low level in case (4). The idea that moving π toward an intermediate value also moves y toward an intermediate value can be interpreted in terms of the information value of punishments.

Recall the formula (36) for the posterior probability of B being the bad type after A experiences a low payoff:

$$\Pi(0) = \frac{\pi \ B(y, \theta_2)}{(1-\pi) \ B(y, \theta_1) + \pi \ B(y, \theta_2)} .$$

This can be written as

$$\Pi(0) = \frac{\pi}{(1-\pi) \ B(y, \theta_1)/B(y, \theta_2) + \pi}$$

This regarded as a function of y for fixed π equals π when $y = 0$ (where both the B functions equal μ) and also when $y = \overline{y}$ (when both the B functions equal λ). For all intermediate y, we have $B(y, \theta_1)/B(y, \theta_2) < 1$ so $\Pi(0) > \pi$. Then $\Pi(0)$ will be maximum at an intermediate y, call it y_0. Intuitively, both the extreme punishments – zero (which deters neither type for any temptation) and $\overline{y}$ (which deters both types for all temptations) are uninformative. Information is maximized for an intermediate punishment.

Maximizing $\Pi(0)$ is equivalent to minimizing $B(y, \theta_1)/B(y, \theta_2)$, or equivalently, $\ln[B(y, \theta_1)] - \ln[B(y, \theta_2)]$. The first-order condition for that is

$$\frac{B_y(y, \theta_1)}{B(y, \theta_1)} - \frac{B_y(y, \theta_2)}{B(y, \theta_2)} = 0$$

or

$$S(y, \theta_1) = S(y, \theta_2) .$$

So coincidentally, the y that is the dividing line between the two cases $S_\theta > 0$ and $S_\theta < 0$ is just the y_0 that maximizes the informativeness of the "bad signal".

There will similarly be another critical value of y that maximizes the informativeness of the good signal. Equation (37) can be written as

$$\begin{aligned}
\Pi(c) &= \frac{\pi \ [1 - B(y, \theta_2)]}{(1-\pi) \ [1 - B(y, \theta_1)] + \pi \ [1 - B(y, \theta_2)]} \\
&= \frac{\pi}{(1-\pi) \ [1 - B(y, \theta_1)]/[1 - B(y, \theta_2)] + \pi}
\end{aligned}$$

So $\Pi(c)$ regarded as a function of y equals π at the extremes $y = 0$ and $\overline{y}$, and is smaller than π in the interior. Its minimum occurs where $[1 - B(y, \theta_1)]/[1 - B(y, \theta_2)]$ is maximized, and the first-order condition for that is

$$\frac{B_y(y, \theta_1)}{1 - B(y, \theta_1)} - \frac{B_y(y, \theta_2)}{1 - B(y, \theta_2)} = 0\,.$$

The resulting y_c say depends on θ_1 and θ_2 and also on the parameters μ and λ that are in the function B, but not on π.

If we regard B somewhat like a cumulative distribution function, B_y/B is like a hazard rate and $B_y/(1 - B)$ the reverse hazard rate. So the conditions that maximize the informativeness of the two types of signals are intuitive counterparts of each other.

The value of information also depends on π. When $\pi = 0$, both $\Pi(0)$ and $\Pi(c)$ are also zero; when $\pi = 1$, both posteriors are also 1. Therefore no new information accrues when the prior is at an extreme, no matter what punishment is being used. It is only for intermediate priors that significant information can be obtained by using punishments.

The dynamic effect is essentially that A will be able to start the next period with updated probabilities, and his current calculation takes into account the discounted continuation value. Therefore when information is more likely to accrue (intermediate π), the dynamic consideration argues in favor of using a more informative punishment (intermediate y). We should again emphasize that the statements are deliberately vague in their use of "high" and "low"; there is no clear cutoff and there may be some ambiguities in an intermediate range, for example because of the two different values of y that maximize information upon A's receiving high versus low payoffs.

Now we can put the dynamic effect together with the static effect that led to the bifurcation property. This contributes a positive term to $dy/d\pi$ when y is high and a negative term when y is low. We can combine these results into a table that shows the signs of the two terms that determine the sign of $dy/d\pi$:

		Probability of the bad type π: Low	Probability of the bad type π: High
Level of punishment y	Low	(1): Static −, Dynamic +	(4): Static −, Dynamic −
	High	(3): Static +, Dynamic −	(2): Static +, Dynamic +

Table 4: Static and Dynamic Effects on the Sign of $dy/d\pi$

When the two effects work in the same direction, the overall effect is clear. Where the two go in opposite directions, the balance depends on δ: the static aspect will be determinative when δ is low, and the dynamic effect when δ is high (close to 1). Therefore we find that there are two cases when the overall result is clear: (2) when π is high and y is high, an increase in π further increases y, and and (4) when π is high and y is low, a further increase in π further lowers y.

In the other two cases, the findings of case studies are borne out only if A's patience is of the appropriate kind. In case (1), when π is low and y is also low, the dynamic effect indicates that an increase in π should lead to higher y, and the dynamic effect will outweigh the static effect which operates in the opposite direction if δ is high. So if the probability that B is bad is low, A is very patient, and the harm to A from B's cheating is sufficiently low and/or the cost to A of actually inflicting punishment is sufficiently high, then a succession of bad experiences will lead A to raise the level of punishment. In case (3), where π is low and y is high, only a very impatient A will behave in conformity with the findings of the case studies.

We emphasize again that the uses of "low" and "high" in reference to values of π and y are deliberately vague, and that in intermediate ranges of these variables there may be reversals of direction. We hope that future research will pin down some of these possibilities more precisely.

References

Shaked, Moshe and J. George Shanthikumar. 1994. *Stochastic Orders and Their Applications.* San Diego, CA: Academic Press.

Stokey, Nancy L., Robert E. Lucas, Jr. and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics.* Cambridge, MA: Harvard University Press.