Carrots vs. Sticks

Dari-Mattiacci, G.; De Geest, G.

DOI
10.1093/oxfordhb/9780199684267.013.41

Publication date
2017

Document Version
Author accepted manuscript

Published in
Oxford Handbook of Law and Economics. - Volume 1

UNIVERSITY OF AMSTERDAM

CARROTS VS. STICKS

Giuseppe Dari-Mattiacci
Gerrit De Geest

Amsterdam Law School Legal Studies Research Paper No. 2016-37
Amsterdam Center for Law & Economics Working Paper Paper No. 2009-13
Washington U. School of Law Working Paper No. 09-09-03

Chapter 21

Carrots vs. Sticks

Giuseppe Dari-Mattiacci and Gerrit De Geest

Forthcoming in Franceco Parisi (ed.) The Oxford Handbook of Law & Economics

21.1 Introduction

Incentives can be generated by either carrots—promises to reward, such as with prizes or bonuses—or sticks—threats to punish, such as with fines or damages.[1] Carrots and sticks are *prima facie equivalent*, because any behavioral change induced by promising compliers a $100 reward can also be obtained by threatening violators with a $100 punishment. Yet, while carrots and sticks seem to produce the same effects, they are not chosen at random; some general patterns can be observed across legal systems. Incentives for careful driving are generally created by holding negligent drivers liable under tort law, rather than by rewarding careful drivers. Likewise, theft is discouraged by penalizing thieves and not by rewarding those who do not steal. Incentives to invent are instead created by rewarding successful inventors with patents or academic prizes rather than by punishing all others.

The goal of this chapter is to draw a broad picture of the differences between carrots and sticks. Our goal is not only to identify the differences but also to examine their causes and the relationships among them. The chapter is organized as follows. In section 21.2, we characterize several versions of the incentive equivalence result. In

[1] A carrot can be defined as a payment from the principal to the agent upon compliance of the agent. A stick is then a payment from the agent to the principal upon violation by the agent. While a carrot can sometimes be rewritten as a mathematically identical stick through the use of entry fees, this is not the case when enforcement is probabilistic. See De Geest and Dari-Mattiacci (2013).

section 21.3, we analyze the differences between carrots and sticks in terms of transaction costs and, in section 21.4, of risk allocation. Next, we consider wealth and budget constraints (section 21.5); distributional effects (section 21.6); activity-level effects (section 21.7); and the principal's incentive to behave opportunistically *ex post* and the agent's incentives to self-report in a biased way (section 21.8). Section 21.9 discusses pre-compensated, annullable, combined, intra-group redistributed, reversible, and strict liability carrots and sticks. Section 21.10 considers two extensions (political risks and behavioral biases). We conclude in section 21.11 by offering guidelines for the optimal use of carrots and sticks.

21.2 The Incentive Equivalence Result

In principle, a carrot has the same incentive effect as a stick of the same magnitude. If a principal wants an agent to do something with an effort cost of $99, the principal can offer the agent a carrot of $100 or threaten the agent with a stick of $100. In both cases, the agent will comply.

Nonetheless there can be cases in which a carrot and a stick of the same magnitude have different incentive effects. The most obvious case is when money has decreasing marginal utility for the agent, and a monetary carrot is used to induce a non-monetary effort.[2] In this case, adding $100 to the agent's wealth changes the total utility less than taking away $100 from the agent's wealth. Since repeated carrots further increase the agent's wealth, carrots may even have a *saturation effect*: there may be a wealth level at which a carrot can no longer incentivize an agent to make a significant effort.[3]

Aside from the decreasing marginal utility of money, the incentive equivalence result holds under a broad set of conditions. To identify these conditions, consider a simple model in which a benevolent ruler (the principal, "she") enforces a rule of conduct over a population of wealth-maximizing risk-neutral individuals (the agents, "he"). Compliance with the rule yields a benefit for society but entails a private effort cost, which is different across individuals. The level of enforcement determines a cut-off level

---

[2] Risk aversion might also be invoked to justify a difference.

[3] Note that saturation may also occur when the effort is monetary but the carrot is overcompensatory, so that the agent receives a rent and becomes wealthier in each round.

of effort so that all individuals with effort cost less than or equal to the cut-off exert effort. We will therefore refer to the cut-off level of effort as *level of enforcement*. To be effective, the rule of conduct is supported either by a stick for violators or by a carrot for compliers.

We take the desired level of enforcement as given and ask the question whether this level can be best achieved through carrots or through sticks.[4] As long as we assume risk neutrality of all parties involved, this question is equivalent to the question whether enforcement is cheaper (or more effective) with carrots or with sticks.[5] Let us first demonstrate three basic incentive equivalence results.

21.2.1 Simple Case: Certain Monitoring with No Errors

Let us introduce the following notation:[6]

e = the effort cost of an individual agent (e^* is the level of enforcement);

c = the reward (carrot);

s = the punishment (stick);

n = the number of agents in the population;

$F(e)$ = the distribution of the cost of effort.

Assume that monitoring occurs with probability equal to one and that the ruler can perfectly and costlessly verify if an agent has exerted effort. Note that the ruler cannot verify the individual's cost of effort and cannot use individualized carrots and sticks.[7]

If the ruler uses carrots, an individual earns $c—e$ upon compliance and 0 upon violation; thus, an individual will comply if his effort cost is $e \leq c$. Similarly, if the ruler uses sticks, an individual pays e upon compliance and s upon violation; thus, he will comply if $e \leq s$. Carrots and sticks are equivalent in terms of incentives: a carrot and a stick of the same magnitude induce the same level of compliance. Therefore, if the ruler wants to induce all individuals with effort cost $e \leq e^*$ to comply—so that, in expectation,

---

[4] The socially optimal level of enforcement is a separate question that is the object of the (extensive) economic literature on the public enforcement of law (see Polinsky and Shavell, 2000). Note that this literature focuses on sticks (because criminal law is a stick-based regime) and that there may be unexplored subtle differences between carrots and sticks in this respect.

[5] Risk adds complications that will be discussed in section 21.4.

[6] Unless otherwise noted, all variables are positive. All probability variables are between 0 and 1.

[7] We discuss individualized carrots and sticks in section 21.6.2.

$nF(e^*)$ individuals comply and $n\left[1 - F\left(e^*\right)\right]$ violate—she can do so equivalently with a carrot or with a stick, so that we have the following incentive equivalence equation:[8]

$$c = s = e^* \qquad (1)$$

21.2.2 Probabilistic Monitoring

The ruler might monitor the behavior of individuals with a probability less than one, so that

p_c = the monitoring probability with carrots;

p_s = the monitoring probability with sticks.

With carrots, an individual will comply with the rule if $e \leq p_c c$. Likewise, with sticks, an individual will comply with the rule if $e \leq p_s s$. Again, carrots and sticks are equivalent with respect to incentives in that the same level of enforcement e^* can be obtained equivalently through carrots or through sticks:

$$p_c c = p_s s = e^* \qquad (2)$$

Note that the combination of probability and sanction is not unique and that if $c > s$, then $p_c < p_s$ and vice versa. If instead carrots and sticks are applied with the same probability $p_c = p_s = p$, then we have the following stronger version of the incentive equivalence equation with probabilistic monitoring:

$$c = s = e^* / p \qquad (3)$$

21.2.3 Enforcement Errors

While verifying whether an individual exerted effort, the ruler might make two types of errors. Even though an individual exerted effort, the ruler might erroneously believe that the individual exerted no effort (a type-I error). Similarly, even though an individual exerted no effort, the ruler might erroneously believe that the individual exerted effort (a type-II error):[9]

ε_I = probability of type-I error;

ε_{II} = probability of type-II error.

---

[8] The level of enforcement e^* is taken as given; its optimal setting is the focus of a large literature on public law enforcement (see note 4).

[9] It is natural to assume that ε_I and ε_{II} are less than ½.

With carrots, an individual earns $p_c(1-\varepsilon_I)c-e$ upon compliance and $p_c\varepsilon_{II}c$ upon violation, since there is a probability ε_I that a complier is not rewarded and a probability ε_{II} that a violator is rewarded. Hence, an individual will comply with the rule if $e \le p_c(1-\varepsilon_I-\varepsilon_{II})c$. Likewise, with sticks, an individual pays $p_s\varepsilon_I s+e$ upon compliance and $p_s(1-\varepsilon_{II})s$ upon violation and will comply with the rule if $e \le p_s(1-\varepsilon_I-\varepsilon_{II})s$.[10] Again, the incentive effects of carrots and sticks are the same and are similarly diluted by errors. If the monitoring probabilities differ, incentive equivalence takes the form

$$p_c c = p_s s = \frac{e^*}{1-\varepsilon_I-\varepsilon_{II}} \quad (4)$$

If instead carrots and sticks are applied with the same probability $p_c = p_s = p$, we have the following version of the incentive equivalence equation with probabilistic monitoring and enforcement errors:

$$c = s = \frac{e^*}{p(1-\varepsilon_I-\varepsilon_{II})} \quad (5)$$

21.3 Transaction Costs

21.3.1 Carrots Are Applied Upon Compliance, Sticks Applied Upon Violation

Assume now that applying carrots and sticks generates a transaction cost. Note that we need to consider only the cost of applying the sanction and not the cost of monitoring the agent; the latter cost depends on the probability of monitoring p (which is only indirectly affected by the choice of carrots or sticks through a possible change in the optimal probability of monitoring).[11] Let:

k_c = the unit cost of applying a carrot;

k_s = the unit cost of applying a stick.

---

[10] Png (1986) was the first to show that type-I and type-II errors dilute deterrence in the same way with respect to sticks. In the literature on public enforcement of law there are several papers identifying differential effects of type-I and type-II errors (for instance Lando, 2006; Garoupa and Rizzolli, 2012; Rizzolli and Saraceno, 2013).

[11] The choice of the optimal level of monitoring (and of monitoring costs) and the optimal mix between probability of monitoring and magnitude of the sanction is treated in the literature on public enforcement of law and we abstract from it here; see note 4.

This set-up is very general and it allows the unit costs to be zero, less than one or even greater than one (meaning that applying a sanction costs more to the ruler than the incentive it produces). With $k_c = k_s = 0$ we focus on the pure-transfer case, where a benevolent ruler only considers the incentive effect of monetary sanctions, which simply move resources and hence yield no social costs. As the unit costs increase, the analysis covers situations in which, for instance, a ruler only bears the costs of paying carrots while not internalizing the costs of sticks ($k_c = 1$, $k_s = 0$), or situations in which both carrots and sticks imply some costs for the ruler $\left(k_c > 0, k_s > 0\right)$. Note that this formulation also covers non-monetary carrots and sticks, which typically have different costs for the ruler and the agents. For instance, the cost k_s can be interpreted as the ratio between the unit disutility of a physical punishment (for instance whipping or incarceration) for an individual and the unit costs of applying that punishment for the ruler (for instance the number of hours that the individual cannot work due to the punishment).

The total expected cost of applying a carrot is

$$p_c(1-\varepsilon_I)ck_c nF(e^*) + p_c\varepsilon_{II}ck_c n\left[1 - F\left(e^*\right)\right] \quad (6)$$

which includes the probability of applying the carrot correctly to compliers $p_c(1-\varepsilon_I)$ and incorrectly to violators $p_c\varepsilon_{II}$. Similarly, with sticks, the total expected cost is

$$p_s\varepsilon_I sk_s nF(e^*) + p_s(1-\varepsilon_{II})sk_s n\left[1 - F\left(e^*\right)\right] \quad (7)$$

Given that the enforcement level is the same in both cases, we can set $cp_c = sp_s$ (incentives equivalence). Thus, comparing the two costs, we have that the costs are less with carrots if

$$\frac{k_c}{k_s} < \frac{\left(1-\varepsilon_I-\varepsilon_{II}\right)\left[1-F\left(e^*\right)\right]+\varepsilon_I}{\left(1-\varepsilon_I-\varepsilon_{II}\right)F\left(e^*\right)+\varepsilon_{II}} \quad (8)$$

Otherwise, the costs are less with sticks. Note that the ratio on the right-hand side depends on two things. First, it decreases if the level of enforcement e^* increases, implying that, if more people comply, carrots become relatively more costly to apply. This is due to the fact that, although there are errors, carrots are mostly paid to compliers. Second, it depends on the balance between type-I and type-II errors. If ε_{II} grows relative to ε_I—that is, if there are more frequent errors with violators than with compliers—then applying carrots becomes relatively more costly than applying sticks. The intuition is that the costs of sanctions are realized when sanctions are applied. Thus, wrong applications of a sanction weigh more on costs than wrong denials of a sanction. With carrots, ε_{II} determines the probability of applying the carrot incorrectly and hence weighs more than ε_I, which determines the probability of wrongly denying a reward. The opposite is true with sticks.

Note that this argument applies irrespective of the population n and of the probability of monitoring p and hence is valid also if there is only one individual or if the probability of monitoring is equal to one. Moreover, to abstract from the difference between type-I and type-II errors and focus in general on the quality of the information that the ruler receives, assume that $\varepsilon_I = \varepsilon_{II} = \varepsilon$. Thus, we have:

$$\frac{k_c}{k_s} < \frac{(1-2\varepsilon)\left[1-F\left(e^*\right)\right]+\varepsilon}{(1-2\varepsilon)F\left(e^*\right)+\varepsilon} \qquad (9)$$

The right-hand side increases in ε if $F\left(e^*\right) > ½$ and decreases in ε otherwise. This implies that the effect of reduced enforcement errors on the costs of carrots and sticks depends on the level of enforcement. If more than half of the population complies, then higher enforcement errors translate into higher costs for sticks, because sticks are now frequently applied by mistake. From a different angle, an improved quality of the information that the ruler gathers makes carrots more desirable only if most of the population complies. The condition in (9) embeds two separate effects that need to be disentangled.

First assume that there is full enforcement, $F(e^*) = 1$. This might be the case if there is only one individual with a known effort cost, so that the expected sanction can be set such that the individual has incentives to comply. Then, carrots yield lower costs if the probability of errors is sufficiently high. That is, as the quality of the signal that the ruler receives deteriorates, carrots become more attractive than sticks (Dari-Mattiacci, 2013).

$$\frac{k_c}{k_s} < \frac{\varepsilon}{1-\varepsilon} \qquad (10)$$

21.3.2 Effect of Equilibrium Compliance (Majoritarian Criterion)

Now assume that there are no errors. Here the optimal choice between carrots and sticks depends on the level of enforcement. Then, (9) becomes

$$\frac{k_c}{k_s} < \frac{1 - F\left(e^*\right)}{F\left(e^*\right)} \quad (11)$$

If the costs of carrots and sticks are the same ($k_c = k_s$), we obtain a pure majoritarian criterion: carrots should be applied if fewer than 50% of the agents comply, sticks otherwise (Wittman, 1984). With different costs ($k_c \neq k_s$), the fractions of compliers and violators receive a weight proportional to these costs.

21.3.3 Effect of Information: Sticks May Punish Those Unable to Comply, Carrots May Reward Those Unable to Violate

The previous result suggests that carrots are only superior when the majority of the agents will violate the rule in equilibrium. This raises the question why this would ever be the case in a rational choice framework. If an agent rationally violates a rule, this means that the carrot or stick is not high enough. But in that case, why does the principal not simply increase the carrot or stick until all agents comply?

The reason is that the principal my not have enough information on the individual effort costs of all agents to determine which of them should follow the rule. Suppose the benefit of compliance is always \$100, while the cost is \$90 for agent A and \$110 for agent B. If the principal is benevolent, she wants only agent A to comply with the rule. If she were fully informed, she would specify that the rule only applies to agent A and set the stick or carrot at \$91 so that agent A would comply. But in the real world, individual effort costs may be difficult to verify. In this case, the principal should take into account that some agents will be unable to comply (for instance, if the rule is "compose good music," most citizens might be unable to comply at reasonable cost). Similarly, there may be agents who are unable to violate (for instance, if the rule is "do not hack computers," most citizens might be unable to violate the rule even if they wanted to).

This suggests that carrots may be desirable when the principal has specification problems, that is, when the principal does not know what can be reasonably expected from the agent. This may help to explain why in an increasingly complex society, carrots are increasingly used (De Geest and Dari-Mattiacci, 2013).[12]

An extreme case of enforcement with information problems is the volunteer's dilemma: it is optimal for society that only one of the bystanders (ideally the one with the lowest effort costs) jumps in the water to save a drowning person. In this case, Leshem and Tabbach (2016) show that carrots are superior to sticks in that they minimize the transfers to be made.

21.4 Risk: Carrots Create Risk for Compliers, Sticks Create Risk for Violators

Individuals might be risk-averse or risk-loving. Obviously, their reaction to probabilistic carrots and sticks changes. Yet, it is important to isolate the effect of risk-aversion from the confounding effects of increased wealth: carrots make individuals richer while sticks make them poorer, so that their behavior might change not only as a result of risk-bearing but also because of endowment effects. In order to isolate the effect of risk-bearing, we focus on exponential utility functions, which have the unique property of exhibiting constant risk preferences, irrespective of wealth.

Starting from risk-aversion, let:

$1 - E^{-wa}$ = a risk-averse individual's utility of wealth, with risk-aversion coefficient $a \geq 0$.[13]

If there are no enforcement errors, with carrots, an individual complies if

$$p_c\left(1-E^{-(w+c-e)a}\right)+\left(1-p_c\right)\left(1-E^{-(w-e)a}\right)\geq 1-E^{-wa} \qquad (12)$$

where the left-hand side is what the individual earns upon compliance and the right-hand side is what he earns upon violation. Likewise, with sticks, an individual complies if

$$1-E^{-(w-e)a}\geq p_s\left(1-E^{-(w-s)a}\right)+\left(1-p_s\right)\left(1-E^{-wa}\right) \qquad (13)$$

With $p_c = p_s = 1$, we find the standard equivalence result: the agent complies if $c = s \geq e$. However, with $p_c = p_s = p < 1$, risk-aversion makes individuals react more strongly to

---

[12] Specification problems and errors in the application of the sanction may even explain why Roman slave-masters used sticks for simple, physical tasks and carrots for complex tasks (such as running a business) (Dari-Mattiacci, 2013).

[13] Not to confuse it with effort, we denote the exponential function as $E^{(\cdot)}$.

sticks than to carrots. From (12) and (13) we have that a carrot and a stick applied with the same probability induce compliance with e^* if

$$\frac{1-E^{-e^*a}}{1-E^{-ca}} = p = \frac{1-E^{-e^*a}}{1-E^{-sa}}\left(\frac{E^{-s}}{E^{-e^*}}\right)^a \qquad (14)$$

which implies $c > s > e^*$. Vice versa, if the carrot and the stick have the same magnitude, the probability of monitoring needs to be greater under carrots, $p_c > p_s$. The left-hand side of (14) indicates that the probability of applying a carrot needs to balance the individual's utility when exerting effort with his utility when receiving the carrot, as in the risk-neutral case. Instead, the right-hand side of (14) indicates that the probability of applying a stick needs to balance the individual's utility when exerting effort with his utility when paying the stick, times an additional factor less than one, which is the ratio of the marginal disutilities of the stick and effort. This factor magnifies the effect of sticks (requires lower probabilities of monitoring). Sticks have a stronger incentive effect than carrots because compliance in the case of sticks is risk-free, while it implies risk in the case of carrots.

The case of risk-loving individuals yields the opposite result. Let

$E^{wl} - 1$ = a risk-loving individual's utility of wealth, with risk-loving coefficient $l \geq 0$.

The analysis is similar to the case of risk-aversion and yields $s > c > e^*$ or, conversely, $p_s > p_c$. This is because now risk is a valued feature and hence carrots, which imply risk for compliers, induce more compliance than sticks.

21.5 Wealth and Budget Constraints

The existing wealth of the principal and agents matters for carrots and sticks in two respects.[14] First, a carrot is a transfer from the principal to the agent; a stick is a transfer from the agent to the principal. Therefore, the principal should be able to finance the carrot and agents should be able to pay the stick. The principal's wealth (or the budget she is able to raise) is thus an upper constraint on the use of carrots; the agent's wealth is an upper constraint on the use of sticks. Second, the application of the carrot or stick itself may be costly to the principal. This is especially the case for a non-monetary stick,

[14] Wealth constraints may be binding when the principal wants to minimize the probability of monitoring, as in Becker (1968).

such as imprisonment. Also in this case, the principal's budget may become a binding constraint. We will now analyze how these constraints affect the use of carrots and sticks.

21.5.1 The Maximum Carrot Is Determined by the Wealth of the Principal, the Maximum Stick by the Wealth of the Agent

Carrots need to be financed and sticks need to be paid. If we assume that it is the principal who finances the carrot and consider a simple setting,[15] with only one agent and one required effort, then the maximum carrot equals the principal's wealth (or maximum budget); similarly, the maximum stick equals the agent's wealth.[16]

But when there are more agents or more required efforts and monitoring is probabilistic, the situation becomes more complex. Suppose that there are ten agents, all with a 10% chance to be monitored and to receive a $1 million bonus if found complying. Does the principal need a budget of $1 million or of $10 million? This depends on whether monitoring is coordinated. If monitoring is organized in such a way that there is only exactly one agent monitored, the principal never has to pay more than $1 million. If monitoring is uncoordinated, so that it may happen that all agents are monitored at the same time, the principal may have to pay $10 million.

To analyze the wealth (and budget) constraints more formally, let:

w = the wealth of an individual agent;[17]

b = the ruler's budget.

Carrots and sticks are bound by the ruler's budget available to apply the carrot or the stick. At the individual level, the maximal stick that can be applied is also bound by the individual level of wealth $s^{max} = w$. In general, one could envisage a boundary for carrots as well, which may be due to saturation, linked somehow to the individual's wealth. In the following, we focus on the ruler's constraints.

From the ruler's perspective, the constraints on the total maximum expenditures in equilibrium can be derived from the total costs (6) and (7). Note, however, that these equations reflect *expected* total costs for the ruler. This implies that costs could be higher

---

[15] In section 21.9.4 we consider intra-group financing of carrots, that is, financing by the violating agents.
[16] For simplicity, we assume away the part of the parties' wealth that needs to be reserved for paying the transaction costs associated with monitoring and with the application of the carrot or stick.
[17] We assume for simplicity that the cost of effort does not reduce an individual's wealth.

or lower, which is important for incentive purposes. For example, a monitoring probability equal to 50% and a carrot equal to \$100 imply that the expected carrot is \$50. However, if the ruler has a budget equal to \$50, the individual will anticipate that the ruler will be able to pay only \$50 (not \$100), and hence his incentives will be diluted. To maintain the incentives unaltered, the ruler's budget needs to be able to cover the worst-case scenario, that is, her budget needs to be equal to the budget that would be necessary if $p_c = 1$, which in this case is \$100. Note that if the probability of monitoring were 10% and the carrot \$500, the expected cost for the ruler would be again \$50, but her budget (for incentive purposes) would need to be \$500—ten times greater. The fraction of complying individuals is also probabilistic if the ruler only knows the probability distribution of the costs of effort. A similar argument applies to errors. In essence, the ruler needs to be able to pay a carrot as if monitoring occurred with probability equal to one, the whole population complied, and there were no type-I errors. Likewise, the ruler needs to be able to apply a stick as if monitoring occurred with probability equal to one, the whole population violated the rule, and there were no type-II errors.

By setting all probabilistic variables in (6) and (7) at their worst-case-scenario level, we obtain the following constraint for carrots:

$$c^{max} = \frac{b}{k_c n} \qquad (15)$$

Similarly, with sticks, the ruler's constraint is

$$s^{max} = \frac{b}{k_s n} \qquad (16)$$

Note that the maximal sanction seems to increase with n, but this effect is illusory, as the maximal sanction is calculated at the individual level. At the population level the total sanction is nc^{max} or ns^{max}, so that the effect of n cancels out from the ruler's perspective.

In the following section we examine how these constraints can be lessened.

21.5.2 Coordinated Monitoring

To see how coordinated monitoring may increase the maximum sanction for both carrots and sticks, consider two monitoring strategies. In one case (uncoordinated monitoring), the ruler randomly decides whether to monitor a single individual by a lottery draw with probability p and then repeats the procedure for each individual. In another case (coordinated monitoring), the ruler randomly draws *pn* individuals from the population and then monitors them.[18] Clearly, the second strategy is constrained by the fact that *pn* must be a discrete number and hence only a discrete number of different probabilities are implementable. Yet, this constraint weakens as *n* grows. If $n = 1$, the only applicable coordinated-monitoring probability is $p = 1$; but if $n = 10$, coordinated monitoring admits ten different probabilities: 1/10, 2/10, . . . , 9/10, 1. As *n* grows bigger this constraint becomes less binding.

In both cases an individual faces a probability *p* of being monitored, hence, incentives are constant across the two treatments. However, from the perspective of the ruler these two strategies are markedly different, because coordinated monitoring lessens the budget constraints identified above. To see why, note that coordinated monitoring transforms a probabilistic variable into a certain number for the ruler. Monitoring each individual with probability *p* yields a distribution of possible outcomes, which includes the worst-case scenario in which (albeit with a very small probability, p^n)[19] all individuals are monitored. Instead, coordinated monitoring yields a certain number of monitored individuals, *pn*. Therefore, under coordinated monitoring the budget constraints in (15) and (16) reduce to:

$$c^{max} = \frac{b}{k_c p_c n} \quad (17)$$

and

$$s^{max} = \frac{b}{k_s p_s n} \quad (18)$$

Whenever $p < 1$, coordinated monitoring makes the constraints less binding (higher maximal sanctions). Moreover, the maximal sanction increases as the probability of monitoring decreases and does so proportionally.

Coordinated monitoring increases the maximal sanction by a factor $1 / p$ both for carrots and for sticks. This effect implies that the minimal probability of monitoring is not bounded by the ruler's budget because the ruler, by reducing the probability of

---

[18] Collective responsibility comes close to coordinated monitoring in that it spreads responsibility across the population of agents (Miceli and Segerson, 2007).

[19] To illustrate, if there are three individuals and the monitoring probability is $p = 1/10$, the probability that all three are monitored is $p^3 = (1/10)^3 = 1/1000$.

monitoring, increases proportionally the maximal sanction applicable and hence can reduce the probability of monitoring at will.[20]

21.5.3 Population Effect

Consider now two scenarios, which derive from two different interpretations of $F(e)$. If $F(e)$ is a probability distribution of the effort costs in the population, then $nF(e^*)$ is the expected number of compliers. If instead $F(e)$ describes the real distribution of effort costs in the population—that is, if $F(e)$ is the ratio of the number of individuals with effort costs equal to or less than e and the total number of individuals—then $nF(e^*)$ is the real number of compliers. The ruler knows more in the second scenario than in the first, although in both scenarios she does not know the effort costs of an individual. (The difference between these two scenarios is particularly clear if one considers a population composed of one individual.)

If the ruler knows the real distribution, then she can predict with certainty how many individuals will be rewarded. If there are no type-II errors,[21] this knowledge further lessens the budget constraint of carrots, which becomes:

$$c^{max} = \frac{b}{k_c p_c nF(e^*)} \qquad (19)$$

(If we do not have coordinated monitoring $p_c = 1$.) The same reasoning does not apply to sticks. The reason is that with sticks the ruler needs to be able to punish all individuals, not only the violators, or the threat of punishment would not be credible. To see this point in the sharpest way, consider a ruler who enforces a rule e^* on an individual with effort cost $e \leq e^*$. The individual complies. If the constraint were determined only by violators, a ruler with no budget would be able to do the job. But this cannot be the case, since a ruler with no budget cannot credibly threaten the application of (costly) sanctions. The argument applies more broadly to populations of any size: the budget has to be large enough to punish all violators and to threaten (credibly) all compliers.

---

[20] This finding is implicit in the observation made in the literature on public enforcement of law that the maximal sanction is optimal even if the punishment is imprisonment (Polinsky and Shavell, 2000).

[21] If there are type-II errors, some non-compliers will be rewarded and hence knowing the exact number of compliers does not help.

The population effect can be summarized as follows: If the ruler knows the population, carrots have to be available to reward only actual compliers, while sticks have to be available to punish all potential violators (not only the actual violators).

21.5.4 The Multiplication Effect of Sticks

The fact that sticks need to be available for all potential violators needs to be reconsidered in light of the fact that individuals might not be able to coordinate their actions. If they are able to coordinate, the analysis of the previous section applies. If they are not able to coordinate, then in fact one stick might be enough to incentivize all potential violators (Dari-Mattiacci and De Geest, 2010).[22]

To see why, consider the extreme case in which a dictator owns only one bullet, but uses this single bullet to make all seven billion world citizens live in slavery. To accomplish this, the dictator goes to the first citizen and presents him with the choice between living in slavery or being shot. The citizen opts for slavery and hence the bullet is not used; therefore, the dictator goes to the second citizen, who also opts for slavery. The dictator repeats this threat seven billion times. At the end of the day, the total effort cost of all seven billion citizens is much larger than the stick (which can destroy only one human life). For all agents taken as a group, it might be better to violate the rule and allow one of them to be punished, yet no agent has an individual incentive to sacrifice himself.[23]

A necessary condition for this multiplication effect to hold is that all agents know the order in which the ruler will monitor and penalize them. If this is the case, the availability of one extra stick (besides those already used up on individuals who have higher costs of effort than the rule requires) is enough to incentivize the first individual who will be monitored. But if the first individual complies, then the stick is not used up and will provide incentives to the second individual and so forth. At the limit, one stick is enough for any number of compliers. Therefore, (18) can be rewritten as follows:

---

[22] Dari-Mattiacci (2009) employs the multiplication effect to show that tort liability (a stick) provides stronger incentives than restitution (a carrot).

[23] Some readers, and even we, might disagree with the value of life implied by the example, but this is not the point here.

$$s^{max} = \frac{b}{k_s p_s [n(1 - F(e^*) + 1)]} \qquad (20)$$

Comparing the new ruler's constraint under sticks with the constraint under carrots in (19), we have that carrots are preferred if

$$n < \frac{k_s}{k_c}\frac{p_s}{p_c}\frac{1 - F(e^*) + \frac{1}{n}}{F(e^*)} \qquad (21)$$

The three terms on the right-hand side of the inequality provide a balance of factors that favor carrots or sticks. Let us start with the simplest scenario in which the costs of carrots and sticks are the same ($k_c = k_s$) and they are applied with the same probability ($p_c = p_s$). Furthermore, let us disregard for simplicity the term $1 / n$ as it becomes very small for large n. Then, the condition in (21) says that the rulers' constraint is less binding with carrots if $n < \frac{1 - F(e^*)}{F(e^*)}$, which is the ratio of violators to compliers.[24] This reveals two general principles. First, the multiplication effect of sticks is stronger (giving a stronger advantage to sticks over carrots) as n grows, since the condition becomes more difficult to satisfy. Second, this advantage also grows if the number of compliers increases. In normal circumstances, compliers outnumber violators and hence the condition becomes $n < 1$, so that carrots always yield a more stringent budget constraint for the ruler than sticks do.[25]

The reason why only sticks have a multiplication effect is as follows. Carrots are applied upon compliance, sticks upon violation. If the agent complies, the carrot is used up but the stick is not. Although a stick can be applied only once, the threat to apply the stick can be repeated several times.

Note that the multiplication effect is another qualification of the incentive equivalence result: a \$100 stick can, under some conditions, have an incentive effect of

[24] Note that this condition can never be satisfied if $F(e)$ is the real distribution of effort in the population.
[25] The same results are obtained if one does not drop the term $1 / n$ from (21). Moreover, the result can be easily generalized to situations in which coordination costs are not fixed (either zero or prohibitive as in our example) but rise with the number of individuals involved in the coordination. In this case, the multiplication effect works as long as there are enough sticks available to punish the group of coordinators. As coordination costs rise, the number of individuals who can be feasibly involved in coordination decreases and so does the number of sticks that need to be available. If coordination costs rise very sharply, the analysis collapses to the basic case described in the text.

more than $100; a $100 carrot, by contrast, can never have an incentive effect of more than $100.

The upside of the multiplication effect is that it makes law enforcement more effective: it allows governing a country of 300 million citizens with relatively few prison cells—it is sufficient that there are some empty prison cells.[26] The downside of the multiplication effect is that sticks carry an inherent risk of abuse, which is absent in carrots.

21.6 Distributional Effects

The primary goal of carrots and sticks is to make an agent follow a certain rule. Nonetheless, carrots and sticks may have significant distributional side effects. Suppose a principal wants to make an agent exert an effort of $90. If the principal threatens the agent with a $100 stick, the agent will comply and become $90 poorer than before the rule. If the principal promises a $100 carrot instead, the agent will also comply but he will become $10 wealthier than before the rule (by receiving $100 compensation for a $90 effort).

21.6.1 Sticks Undercompensate, Carrots May Overcompensate

Sticks inherently undercompensate because the agent is always worse off compared to the status quo: either he incurs an effort cost or a penalty. (Note that *indirectly* the agent may receive compensation under a stick regime in the form of receiving the benefit of rule compliance by the other agents). Carrots, by contrast, have a built-in compensation mechanism in that they always allow the agent to opt for the status quo, in which the agent does not do the effort and simply does not receive the carrot.

While carrots can never be undercompensatory (because then they would violate the incentive compatibility constraint), the question remains why they would ever be overcompensatory. Why cannot the principal set the carrot equal to the effort cost, so that there is no overcompensation?

A first reason is that the principal may have imperfect information on the individual effort cost. In this case, the carrot may be set higher than strictly required to

---

[26] Bar-Gill and Ben-Shahar (2009) explain that prosecutors are able to strike plea-bargaining deals with most defendants despite their limited resources because they can threaten each individual defendant to go to trial.

induce performance; the agent's rent is then essentially an information rent. This is especially the case if agents have varying effort costs and the principal knows only the distribution $F(e)$ of the cost (but not the individual costs). In this case, agents with a relatively low effort receive relatively high rents.

A second reason why carrots may be overcompensatory is that the principal makes monitoring errors by mistakenly concluding that some violating agents complied. Since violation leads to an (even small) expected rent, compliance needs to include this expected rent too in order to give sufficient incentives to comply. Note that errors create fixed rents or impose fixed costs that are equal for all individuals in expected terms; therefore, the only distributional effects *within* the group of agents derive from differences in effort costs.

With carrots, compliers and violators earn:

$$\begin{array}{l} p_c(1-\varepsilon_I)c - e \geq p_c\varepsilon_{II}c\left(compliers\right) \\ p_c\varepsilon_{II}c\left(violators\right) \end{array} \qquad (22)$$

The pay-offs from complying and violating are the same only for individuals with $e = e^*$. Given that the optimal carrot is such that $e^* = p_c(1-\varepsilon_I-\varepsilon_{II})c$ and compliers have $e \leq e^*$, the pay-off of compliers is greater than zero and increases as e decreases, as indicated by the first inequality. Moreover, all individuals earn a rent equal to $p_c\varepsilon_{II}c$. The intuition is that, due to type-II errors, some violators may earn a carrot. To restore incentives, the pay-off of compliers needs to be increased accordingly, thereby generating a rent for compliers equal (at the margin) to the fraction of carrots paid to violators. Moreover, since the carrot is the same for all individuals, compliers below the margin earn an even greater rent, which increases as their effort cost decreases.

With sticks, compliers and violators pay:

$$\begin{array}{l} p_s\varepsilon_I s + e \leq p_s\varepsilon_I s + e^*\left(compliers\right) \\ p_s(1-\varepsilon_{II})s = p_s\varepsilon_I s + e^*\left(violators\right) \end{array} \qquad (23)$$

Again, the pay-offs from complying and violating are the same only for individuals with $e = e^*$. Given that the optimal stick is such that $e^* = p_s(1-\varepsilon_I-\varepsilon_{II})s$ and compliers have $e \leq e^*$, the cost for compliers is greater than zero and increases in e up to $p_s\varepsilon_I s + e^*$. The intuition is that, due to type-I errors, some compliers are punished. To restore incentives, the expected stick has to be raised above e^*. These greater sticks also fall onto violators, generating a fixed cost for all individuals, which is greater than e^* if there are type-I errors.

21.6.2 General Carrots and Sticks Make Compliers with a High Effort Cost Relatively Poorer Compared to Compliers with a Low Effort Cost. This Effect Can Be Removed Through Individualized Carrots but Not Through Individualized Sticks

General carrots make compliers with lower effort cost richer. If the principal knows the individuals' effort costs, carrots can be made individual-specific, so that each individual receives a carrot that is just enough to incentivize him. This removes the distributional effects that were caused by the general carrot, both with respect to non-agents—that is, individuals not subject to enforcement—and agents.

Remarkably, individualizing sticks does not make their distributional effects disappear. To illustrate, suppose that agent A has an effort cost of $80 while agent B has an effort cost of $90. Instead of threatening all agents with a $100 stick, the principal now threatens agent A with a $81 stick and agent B with a $91 stick. Both agents comply. At the end of the day, compliance still made agent A $80 poorer and agent B $90 poorer. The reason why the distributional effects of individualized sticks are the same as of general sticks is that sticks are meant not to be applied. If they are not applied, their magnitude cannot change the distribution.

21.7 Activity-Level Distortions Caused by Distributional Effects

We have seen that carrots and sticks have different distributional effects. These distributional effects may in turn cause activity-level effects. Stick regimes naturally undercompensate and therefore also tend to violate the participation constraints of agents. Carrots, on the other hand, are either fully compensatory or overcompensatory. In the latter case, they generate a rent for the participating agent. Therefore, carrots and sticks also generate different incentives for entry into the regulated activity.

Carrots attract entry of compliers, and incentives to enter increase as e decreases. If there are type-II errors, carrots generate a fixed rent from entry for both compliers and violators. Sticks discourage entry of compliers and violators. The complier's incentives not to enter increase as e increases. If there are type-I errors, sticks generate a fixed cost of entry for both compliers and violators.

Because of these activity-level effects, Wittman (1984) argued that, as a rule of thumb, carrots need to be used for positive externalities and sticks for negative externalities. Indeed, the overcompensatory or undercompensatory effects can generate desirable activity-level effects when there are positive or negative externalities associated with the activity (Dari-Mattiacci, 2009).

21.8 The Principal's Opportunism and the Agent's Distorted Incentives to Self-Report Caused by the Distributional Effects

21.8.1 The Principal's Incentives to Behave Opportunistically

So far we have assumed that after the agent has complied, the principal enforces the rule as announced, that is, she pays the promised carrot, or she does not apply the stick. In practice, however, a principal may behave opportunistically for two reasons. First, under a carrots regime, the principal may want to keep the bonus for herself. Second, under a sticks regime, the principal may falsely declare that the agent violated the rule, in order to receive the proceeds of the fine.

More formally, assume the following timeline: at t_0, the rule is announced, at t_1 the agent complies with or violates the rule, at t_2 the principal monitors and applies the carrot or stick. As always, opportunism is caused by the timing of the actions; the principal is that party who acts last and who therefore may benefit from not acting as announced.

It is nonetheless important to see the relationship between this opportunism and the distributional effects of carrots and sticks. Carrots increase the agent's wealth, sticks increase the principal's wealth (if the principal receives the proceeds of the fines). Therefore, the principal may not want to apply the carrots and may want to apply the sticks.

Because of this danger of opportunism, both carrots and sticks require commitment by the principal. If the principal cannot commit, the agent will violate the rule, either because he expects not to receive the carrot at t_2 or because he expects to be punished at t_2 even if he complies at t_1.

But opportunism is symmetric here: it exists irrespective of whether carrots or sticks are used. Yet the symmetry may disappear if special variants of carrots or sticks are used, as will be explained below.

21.8.2 The Agent's Incentives to Self-Report

Agents have a natural incentive to self-report compliance under a carrots regime. Indeed, if they are found complying, they receive the carrot. By contrast, agents do not have natural incentives to self-report a violation under a sticks regime since, if they are found violating the rule, they receive a penalty.[27]

Another way of looking at this is that, in a probabilistic enforcement regime, agents have an incentive to manipulate the probability of monitoring p. Under a carrots regime, complying agents try to increase p; under a sticks regime, violating agents try to decrease p.

Principals can use these features of carrots to increase p when p is naturally low, for instance because the behavior is to a large extent non-verifiable without the cooperation of the agent.[28]

21.9 Special Types of Carrots and Sticks

21.9.1 Pre-Compensated Carrots and Sticks

The distributional effect of sticks and carrots can be cancelled out through transfer payments. Stick regimes, which are naturally unattractive for agents, can be made attractive by offering agents an entry fee. Carrot regimes that generate rents for agents can be made less attractive to agents by requiring them to pay an entry fee; as long as the entry fee does not exceed the rents, the participation constraint of the agent is still met.[29]

---

[27] Incentives for self-reporting in a stick regime can be created by leniency programs that annul the punishment for those who report themselves to the authorities. Such annullable sticks, however, can be seen as precompensated carrots. See section 21.9.2.

[28] Tabbach (2010) shows that when manipulating p is costly for the agent (avoidance) it might be desirable that agents try to avoid punishment.

[29] An example can be found in Becker and Stigler (1974), who suggested to pay efficiency wages (which are technically annullable bonuses) to police officers in order to give them an incentive not to be corrupt, and to let them pay a fee for the right to become a police officer in order to cancel out the rents.

When the distributional effects are removed in this way, the activity-level distortions caused by the distributional effects are removed as well.

In practice, however, the legal system may forbid entry fees, especially when they are paid by employees. One concern is indeed that entry fees (paid by the agent) give the principal an additional incentive to be opportunistic. Similarly, entry bonuses paid to the agent (at t_0) give the agent an incentive to be opportunistic.

21.9.2 Annullable Carrots and Sticks

Annullable carrots are carrots that are paid *unless* the agent is monitored and found violating. They differ from normal carrots, which are paid *if* the agent is monitored and found complying. Similarly, *annullable sticks* are sticks that are applied *unless* the agent has been monitored and found complying. In essence, an annullable carrot is a threat to take back, an annullable stick a promise to give back (De Geest, Dari-Mattiacci, and Siegers, 2009).

Annullable carrots are mathematically identical to normal carrots if the probability of monitoring is 100%. Indeed, the difference between the normal and the annullable variants consists in what happens in case of no monitoring. If all agents are monitored all the time (p = 1), the difference disappears. Consider the case in which an employee receives a bonus unless he is underperforming to the case in which the employee receives the bonus only if he is found performing. If monitoring takes place with certainty, the employee will receive a bonus if he complies and no bonus if he violates in both scenarios. Yet there is an important difference if the probability of monitoring is less than 100%. In those cases, violating employees who were lucky enough not to be monitored still receive a bonus under annullable carrots.

An example of annullable carrots can be found in the efficiency wages literature (Shapiro and Stiglitz, 1984; Becker and Stigler, 1974). Here it has been argued that overpaying an employee can make sense because it gives the employee an extra incentive not to be fired. Indeed, the overpayment can be seen as a carrot that is paid unless the employee is monitored and found shirking. This extra incentive not to shirk may allow the employer to lower the monitoring levels and save on monitoring costs.

This indicates that annullable carrots have a source of ineffectiveness that is absent in regular carrots. Regular carrots are only paid to agents who have been monitored (and found complying). Annullable carrots are paid not only to agents who have been monitored (and found complying) but also to agents who have not been monitored (and who may have complied or shirked). Paying a bonus to an agent who has not been monitored obviously has no incentive effect, because incentives are generated by the wedge between what compliers and violators receive. Therefore, a carrot paid to non-monitored agents is, from the viewpoint of the employer, "wasted money."

This wasted money may indirectly lead to inefficiency. Indeed, a profit-maximizing employer faces a trade-off between paying rents and bearing monitoring costs: the higher the overpayment, the lower the monitoring costs can be. But if rents have to be paid also to the agents who are not monitored, the employer will adopt inefficiently high monitoring levels in order to cut the frequency with which such rents are paid. This monitoring-level distortion is not present in a normal carrots regime, in which the bonus is only paid to monitored, complying agents (De Geest, Dari-Mattiacci, and Siegers, 2009).

If annullable carrots are intrinsically wasteful, why would rational principals ever use them? The answer is that the annullable variant changes the timing of the payments and therefore also the problem of opportunism. More specifically, annullable carrots give the principal an incentive to monitor as promised. In contrast, under normal carrots, the employer may *ex post* (after the agent has performed) decide not to monitor in order not to have to pay the agent a bonus. As a result, annullable carrots (such as efficiency wages) can be rational choices when the principal cannot credibly commit to paying carrots with a certain probability.[30]

Another way of looking at annullable carrots is to see them as fully pre-compensated sticks. Indeed, agents first virtually receive full pre-compensation c and virtually pay back c (that is, pay s) if found violating. Since the lump sum payment is sunk, this suggests that annullable carrots will have many of the characteristics of normal sticks. And indeed, they have the same risk properties (they create risk for violators), and

---

[30] The reason why the principal may offer a device that protects the agent against opportunism by the principal is that the principal must meet the participation constraint of the agent. Note that efficiency wages do not prevent opportunism consisting of falsely declaring that the employee shirked or flat-out refusing to pay the bonus to monitored non-shirkers. But the legal system may be able to observe such forms of opportunism, while opportunism consisting in lowering the probabilities of monitoring may be non-verifiable.

treat non-monitored agents in the same way as monitored compliers—just like normal sticks do. Similarly, annullable sticks have the risk properties of normal carrots.

Yet the lump sum payment changes some of the distributional distortions. While normal sticks usually underpay (that is, they underpay unless there is sufficient in-kind compensation in the form of the received benefit), annullable carrots do compensate the agents for their effort costs—as a matter of fact, they have to pay the same overcompensation to monitored compliers as normal carrots. But while normal carrots pay no compensation to non-monitored agents (and therefore correctly compensate in expected terms, at least if effort costs are identical), annullable carrots pay the same overcompensation to non-monitored agents. Therefore, their distributional characteristics are different from those of normal sticks and carrots. Similarly, the distributional effects of annullable sticks are greater than those of normal sticks.

The annullable variants also generate more transaction costs, though the analysis depends on whether or not the original payment and the annihilation of that payment are set off against each other. Without set-off, the annullable variants clearly generate more transaction costs than their normal variants: they require a lump sum transaction in 100% of the cases and on top of that there will be an annihilation transaction for some of them. But even with set-off (where the lump sum pre-compensation is only virtually paid beforehand, but effectively paid *ex post*, in order to allow for transaction cost-saving set-offs) annullable carrots and sticks generate higher transaction costs because they also require transactions with non-monitored agents.[31]

Since annullable carrots can be seen as pre-compensated sticks, they will also make sense in most cases in which sticks make sense. For instance, when only shirking can be proven, (normal) sticks can work but (normal) carrots cannot. In this case, annullable carrots can work as well. An example is corruption, where typically the only parties who have information on the corrupt act are the two parties involved (the bribed agent and the bribing third party). Therefore, it may be possible to prove corruption only

---

[31] This can be illustrated with a numerical example. Suppose 10% monitoring probability and 20% violators. In this case a carrot is paid to 8% of the agents, a stick is paid by 2% of the agents, an annullable carrot without set-off requires payment to or from 100% + 2% of the agents (100% receive the carrot, and 2% have to give it back), an annullable carrot with set-off requires payment to 98% of the agents, an annullable stick without set-off requires payment to or from 100% + 8% of the agents, and an annullable stick with set-off requires payment to 92% of the agents.

when the third party betrays the agent or when the principal happens to acquire the relevant information. Since the quality of the information is such that monitoring can only show that an agent is corrupt, but never reveal that the agent is honest, the fact that an agent has not been found to be corrupt during a monitoring session does not necessarily prove that he is honest.

21.9.3 Combined Carrots and Sticks

The combined use of carrots and sticks increases the potential maximum sanction/reward and therefore may create stronger incentives than their single use. Indeed, the theoretical maximum of the sanction/reward under the combined use now becomes the wealth of the agent plus the total wealth of the principal. The combined use, however, also leads to higher transaction costs because a transfer is needed both in case of compliance and in case of violation. In addition, the combined use also changes the risk properties, by creating risks for both compliers and violators.

21.9.4 Intra-Group Financed Carrots and Intra-Group Redistributed Sticks

So far we have assumed that it is the principal who both finances carrots and receives the proceeds of the sticks. Now consider the case in which it is the violating agents who finance the carrots for compliers, or the complying agents who receive the sticks paid by violators.

An interesting question in such a regime is what happens when all agents comply. Since no agent is left to finance the carrots, all complying agents may have to finance the carrot, in which case no complier receives a net carrot. Similarly, if all agents violate, the proceeds of the fines may have to be redistributed to all agents, in which case no violator pays a net stick. But the fact that no agent pays or receives anything in this equilibrium does not mean that there are no incentives to comply. Indeed, what matters for incentives is that the first agent to comply or violate faces a change in pay-offs—which is the case here.

This intra-group redistribution of carrots and sticks has three benefits. First, it removes the principal's incentive to behave opportunistically by making the principal a neutral rather than a financially interested party.[32] Second, for the reasons explained in the previous paragraphs, there are no transaction costs in the equilibrium in which all agents comply or violate, which is a remarkable property for a regime that has an (explicit or implicit) carrot component. Third and foremost, intra-group redistribution strengthens incentives by creating combined sanctions—carrots for the compliers and sticks for the violators.

Yet, the marginal incentives now become a function of the behavior of the others. Consider first a stick system with intra-group redistribution of the proceeds of the stick (and in which individual sticks remain constant). The more agents violate, the more attractive it becomes to comply because the stick proceeds are both larger and have to be split among fewer compliers; if all others violate, the incentives to comply are at their strongest. This means that such a regime can be particularly effective for weakest-link offenses, such as cartels, in which it is sufficient that one individual cooperates with the competition authority.

Contrast this to a carrot system with intra-group redistribution (assuming again that the individual carrots remain constant). The more agents cooperate, the less attractive it now becomes to violate because the total carrots are both larger and need to be financed by fewer violators; if all others comply, the incentives to comply are the strongest. Therefore, such a regime is most effective for rules that need to be followed by all agents.

21.9.5 Reversible Carrots

Reversible rewards (Ben-Shahar and Bradford, 2012) are amounts of money that will either be used to reward the agent (if he complies with the rule) or to finance his punishment (if he violates the rule). The rewards are "reversible" because they are reversed into stick enforcement funds if the agent violates the norm.

Reversible rewards are a special type of combined sanctions. One special feature is that the reward is not combined with a stick of the same magnitude but with a stick enforcement budget of the same magnitude. This budget may be lower or higher than the stick it enforces. Therefore, the incentive effect may be more (or less) than twice the incentive effect of the reward alone. Another feature is that the budget the principal is

---

[32] It often occurs in firms, organizations, and commitment contracts that penalties are paid to third parties (Ayres, 2010).

going to spend is independent of the choice of the agent. This fixed character of the amount makes it easier for the principal to pre-commit to spending it (for instance by transferring the amount to a fund controlled by a third party). This pre-commitment reduces the danger of *ex post* opportunism by the principal, which in turn makes the scheme more credible and effective.

Reversible rewards can be useful when the enforcement costs associated with sticks are significant but cannot be financed through the proceeds of the stick (for instance because the stick is non-monetary). An example may be the economic boycott of another country, which is often as costly to the boycotter as to the boycotted.

21.9.6 Strict Liability Carrots and Sticks

So far our analysis has focused on fault-based rules, in which the principal specifies good or bad behavior, and applies sanctions, possibly with errors, after observing good or bad behavior.

Now consider strict liability rules, which measure only output but do not specify optimal input. Such rules can make sense when specification problems become too large because effort costs are insufficiently verifiable, while output is still sufficiently verifiable.

Under strict liability, sticks may also have to be paid upon "compliance." Indeed, even if the agent reduced the output (for instance pollution) to the optimal level (by taking optimal care), there is still a "sanction" applied as long as the output is not zero. As a result, sticks may lose a comparative advantage—that they do not have to be applied upon compliance. Moreover, "compliers" (who can be defined here as those agents who do what is optimal from the perspective of the principal) still face a risk (though a lower one than "violators") if monitoring of output is probabilistic.

Another difficulty with strict liability is that defining carrots or sticks requires defining a baseline at which zero has to be paid. To illustrate, suppose that carrots would be used to reduce pollution. At what point should carrots start to be given? At the maximum possible level of pollution? But this may be infinite. This is why Lazear (1991) suggested (in the context of bonuses or penalties for workers) that sticks should be used if the maximum performance can more easily be defined, and carrots when the minimum performance can more easily be defined.

21.10 Extensions

21.10.1 Political Risks

So far, we have assumed that the principal chooses the optimal type of enforcement mechanism. However, when the principal is a political institution, public choice mechanisms may distort the decision-making process. Carrots (in the form of subsidies) may be more distortive in this respect, because well-organized groups may benefit more from obtaining a carrot for their own group than from obtaining a stick (Galle, 2012).

21.10.2 Behavioral Economics

Behavioral biases may further complicate the analysis. For instance, when agents have loss aversion (Kahneman and Tversky, 1979), negative incentives may be more effective than positive incentives. Or when agents are imperfectly aware of the rule, carrots may become relatively more effective because they tend to be more salient (Galle, 2014).[33]

Some findings in psychology seem to favor the use of carrots. For instance, when behavior is followed by a positive consequence ("reinforcement"), it tends to be repeated; when it is followed by a negative consequence, it tends to be avoided (Thorndike, 1913; Ferster and Skinner, 1957). These results are in line with the activity-level distortions caused by the different distributional effects of carrots and sticks. Similarly, the general tendency among psychologists to be more favorable toward carrots than toward sticks (see Kohn, 1993) is compatible with our observation that, in an increasingly complex society, parents and employers increasingly face specification problems.

There is a large literature in experimental economics studying carrots or sticks, but only a relatively small fraction of it deals with the choice between carrots and sticks.[34] Most of the effects that we identify have not been tested empirically. Finally, a related literature studies the emergence of preferences for punishing or rewarding in an evolutionary model (Herold, 2012).

---

[33] More specifically, Galle (2014) considers a trade-off between moral hazard and risk-spreading and argues that if government can control the salience of carrots, it may be able to mitigate some of the moral hazard carrots usually present. This would shift the optimal mix of instruments farther towards carrots.

[34] A non-exhaustive list includes Abbink, Irlenbusch, and Renner (2000); Offerman (2002); Andreoni, Harbaugh, and Westerlund (2003); Brooks, Stremitzer, and Tontrup (2012); Nosenzo et al. (2016).

21.11 Conclusion: The Optimal Choice Between Carrots and Sticks

Overall, should carrots or sticks be used? Sticks have an intrinsic advantage over carrots: they do not have to be applied if the agent complies. Moreover, probabilistic sticks create risks for violators rather than for compliers, which further increases their effectiveness when agents are risk-averse. Because of these advantages, sticks will be superior in simple settings, in which the principal has sufficient information about the agent's effort cost to make sure that the agent should and will comply.

Yet, carrots may be superior in two cases (De Geest and Dari-Mattiacci, 2013). The first is when the principal faces specification problems, that is, when she does not know what to expect from each individual citizen because she does not know their effort costs. In those cases, sticks may punish agents who are unable to comply with the norm at reasonable costs, which leads to higher transaction costs, risk costs, and undesirable wealth effects. Specification problems may explain why carrots (in the form of copyrights) are used to incentivize composing music (the legal system does not know which individuals are able to do this) or rescuing at sea (the legal system does not know which part of the cargo of a sinking ship should be rescued in emergency situations with time constraints).

The second case in which carrots may be superior is when a significantly higher effort is required from some individual agents than from others (for instance, when only some families need to send a family member to the army, or only some families to sacrifice land for a highway project). In such cases of singling-out, sticks would cause significant unintended distributional distortions (impoverishing those from whom much is required). This may help to explain why carrots are increasingly used in complex societies. When agents specialize, specification and singling-out problems increase.

Acknowledgements

We thank Brian Galle, Theo Offerman, Alexander Stremitzer, and Donald Wittman for comments. We also thank Megan Lasswell for research assistance.

References

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner (2000). "The Moonlighting Game: An Experimental Study on Reciprocity and Retribution." *Journal of Economic Behavior & Organization* 42: 265–77.

Andreoni, James, William Harbaugh, and Lise Westerlund (2003). "The Carrot or the Stick: Rewards, Punishment, and Cooperation." *American Economic Review* 93: 893–902.

Ayres, Ian (2010). *Carrots and Sticks: Unlock the Power of Incentives to Get Things Done*. New York: Random House.

Bar-Gill, Oren and Omri Ben-Shahar (2009). "The Prisoners' (Plea Bargaining) Dilemma." *Journal of Legal Analysis* 1: 737–773.

Becker, Gary S. (1968). "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76: 169–217.

Becker, Gary S. and George Stigler (1974). "Law Enforcement, Malfeasance, and Compensation of Enforcers." *Journal of Legal Studies* 3: 1–18.

Ben-shahar, Omri and Anu Bradford. (2012). Reversible Rewards. *American Law and Economics Review*, 15: 156–186.

Brooks, Richard R. W., Alexander Stremitzer, and Stephan Tontrup (2012). "Framing Contracts: Why Loss Framing Increases Effort." *Journal of Institutional and Theoretical Economics* 168: 62–82.

Dari-Mattiacci, Giuseppe (2009). "Negative Liability." *Journal of Legal Studies* 38: 21–60.

Dari-Mattiacci, Giuseppe (2013). "Slavery and Information." *Journal of Economic History* 73: 79–116.

Dari-Mattiacci, Giuseppe and Gerrit De Geest (2010). "Carrots, Sticks, and the Multiplication Effect." *Journal of Law, Economics, and Organization* 26: 365–384.

De Geest, Gerrit and Giuseppe Dari-Mattiacci (2013). "The Rise of Carrots and the Decline of Sticks." *University of Chicago Law Review* 80: 341–392.

De Geest, Gerrit, Giuseppe Dari-Mattiacci, and Jacques J. Siegers (2009). "Annullable Bonuses and Penalties." *International Review of Law and Economics* 29: 349–359.

Ferster, Charles B. and Burrhus Frederic Skinner (1957). *Schedules of Reinforcement*. Englewood Cliffs, NJ: Prentice-Hall.

Galle, Brian D. (2012). "The Tragedy of the Carrots: Economics and Politics in the Choice of Price Instruments." *Stanford Law Review* 64: 797–850.

Galle, Brian D. (2014). "Carrots, Sticks, and Salience." *Tax Law Review* 66: 53–110.

Garoupa, Nuno and Matteo Rizzolli (2012). "Wrongful Convictions Do Lower Deterrence." *Journal of Institutional and Theoretical Economics* 168: 224–231.

Herold, Florian (2012). "Carrot or Stick? The Evolution of Reciprocal Preferences in a Haystack Model." *American Economic Review* 102: 914–940.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47: 263–291.

Kohn, Alfie (1993), *Punished by Rewards*, Boston: Houghton Mifflin.

Lando, H. (2006). "Does Wrongful Conviction Lower Deterrence?" *Journal of Legal Studies* 35: 327–338.

Lazear, Edward P. (1991). "Labor Economics and the Psychology of Organizations." *Journal of Economics Perspectives* 5: 89–110.

Leshem, Shmuel and Avraham Tabbach (2016). "Solving the Volunteer's Dilemma: The Superiority of Rewards over Punishments." *American Law and Economics Review* 18: 1–32.

Miceli, Thomas and Kathleen Segerson (2007). "Punishing the Innocent along with the Guilty: The Economics of Individual versus Group Punishment." *Journal of Legal Studies* 36: 81–106.

Nosenzo, Daniele, Theo Offerman, Martin Sefton, and Ailko van der Veen (2014). "Encouraging Compliance: Bonuses Versus Fines in Inspection Games." *Journal of Law, Economics, & Organization* 30: 623–648.

Nosenzo, Daniele, Theo Offerman, Martin Sefton, and Ailko van der Veen (2016) "Discretionary Sanctions and Rewards in the Repeated Inspection Game." *Management Science* 62: 512–517.

Offerman, Theo (2002). "Hurting Hurts More Than Helping Helps." *European Economic Review* 46: 1423–37.

Png, I. P. L. (1986). Optimal Subsidies and Damages in the Presence of Judicial Error. *International Review of Law and Economics*, 6: 101–105.

Polinsky, A. Mitch and Steven Shavell (2000). "The Economic Theory of Public Enforcement of Law." *Journal of Economic Literature* 38: 45–76.

Rizzolli, M. and M. Saraceno (2013). "Better that Ten Guilty Persons Escape: Punishment Costs Explain the Standard of Evidence." *Public Choice* 155: 395–411.

Shapiro, Carl, and Joseph E. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74: 433–444.

Tabbach, Avraham (2010). "The Social Desirability of Punishment Avoidance." *Journal of Law, Economics and Organization* 26: 265–289.

Thorndike, Edward Lee (1913). *Educational Psychology*. New York: Teachers College Press.

Wittman, Donald A. (1984). "Liability for Harm or Restitution of Benefit?" *Journal of Legal Studies* 13: 57–80.