# Exploring Mechanisms of Institutional Design wich Promote Compliance*

Jordi Tena-Sánchez

Universitat Autònma de Barcelona
jordi.tena@uab.cat

## Abstract

This paper explores institutional design strategies that promote compliance by regulated agents. Without a doubt, this is one of the central topics in contemporary political theory. Many illustrious traditions in political theory have held that institutions should be designed in such a way that they are capable of ensuring compliance of a nation of devils. This paper argues that such a viewpoint is erroneous and that effective regulatory institutions must assume virtue but must also be prepared for non-compliance; a view shared by some of the most renowned contemporary authors in the field of institutional design. In what follows, the contributions of scholars such as P. Pettit, I Ayres, J. Braithwaite, V. Braithwaite, T. Makkai, N. Gunningham, or P. Grabosky, among others, are reviewed. The possible problems and limitations that may arise with these proposals are then discussed and future lines of work suggested.

**Keywords:** Institutional Design; Compliance; Deviant-Centred and Complier-Centred Strategies; Enforcement and Strengths-Based Pyramids.

**Resumen:** *Explorando mecanismos de diseño institucional que promuevan el cumplimiento*

En el presente trabajo se exploran estrategias de diseño institucional que promuevan el cumplimiento por parte de los agentes regulados. Éste es sin duda uno de los temas centrales en las ciencias sociales contemporáneas. Desde algunas ilustres tradiciones de teoría política se ha sostenido que las instituciones deben diseñarse de modo tal que sean capaces de asegurar el cumplimiento de un pueblo de demonios. En el presente trabajo se sostiene que dicho punto de vista es erróneo y que las instituciones reguladoras efectivas deben asumir la virtud pero estar preparadas para el incumplimiento. Éste es el punto de vista de algunos de los autores contemporáneos más notables en el terreno del diseño institucional. En lo que sigue, entre otras, se revisan las aportaciones de escolares como P. Pettit, I Ayres, J. Braithwaite, V. Braithwaite, T. Makkai, N. Gunningham o P. Grabosky entre otros.

Posteriormente, se discuten algunos posibles problemas y limitaciones de este tipo de propuestas y se señalan algunas líneas para el desarrollo futuro de las mismas.

**Palabras clave:** Diseño Institucional; Cumplimiento; Estrategias Centradas en el Cumplimiento; Estrategias Centradas en la Desviación; Pirámides de Aplicación y Pirámides Basadas en Fortalezas.

## Introduction

The aim of this paper is to explore institutional design strategies which promote compliance. More specifically, this paper will attempt to identify designs for laws, policies, regulations, by-laws and formal rules in general, aimed at either promoting or eradicating certain kinds of behaviour with a view to ascertaining which of these designs achieves the highest levels of compliance.

The design of institutional strategies for promoting compliance is one of the most important topics in modern political theory and closely linked to many other debates within the discipline such as justice or governance. While it is unquestionable that factors such as network governance, citizen participation or fairness in decision-making processes are fundamental to compliance, this paper focuses exclusively on the impact that design of institutional mechanisms has on compliance.

The paper has two parts: the first one offers an overview of existing literature; the second part takes sides on that matters. In line with the specialised literature, a distinction is made between institutional design strategies centred on sanctions, deviance or deterrence, and strategies centred on compliance. While the former are characteristic? for conceiving of agents as villains, the latter assume virtue[1] but are prepared for non-compliance. In what follows, some problems that arise with the first type of strategies are discussed, while an overview is made of some of the most important contributions to the debate on the second type. Specifically, the contributions of authors such as P. Pettit, I Ayres, J. Braithwaite, V. Braithwaite, T. Makkai, N. Gunningham, or P. Grabosky, among others, are reviewed. Possible problems and limitations to proposals of this kind are then discussed and future lines of work suggested. The paper concludes with a summary of the main arguments.

---

1. I use "virtuous" and "villainous", freely, to make reference to comply and deceive. However, certainly, sometimes virtue may ask us to break the law, etc. I leave these complications aside.

## Institutional design strategies

### *Compliance and sanction in classic texts*

The main works of reference concerning the institutional design strategies applied by relevant organisations focus on drawing up ideal types (Grabosky and Braithwaite, 1986). This approach will also be taken in this paper. Generally speaking, these types are normally drawn up on the basis of their position in the formality continuum suggested in the seminal work by Black (1976) (Grabosky and Braithwaite, 1986). Once again, this approach will also be taken in this paper. Some of the classic works in this area include those of Reiss (1984) and Hawkins (1984). These authors draw a distinction between deterrence (Reiss) or sanction (Hawkins) strategies and compliance strategies. In general terms, deterrence or sanction strategies seek to detect infringements of the law by means of surveillance and control in order to subsequently apply formal penalties which, in principle, should bring the punished back onto the path of compliance. On the other hand, compliance strategies reflect a more informal regulatory style focused more on persuasion (Grabosky and Braithwaite, 1986). This basic distinction between complier-centred strategies and strategies centred around deterrence or sanction is also followed in this paper.

### *P. Pettit and deviant-centred and complier-centred strategies*
P. Pettit (1996, 1997, 2002) draws a distinction between *deviant-centred strategies* and *complier-centred strategies*.

### *Deviant-centred strategies*
Illustrious traditional thinking views human beings as intrinsically corrupt or, at least, argues that this should be our starting premise when setting out to design institutions. This approach is common in institutional design and serves as a basis for deviant-centred strategies (Pettit, 1996, 1997, 2002). It is assumed that since selfishness may lead an individual to behave badly, institutions must be designed in such a way that compliance becomes the rational choice. In an ideal situation, suitable motivators would have to be sought for each individual but, given that this is impossible in a world such as ours, we must look for appropriate sanctions (both positive —like rewards— and negative —like punishments) that will at least motivate the most selfish, most shameless individual. A strategy of this nature would, therefore, offer a far higher dose of motivation than required by most individuals. The problem with this strategy is the image which it projects of its targets, who are portrayed as veritable villains. In fact, it is not the penalties but the rewards that may be perceived as offensive by the people they are aimed at, who could become demoralised and feel their virtue undermined (Ayres and Braithwaite, 1992, Bardach and Kahan, 1982, Pettit, 1996, 1997, 2002). A strategy of this type could therefore give rise to numerous counter-productive effects. In the following section, some of the most common of these effects are reviewed briefly.

Counter-productive effects of deviant-centred strategies

In a reasonably civic context, human beings generally comply with the requirement to cooperate in a more or less unquestioning manner. Indeed, certain options do not even enter into our range of alternatives. For instance, when we are at the public library browsing through the latest edition of the reference manual on rational choice theory and, in an unguarded moment, we are seized with the temptation to steal it, we will not even bother to assess the probability of taking it without being caught. The mere fact that this idea has even crossed our mind will probably make us feel guilty and think badly of ourselves.

However, the arrival on the scene of significant rewards or punishments for certain acts may mean that options that were originally excluded now become part of our range of available choices. Such an event could cause us to move from a state of non-egocentric deliberation to one of egocentric deliberation (Pettit, 1996, 1997, 2002). If, for example, my salary is dramatically increased by my employer, I might think that up to now I had been working for much less than I deserved. In the same way, new punishments may serve to alert me to the possibility of promoting my own personal advantage.

Another way in which further sanctions may actually reduce virtue is by marginalising or eliminating my non-egocentric motives instead of, as in the previous case, making me more aware of my own personal interests. If all incentives take the form of positive or negative sanctions that have an influence on my own selfish interests, they may lead me to think in terms of these selfish interests even though previously I did not do so. Sanctions that are harsh enough to influence a wrongdoer may actually eclipse the virtue of ordinary citizens. If, for example, the state decides to offer me a substantial reward for something that I previously did out of altruism, my altruistic motivation will probably be undermined.

A third way of producing the same effect is through the mechanism of the *self-fulfilling prophecy*. If, for example, tough controls are imposed on a certain group of workers who, up to that time, had generally performed their duties in a reasonable manner, this will project a stereotyped and negative image of these workers, thus leading other people to see them as villains who comply only under the threat of harsh sanctions. This can cause adverse effects such as *labelling* or *provocation* which could eventually lead the workers to adopt the kind of behaviour associated with such a stereotype.

In the three cases discussed up to this point, the imposition of tough sanctions will undermine virtue by changing the basis of assessment in different ways. However, deviant-centred strategies could have two further kinds of counter-productive effects on compliance and virtue which, while similar to the others, should be treated differently.

A fourth effect is to *divert attention to the possibility of desertion*. Pettit (1996) offers a highly illustrative example of this effect: Imagine a worker who is continually obliged to watch the clock under the threat of heavy sanctions if he arrives late to work. Firstly, he will probably become demoralised and lose a great deal of his virtuous motivation as a result of the *self-fulfilling prophecy*

effect mentioned above. However, the fact that he is obliged to pay so much attention to this issue will also make him more aware of the possibilities available to him to avoid complying with his duties without being found out.

The last effect on motivation that I want to emphasise here is that new sanctions act as a kind of signal that while I comply systematically with my obligations, there are others who do not (Pettit, 1996, 1997 and 2002): My awareness of the non-compliance of others undermines the basis for my own compliance.

At this point I would like to stress that, however effective sanctions may be, a state of egocentric deliberation will never lead to greater compliance than one in which players are reasonably motivated to comply. This is because in almost all real-life situations, players will always have some opportunity to avoid complying without being found out. If the threat of punishment or the expectation of some reward are the only things motivating them to do their bit, whenever they think they have a chance to avoid complying, while escaping punishment and receiving the reward, they will not hesitate to seize it. Thus, the imposition of strong positive or negative sanctions designed to increase levels of compliance could actually have the paradoxical effect of reducing them.

The effects mentioned up to this point relate to the different mechanisms by which the imposition of sanctions deactivates virtue, but deviant-centred strategies can also have a negative effect on choice, leading to the *adverse selection* effect. Imposition of very severe penalties may discourage the virtuous from becoming involved in a particular field of activity, while very high rewards may attract the villainous.

On the other hand, the potential counter-productive effects of such strategies are not necessary limited to the fact that agents are demotivated and compliance is reduced. In a well-known study on regulation, Bardach and Kahan (1982) argue that a strategy based chiefly on sanctions can lead to defiance and an organised culture of resistance and counterattack. Indeed, the agents could even organise themselves into groups to avoid regulation. Regulated enterprises can, for example, share information about legal loopholes and the best way to avoid compliance. Resistance against regulation can even mean that agents who oppose the regulation will resort to organising themselves politically through lobbies to revoke the regulation in question. In general, this leads to a sort of cat and mouse logic by which enterprises undermine the spirit of the law and attempt to exploit the legal loopholes, while the state finds itself obliged to dictate a greater number and more specific laws in an attempt to tighten up those loopholes. According to Bardach and Kahan, this type of logic may result in a final package of incoherent rules and/or legalism that focuses on the most visible, simple and specific violations of the law, while neglecting systemic problems that lie at the bottom of the issue. These effects have also been identified in a large number of later studies on different spheres of regulation[2].

2.   See, for example, Sherman (1993) or Braithwaite, K. Murphy and M. Reinhart (2007).

*Complier-centred strategies*

Unlike deviant-centred strategies, *complier-centred* institutional design strategies work on the premise that human beings are neither virtuous nor villainous by nature. Instead, they perceive people as generally motivated to comply with the rules of cooperation[3] while, at the same time, intrinsically corruptible. Indeed, findings in fields such as psychology, evolutionary biology or cognitive science have shown this premise to be plausible[4].

Thus, institutional design should be directed primarily towards ordinary people who are reasonably motivated to comply, and not towards the villainous (Ayres and Braithwaite, 1992, Gunningham and Grabosky, 1998, Pettit, 1996, 1997, 2002). I will now examine this strategy point by point.

First Principle: Screening devices before sanctions

Pettit (1993, 1996, 1997, 2002) suggests that before formulating sanctions, all the available options for screening should be explored. The possibility of screening out all undesirable options and individuals from the outset is certainly worth investigating. The best-known screening mechanism entails sifting all the agents involved. For example, there are institutional mechanisms that allow the different sides in a legal action to reject members of a jury or tribunal if they believe that the members may have a particular interest in the outcome of the case or if there is a risk that they may not be impartial for some other reason. This mechanism does not imply that these people are going to behave improperly, merely that they have a strong incentive to do so. Sifting them out avoids the need to design strict penalties and monitoring mechanisms to guarantee compliance.

The screening of agents and options may also require the contrary, that is, positive sifting which includes certain agents or options that are excluded. For example, Pettit (1996, 1997, 2002) mentions systems which offer individuals the possibility of expressing their complaints or introducing options for a court or parliament to obtain expert advice on a particular issue as examples of positive sifting of options. Once the possibility of applying screens is explored, the second step of the strategy consists of applying sanctions.

Second Principle: Non-alienating sanctions to control the agents selected

The second principle of complier-centred strategies consists of applying sanctions in a deliberately encouraging way (Pettit, 1996, 1997, 2002). Generally, this means trying - to the greatest possible extent - to avoid sanctions that offer very high rewards or very harsh punishments such as those mentioned above with regard to deviant-centred strategies. Sanctions should instead be neutrally motivated, i.e. they should not presuppose that the people at whom they are directed are wicked, so that these people do not regard the sanctions

---

3. I leave aside questions such as who define rules of cooperation, etc.
4. See, for example, Tena (2010).

as an affront. There is no reason to suppose that sanctions of this type should have negative effects of any kind on the motivation of individuals.

There is a second requirement that should, as far as possible, be met by sanctions, and this goes beyond motivational neutrality. Wherever possible, sanctions should appear to be optimistically motivated (Pettit, 1997). They should project a good image of the people they are aimed at, who must feel that they are treated with esteem and trust, and be given the opportunity to live up to that[5].

Third Principle: Sanctions suitable for offenders
who repeatedly behave as knaves

The third principle of complier-centred strategies relates to the need for structural sanctions to tackle wrongdoing (Pettit, 1996, 1997, 2002). As stated earlier, although the majority of people can generally be expected to be motivated to comply, there is always the possibility of encountering wrongdoers; those devils who filled Kant with such fear. It has been argued that institutional design should primarily be directed towards compliers but, by the same token, it should provide sanctions against wrongdoing if, as mentioned earlier, the institution does not want to end up succumbing to it (Ayres and Braithwaite, 1992, Pettit, 1996,1997, 2002). In this case, sanctions that are sufficient to motivate those who are entirely villainous should indeed be contemplated, but in contrast to what happens with *deviant-centred* strategies, these sanctions should be designed in such a way that is neither threatening nor offensive to those who were originally motivated to comply (Ayres and Braithwaite, 1992, Pettit, 1996, 1997, 2002).

Clearly, no institutional design applied in the real world would result in a perfect degree of compliance, but there are various kinds of measures that can be taken in order to achieve good levels of compliance. Building on the work by Ayres and Braithwaite (1992), Pettit (1996, 1997, 2002) recommends establishing a sliding scale of increasingly tougher sanctions. In their classic work, Ayres and Braithwaite propose the use of enforcement or responsive regulation pyramids. In recent years, numerous authors have proposed different types of regulatory pyramids. In what follows I will review some of the most important proposals focusing chiefly on the contributions of John Braithwaite. As pointed out above, the analysis will centre on the question of institutional design, leaving aside more general contributions that have more to do with the sphere of governance.

---

5.  This second principle ties in partially with what has come to be known as *procedural justice*. In this sense, numerous authors argue that when regulated agents perceive virtues in the regulatory process (for example when they are treated with respect and fairness), compliance tends to improve. Although their results are not very conclusive, some well-known empirical works on this issue in a variety of spheres such as nursing-home regulation, tax compliance or criminology include those of Braithwaite and Makkai (1994, 1996), Braithwaite, Murphy and Reinhart (2007) or Sherman (1993, 2003).

*Regulation through pyramids*

Responsive regulation and enforcement pyramids

Ayres and Braithwaite (1992) coined the concept of "responsive regulation" to refer to their proposal of regulation strategy. This proposal was originally conceived to deal with the regulation of business by governmental agencies. However, there is nothing in the proposal that necessarily confines it to this original field, and in subsequent years Braithwaite worked on a variety of projects to apply the strategy in areas as diverse as criminology (2002; Braithwaite and Pettit, 1990), fiscal compliance (2003) or development in third-world countries (2006). In general terms, responsive regulation is a universalistic strategy which should, in principle, be applicable to any area of institutional design. As a strategy, it should be included in the family of complier-centred strategies.

These two authors advocated the use of an enforcement pyramid. At the base of this pyramid we find sanctions or regulation strategies based on consideration and respect. Higher up the pyramid, the measures become increasingly harsh. The basic logic behind the system consists of amiable regulation for common citizens, with exceptional measures reserved for specific wrongdoers. The content of the sanctions pyramid will vary from one context to another, just as different sanctions will be appropriate to different areas. Let us look at an example: Diagram 1, depicts a case in which a strategy is applied to corporate regulation. Observe how the measures become gradually harsher, ranging from persuasion to withdrawal of the licence.

According to this strategy, the principle of "minimal sufficiency" of sanctions should be followed at all times. In the beginning, the regulator must "speak softly" to try to persuade agents to comply of their own free will. If the agents do not comply, the regulator must gradually move up through the sanctions pyramid, but never move on to the next level until all the possibilities of the previous level have been exhausted.

At the top of the pyramid stands the "benign big gun". As can be seen from the previous example, the last level in the pyramid consists of a drastic and definitive incapacitating measure which, depending on the context once again, could involve imprisonment, withdrawal of the licence or other types of exclusion measures.

A *tit for tat* logic should be followed at all times. Once we begin using a persuasion strategy, we must always reward compliance (even in habitual villains) and we must always punish failure to comply (even in those who are habitually virtuous). In this way, cooperation can be expected to become the dominant strategy for the majority. It is worth emphasising this aspect of the strategy. Indeed, the tit-for-tat strategy can prove to be an extremely productive regulatory strategy. In a celebrated work, R. Axelrod (1984) demonstrated that, under the right conditions, this kind of strategy could lead to cooperation among selfish agents. We are now able to make the positive assertion that not only is this the most suitable strategy for engendering cooperation between the

selfish, but it also works well with common individuals who, as observed, generally display conditional motivations regarding compliance. Tit for tat suggests that regulators should at first cooperate (Axelrod, 1984). In this regard, it is a friendly strategy that will not demotivate the majority who are willing to comply from the outset. Given that most people are generally willing to cooperate so long as they are assured that others will also comply; the tit for tat strategy is also useful for dealing with them in a secondary manner. Because this strategy provides for sanctions on non-compliers, it will not only *not* demotivate those who do comply, but also give them the confidence they need to cooperate. The fact that there is no threat of sanctions from the very beginning, but is held in reserve, means that compliers will not feel that they have been hard done by. In addition, tit for tat will also be the best strategy for combating those who are not motivated to comply at a given moment and prefer to try and exploit any opportunity to gain an illicit advantage.

However, we may also come across players who refuse to cooperate even when this means facing costs. Such cases must continue to be punished with sanctions that become increasingly severe until compliance is finally achieved or the "benign big gun" is reached. However, it should come as no surprise that the benign big gun, given that it is a draconian measure, is rarely if ever imposed.

**Diagram 1.** Example of an enforcement pyramid. Source: Ayres and Braithwaite, 1992.
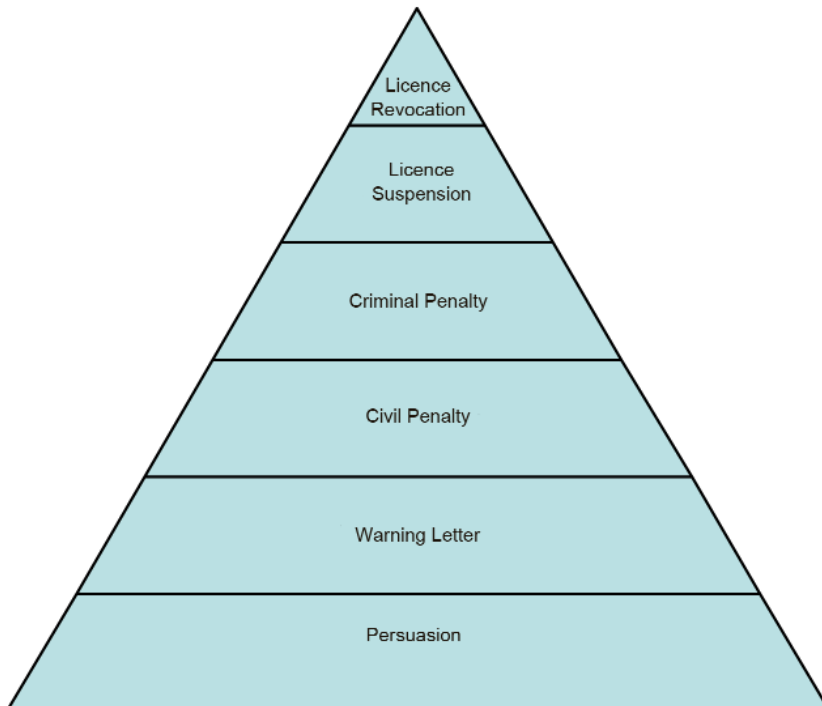
A benign big gun is benign only up to the point when it is fired. When a benign big gun is "reached," it is no longer benign. The amiable measures at the bottom of the pyramid will serve no purpose if the regulator does not hold big guns that are kept in the background. But when such measures do exist, one expects citizens to comply without having to resort to them.

Note that, in the same situation, there may be (and can be expected to be in practically all situations) players at different levels within the pyramid. The majority can be treated "gently", while some specific compulsive offenders may be subjected to drastic punishment. Moreover, it is always better to have a pyramid than a single sanction, however harsh the sanction may be, partly because attempts to apply very harsh sanctions for minor offences end up causing under-regulation, as the authorities responsible for applying the penalties will refrain from doing so (Grabosky, 1995, Sunstein, 1990, 1990b).

As mentioned earlier, despite the fact that the nature of sanctions will vary from one scenario to another, Ayres and Braithwaite suggest that we should follow a general "super-punishments" pattern. These should take the form of a brief period of punishment ("stick") followed by a longer period of reinsertion ("carrot"). In addition, the offender being punished is urged to collaborate in his own punishment during the sanction period. The more he collaborates, the shorter this period will be and the quicker he will move on to reinsertion.

*Some additional contributions*

In a well-known study on environmental policy design, Gunningham and Grabosky (1998) provide some additional advice that should be taken into account when using pyramids in regulation[6]. Firstly, in the event that an instrument (or a combination of instruments) that appears to be viable in its entirety finally turns out not to be so, the authors propose sequencing the instrument, that is, to enable escalation from the preferred least interventionist option, and if this fails, to move up the pyramid towards more interventionist measures.

Secondly, the authors recommend reducing the level of discretion in such a manner that the jump from one level to another of the pyramid is as automatic as possible and does not depend too heavily on the regulator's own criterion. Reducing the level of discretion lowers the risk of lobbying by the most powerful agents.

Gunningham and Grabosky also point to the need for *triggers* and *buffer zones* between the different levels of the pyramid. Firstly, there should be triggers that automatically alert authorities when the least interventionist measures have failed. Among the elements mentioned in their study, the

6. Gunningham and Grabosky specifically refer to an expanded regulatory pyramid model that would include multiple instruments and parties. Given that the authors' proposal goes far beyond the debate on the most adequate designs for institutional measures and is directed more at general issues having to do with governance (such as the need to incorporate third parties in regulatory processes), I leave much of their proposal aside and focus only on those elements which can be incorporated into an enforcement pyramid.

authors highlight random inspections, independent audits or mechanisms for industry association reporting as possible triggers to alert authorities to non-compliance. Secondly, the aim of regulation should be to prevent non-compliance rather than to punish it when it has already occurred. Thus, buffer zones should be set up between the point at which the trigger goes off and the level of harm that is being monitored. The higher the level of compliance required, the larger the buffer zone needs to be.

Finally, Gunningham and Grabosky (1998: 407) refer to the need for *circuit breakers*. Circuit breakers "tend to be low intervention instruments, introduced in anticipation that certain high intervention instruments, introduced in isolation, have a high chance of failure." Although the circuit-breaker concept is related to the idea of instrument sequencing, there is a fundamental difference between them. Escalating a pyramid towards increasingly harsher measures occurs only when lower measures have failed. In contrast, a circuit breaker is a short-term option that will eventually be replaced by another more conventional measure.

In spite of the fact that this idea was already discussed more or less explicitly in the work of Ayres and Braithwaite, it is worth highlighting that numerous authors such as Black (1998), Braithwaite (2008), Braithwaite and Makkai (1994) or Braithwaite, Murphy and Reinhart (2007) have argued in favour of regulators establishing processes of dialogue with the regulated. Dialogue, insofar as it is a fairly non-intrusive instrument and emits a clear signal that the regulator holds the regulated in high esteem, is an ideal instrument for the base of an enforcement pyramid. According to Braithwaite, Murphy and Reinhart, in the case of tax compliance, such processes can be carried out through various methods of communication such as intermediaries, the sending of letters, etc. According to the authors, the objective is three-fold: First, regulators must try to morally engage agents in the aims of the policy in question. Second, they must reach a consensus as to when it is legitimate for taxpayers to reduce the tax paid to government and when such an attitude constitutes a lack of civility. Finally, by talking to taxpayers, the regulator has the opportunity to deal with the possible injustices of the measures to be implemented (Braithwaite, Murphy and Reinhart 2007: page(s?)).

*Incentives in an enforcement pyramid and the strengths-based pyramid*
So much for the basic elements of enforcement pyramids. Nevertheless, there is more to be said about the use of positive and negative sanctions in this strategy. According to Braithwaite (2002b), punishments are more effective than rewards as part of a regulatory strategy, at least in the case of explicit punishments of a material nature.

When rewards are introduced at the base of the pyramid in order to encourage compliance, they often end up generating perverse incentives. In the first place, those who were originally motivated may switch to a system of egocentric deliberation. Generally speaking, it is highly likely that they will set out to gain maximum access to the available incentives at a minimum cost.

They will try to make the regulator believe that they are complying to a greater degree than they really are, they will begin to seek incentives to do things that they would have done anyway (even if no such incentives had been there) or they will begin to comply only in those areas in which they are better rewarded, ignoring the others. Drawing on a concept from McBarnet and Whelan (1999), Braithwaite calls this kind of strategy "creative compliance".

As mentioned earlier, both rewards and punishments run the risk of demotivating those who were initially motivated. In this regard, Braithwaite maintains that the threat of punishment is more demotivating than a reward. There are, nevertheless, powerful reasons to have more faith in punishments than in rewards within an enforcement pyramid. At the base of the pyramid both punishments and rewards can generate adverse effects. In the middle sections, however, it is better to rely on punishments. The other way round it would give incentives to not comply at the beginning with the subsequent possibility of gaining access to incentives aimed at bringing non-compliers back onto the path of compliance. In the same line, Gunningham and Grabosky (1998) hold that command-and-control regulation and broad-based economic incentives are an inherently counterproductive combination.

However, if we broaden our approach and look beyond material incentives, as suggested, intangible rewards based on respect will have positive effects at all levels of the pyramid. Showing esteem, trust, respect and consideration for the regulated will be a good strategy both at the base of the pyramid and when reinserting those who are now complying after having been punished.

Nonetheless, the difficulties involved in integrating the tangible rewards within the strategy has been one of the main critiques of responsive regulation, since we know that giving rewards generally functions better than threatening to sanction. In this sense, Braithwaite, Makkai and Braithwaite (2007) and Braithwaite (2008) have developed the strengths-based pyramid to complement the traditional enforcement pyramid.

Strengths-based intervention consists of trying to build the strengths of a given agent. These strengths are developed in such a way that they can ultimately expand and overcome other weaknesses or compensate for them.

Table 1 shows the main design principles that distinguish both types of pyramid.

**Table 1.** Distinguishing the design principles of a regulatory enforcement pyramid and a strengths-based pyramid. Source: Braithwaite, Makkai and Braithwaite, 2007
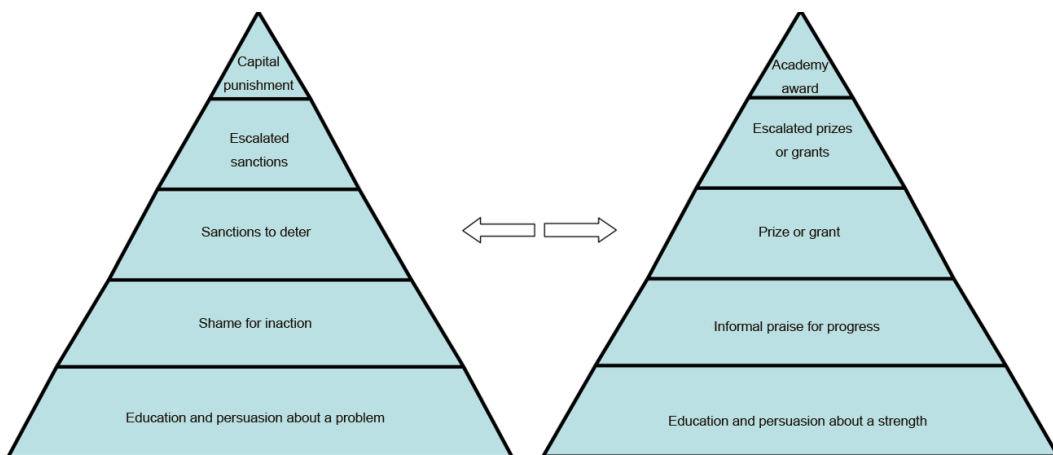
| Regulatory pyramid | Strengths-based pyramid |
| --- | --- |
| Risk assessment | Opportunities assessment |
| Fear | Hope |
| Prompt response before problem escalates | Wait patiently to support strengths that bubble up from below |
| Pushing standards above a floor | Pulling standards through a ceiling |

While the enforcement pyramid attempts to respond to fears over risks, the strengths-based pyramid tries to foster opportunities. In this way, the strengths-based pyramid is not designed to ensure minimum standards, but to maximise quality by pulling standards through a ceiling. The aim of combining both pyramids in regulation is to continually improve compliance. For example, in their study on nursing-home regulation, Braithwaite, Makkai and Braithwaite show that, from this new viewpoint, organisations must not only have a system to manage and detect risk that permits them to identify and resolve the main problems, but also a system that permits them to identify the principal opportunities for improvement and to work from there. Diagram 2 provides an example of how both pyramids can be compared and applied in an integrated manner.

The two pyramids are linked at the base. The remaining steps of the pyramids, however, contain alternative rather than complementary strategies. The strengths-based pyramid can alternate between grants and prizes. In general, the prizes build strengths among those who are more excellent. Nonetheless, there is no problem in awarding a prize to an agent for a particular accomplishment even when the agent has achieved poor results in other dimensions of compliance. On the contrary, it may contribute to improving the agent's motivation to comply in other areas. Grants may be associated to prizes on certain occasions. However, it generally makes more sense to dedicate these resources to agents that are building strengths from a position of weakness. In spite of this, for the agent in question to receive the prize he must have already begun to build the strength, and the grant should provide him with the necessary resources to continue in his pursuit.

Combining an enforcement pyramid with a strengths-based pyramid poses no problems in terms of the use of rewards mentioned above. An agent who lets

**Diagram 2.** A possible approach to integrating a strengths-based pyramid with a regulatory pyramid based on the work on nursing-home regulation by Braithwaite et al. (2007) Source: Braithwaite, 2008.

a problem get out of hand will receive neither a grant nor a prize, but a punishment. If a disobedient agent is pushed to the bottom of the enforcement pyramid, he may access the strengths-based pyramid if he continues to improve.

### Critiques and limitations of complier-centred strategies

In this section, some problems of complier-centred strategies are discussed:

### The absence of adequate instruments

In the work on environmentalpolicy design mentioned above, Gunningham and Grabosky (1998) examine the possible problems and limitations involved in using enforcement pyramids in regulation.

The first problem is to determine what type of instruments can be used in this way. Controlled escalation is only possible when there are instruments that permit a graduated, responsive and interactive escalation. According to Gunningham and Grabosky, the most appropriate instruments in this regard would be self-regulation and command and control. Yet there are also instruments that are less amenable to such a strategy. This is the case of economic instruments. In some circumstances the problem may stem from the fact that escalation is only possible up to a certain level given that the appropriate instruments to go beyond that level or agents capable of applying them are absent.

In these circumstances, Gunningham and Grabosky suggest that the solution may be for a government to fill the gaps between the different levels of the pyramid. In these cases, the government must act as a catalyst to aid second and third parties in climbing the pyramid. To achieve this, the authors propose a three-sided pyramid (expanded-enforcement-pyramid model) in which a large number of instruments and parties are combined in an attempt to overcome the limitations of Braithwaite's enforcement pyramid. Nonetheless, their criticism of responsive regulation does not appear to be completely in line with the reality on this point. In their work published in 1992, Ayres and Braithwaite dedicated an entire chapter to the question of tripartism and the need to incorporate third-parties in the regulatory process that were neither government institutions nor enterprises. In recent years, however, Braithwaite (2004, 2006, 2008), Braithwaite, Makkai and Braithwaite (2007) or Drahos (2004) have developed the concept of pyramids of networked escalation. In this type of pyramid, different types of agents work in a coordinated manner to design and implement the measures.

I will not go into this issue here, however, as it goes beyond the scope and objectives of the paper to engage in more general debates about governance. As said above, this paper centres on the design of specific institutional measures and not on who designs or applies them or what procedures are used to carry them out. Instead, I argue that the solutions proposed by the authors are satisfactory, but that if in a specific case it were not possible to apply an enforcement pyramid due to the absence of adequate instruments, resorting to more interventionist measures could be justified.

### Impossibility of applying the tit-for-tat strategy

Gunningham and Grabosky argue that in contexts in which there are a large number of small agents, and the interactions between them and the regulator are therefore infrequent, a first response that is more interventionist could again be justified. At this point, different solutions are possible. In contexts where agents of different sizes come together, the regulator may attempt to make larger agents responsible for the conduct of smaller ones (Braithwaite, 2008, Gunningham and Grabosky, 1998). In the chemical industry, for example, there are a great number of small companies that depend on the large multinationals from which they buy or sell their products. Regulatory bodies must make multinationals accountable (even if it is only in terms of reputation) for the conditions under which the products they buy are produced, or for the destination of their sales, etc. Another possible solution would be to use pyramids of networked escalation in such a way that the regulatory body seeks partners that will aid them in ensuring compliance with the regulation (Braithwaite, 2008). Again, I will not go into detail here about this type of solution for the reasons stated above. Simply let me state that the regulatory body should lay the grounds for the pyramid to function properly, that is, for instance, the necessary measures must be put in place so that the tit for tat between the regulator and the regulated operates in a correct manner. More specifically, the solution proposed by Braithwaite is satisfactory. However, as in the above case when there are interim situations in which this is not possible, recourse to more severe measures could be justified in the first instance.

### Situations in which non-compliance could lead to irreversible or catastrophic losses

Finally, Gunningham and Grabosky (1998) argue that using the pyramid could be a bad strategy in situations where non-compliance could lead to an irreversible and catastrophic loss. Under circumstances of this kind, the authors argue that it is preferable to take a horizontal approach using a wide range of instruments including underpinning regulation in a simultaneous rather than a sequential manner.

Braithwaite (2002) accepts this possibility. He gives an example in which a hijacker has taken hostages. The hijacker has killed some of the hostages and seems ready to kill more. A police marksman then has a chance to kill the hijacker with a single shot without endangering the hostages. In such a case, common sense dictates that he should do so, even though persuasion and subsequent imposition of restorative justice could also be expected to end up working in this case. Moving on from the anecdotal to the general principle, we might concede that responsible regulation could be bad advice under circumstances of this kind.

### The link between complier-centred strategies and republican ethical-political doctrine

A further controversy that arises with these institutional design strategies is that of their relationship to normative theories. Obviously, these strategies are normative in a certain sense: they inform the regulator about what to do if he

wants to achieve certain objectives. This is the "technological" sense of normativity (Noguera, 2009). However, the mere fact that these strategies are normative from the technological standpoint does not mean that they are normative in an ethical-political sense. Some of authors who defend them, however, seem to argue that such strategies are related to specific ethical-political doctrines, namely civic republicanism. To be more precise, arguing that these strategies "are related" with specific doctrines can be interpreted in two ways. First, it can mean that assuming a republican perspective entails adopting these strategies. Second, it can mean that adopting these strategies entails assuming a republican ethical-political doctrine. Some of authors who defend these strategies seem to sympathise with both points of view.

For example, Braithwaite and Makkai (1994) hold that conceiving of individuals as rational and selfish agents is based on a liberal conception of individual action in an organisational world, whereas conceiving of them as "citizens", as agents willing to respond to the trust deposited in them, is based on a civic republican concept. According to this point of view, we would be forced to conclude that applying a strategy of responsible regulation would involve assuming a republican ethical-political doctrine, and that assuming a republican ethical-political doctrine would entail applying a strategy of responsible regulation. It would also seem to suggest that assuming a liberal ethical-political doctrine would entail applying a deviant-centred strategy and, even maybe, applying a deviant-centred strategy would imply assuming a liberal ethical-political doctrine.

Braithwaite, Makkai and Braithwaite (2007) initially take a more moderate position but eventually reach a similar conclusion. In the beginning, the authors stress that one critique of the *networked enforcement pyramid* (and, more so, of the *enforcement pyramid*) is that it can constrain the freedom of citizens. The authors accept that discipline networks can both constrain liberty and enable it. In this sense, some kind of normative theory such as a republican theory of freedom as non-domination is necessary to guide the design of such networks. Nonetheless, further on (Braithwaite, Makkai and Braithwaite, 2007: 323) the authors once again attempt to analytically link their proposals to a specific doctrine, namely civic republicanism. The authors argue that linking a *strengths-based pyramid* to the *networked enforcement pyramid* constitutes a mechanism of control over tyranny. From the moment strengths are built, the capacity to resist domination is constructed. Thus, the authors once more seem to suggest the existence of a two-fold analytical tie between the proposed institutional design strategies and republican doctrine. On the one hand, assuming a republican ethical-political doctrine would appear to largely determine which institutional design strategy is chosen. On the other hand, assuming the institutional design strategies proposed by the authors would seem to once again entail the acceptance of a republican doctrine. At the same time, such regulatory strategies would appear to be incompatible with other ethical-political doctrines. For example, it would seem that this strategy could not be used to motivate the civil servants of a fascist dictatorship.

In contrast, this paper takes a sceptical position regarding the arguments of the above authors. Our discussion centres chiefly on the second argument as it is the most relevant to institutional design. This is so because if being republican determines whether or not these strategies are chosen is a question that should solely concern authors who share a republican doctrine, whereas if accepting the goodness of such strategies means being republican, it should be of concern to all those interested in institutional design.

On the one hand, as argued above (and as Braithwaite, Makkai and Braithwaite (2007) acknowledge), it seems plausible to state that an enforcement pyramid can both promote and constrain freedom. Without going further, it is difficult to see why a pyramid of this kind could not be used to motivate agents who engage in some type of activity that is harmful to the common good. This type of pyramid could be used, for example, to motivate workers in a highly contaminating factory or in one that manufactures weapons of mass destruction, or civil servants who work in a prison under a dictatorship that oppresses political dissent, among others. Obviously, this argument does not only affect enforcement pyramids. A totalitarian regime could, for example, apply screens such as those proposed by Pettit (….) to ensure that people suspected of opposing the regime do not hold public office. The list of examples is endless.

On the other hand, it is difficult to see why this strategy could not perfectly well be applicable in order to turn a decent into an indecent society (or into a decent but not republican society), one that wants to elicit compliance by a majority to norms that are not fair, such as norms that discriminate against minorities, without wanting to do away with the "civic virtue" of the majority. The tit-for-tat logic, for instance, also seems to be perfectly compatible with the advent of a tyranny of the majority; a "soft" totalitarianism as described by Tocqueville (…).

The above authors could respond to these critiques in a number of ways. First, as pointed out above, the authors propose situating dialogue at the base of the pyramid. For authors like Braithwaite, Murphy and Reinhart (2007), dialogue is not only an amiable and unintrusive instrument that may encourage voluntary compliance as well as correct inefficient or unreasonable aspects of the measures to be applied, but it also paves the way for citizens to dispute the decisions made by those who govern them, which is a constitutive element of the republican notion of *freedom as non-domination* (Pettit, 1999). So, situating dialogue at the base of the pyramid would be a way of promoting *freedom as non-domination.*

Nonetheless, the existence of processes of dialogue, in spite of the fact that they may give rise to more reasonable and fair decision-making processes as well as the voluntary compliance of citizens, is no guarantee that the decisions made during the process of dialogue will promote the common good or much less republican freedom. In fact, this seems to be a somewhat circular argument that is common to some republican conceptions of democracy, deliberation and dispute. Indeed, such an argument would seem to contend that the decisions

made during these processes of dialogue are the right ones for the mere reason that the right decision is that which is made in a process of dialogue of this kind.

Braithwaite, Makkai and Braithwaite (2007) could provide a second response. As I have said, combining a *strengths-based pyramid* with the *networked enforcement pyramid* constitutes a mechanism of control over tyranny since by building strengths, the capacity to resist domination is also built. Once again, however, we must ask why it is only possible to develop strengths in the weakest agents and why such strengths should always aid them in resisting tyranny.

Finally, it also seems to be a complicated task to argue that assuming a liberal doctrine implies assuming that individuals are rational and selfish, or that such a concept should be the premise for institutional design. The contemporary political liberalism of authors such as Rawls or Van Parijs is not committed to any specific type of institutional design (though obviously can be incompatible with some). This liberalism can be committed to a particular concept of justice, for example, but deciding which institutional design is the best for developing the concept is a purely instrumental question that should be determined empirically[7].

In short, the argument put forward by the authors regarding this point seems to fall into the fallacy that all good things (full compliance, efficiency, republican freedom, etc.) come in the same package. In contrast, this paper takes the stand that the institutional design strategies being defended are independent of concrete ethical-political theories. Thus, applying these strategies can ensure a good level of compliance and achieve the goals of a given policy. Nonetheless, what those policy goals should be is a question that must be determined by means of an external and independent normative theory. In the words of Gunningham and Grabosky (1998: 377) (in reference to environmental policy design):

> The purpose of design processes and principles is to guide policymakers in developing an optimal policy mix to achieve *a given environmental goal* (or goals), not to determine what the goal should be. Fundamentally, the design processes and principles do *not* address the issue of how governments should establish environmental policy goals. Rather, it is assumed that government has already determined a specific environmental goal or goals. It is only after this goal has been articulated that our set of design processes/principles can come into play.

*Limited informative power of the proposals*

A problem affecting all the strategies discussed in this paper (albeit to a different degree) is their excessive vagueness and limited informative power.

---

7. Makkai and Braithwaite might answer that (atomist) liberalism they criticize is not Rawls' or Van Parijs' liberalism. However, it would be also possible to ask what liberal authors fit nowadays the type criticized by Makkai and Braithwaite.

The proposals are too general, thus entailing both a normative and an empirical problem. At the normative level, the fact that the proposals are too vague generates problems when recommending specific policies in each particular case. At the empirical level, the problem has to do with the fact that it is difficult to decide *a priori* whether or not a given measure is based on compliance or to what extent. In this sense there is an undesirably wide margin that allows us to read actual laws and policies from the point of view of one ideal type or another, depending on our own interests. This problem does not affect all the proposals examined in an equal manner. Specifically, the problem affects the proposals of Pettit (….) and that of the strengths-based pyramid to a larger degree, while affecting the enforcement pyramid to a much lesser degree. It is therefore necessary to continue working to develop more informative concepts and hypotheses.

*Cases and agents with which the proposals work*
Another problem concerns the type of agents and cases for which the strategy is expected to work. As said above, the recommendations of the strategy should take into account the particular idiosyncrasies of each specific case. For example, in the case of Ayres and Braithwaite (1992), the content of the regulatory pyramid will change according to the sphere to be regulated. However, at least in principle, the basic pillars of the proposal should aim to be universal. While certainly an appealing statement, it is one that must be thoroughly explored at the empirical level. Likewise, the strategy may function differently not as a result of the type of situation (or not only) but also due to the type of agents involved.

In recent years a vast amount of empirical work has been done that should help to shed some light on these issues. Yet, given the general difficulties involved in gathering and analysing social science data, specially in areas like institutional design, the results are, in general, not very conclusive[8].

Although fieldwork may be very valuable here, so could other types of approaches such as experimental designs or multi-agent simulation models. Such approaches should involve empirically exploring different types of circumstances, while taking into account relevant variables and determining if the strategy works well across the board, and if it does not, determining which characteristics and mechanisms best explain the different results achieved.

**Conclusions**

This paper has attempted to explore institutional design strategies which promote compliance. More specifically, it has tried to identify designs for laws, policies, regulations, by-laws and formal rules in general aimed at either

---

8.   See, for example, Braithwaite, Murphy and Reinhart (2007) or Sherman (1993, 2003).

promoting or eradicating certain kinds of behaviour in order to ascertain which of these designs achieves the highest levels of compliance.

In line with the specialised literature, a distinction has been made between institutional design strategies centred on sanctions, deviations or deterrence and compliance-centred strategies. The main characteristic of the first type of strategies is that they view agents as villains, while those of the second type assume virtue but are prepared for non-compliance. Some of the problems that arise in the first type of strategies were then discussed and an overview given of some of the most important contributions to the debate on the second type. Specifically, contributions by authors such as P. Pettit, I Ayres, J. Braithwaite, V. Braithwaite, T. Makkai, N. Gunningham or P. Grabosky, among others, were considered. Finally, the possible problems and limitations of these proposals were explored, and possible lines of work to be undertaken in the future were suggested.

## References

AXELROD, Robert (1984). *The Evolution of Cooperation*. New York: Basic Books.

AYRES, Ian and BRAITHWAITE, John (1992). *Responsive Regulation*. New York: Oxford University Press.

BARDACH, Eugene and KAHAN, Robert A. (1982). *Going by the Book. The Problem of Regulatory Unreasonableness*. Philadelphia: Temple University Press.

BLACK, Donald (1976). *The Behavior of Law.* New York: Academic Press.

BLACK, Julia (1998). «Talking about Regulation». *Public Law* (Spring), 77-105.

BRAITHWAITE , John (2002). *Restorative Justice and Responsive Regulation*. Oxford: Oxford University Press.

— (2002b). «Rewards and Regulation». *Journal of Law and Society*, 29 (1), 12-26.

— (2003). «Meta-Risk Management and Responsive Regulation for Tax System Integrity». *Law and Policy*, 25 (1).

— (2004). «Methods of Power for Development: Weapons of the Weak, Weapons of the Strong». *Michigan Journal of International Law*, (26), 296-330.

— (2006). «Responsive Regulation and Developing Economies». *World Development*, 34 (5), 884-898.

— (2008). *Regulatory Capitalism. How It Works, Ideas for Making It Work Better*. Cheltenham: Edward Elgar.

BRAITHWAITE, John and MAKKAI, Toni (1994). «Trust and Compliance». *Policing and Society*, (4), 1-12.

— (1996). «Procedural Justice and Regulatory Compliance». *Law and Human Behavior*, 20, (1), 83-98.

BRAITHWAITE, John; MAKKAI, Toni and BRAITHWAITE, Valerie (2007). *Regulating Aged Care. Ritualism and the New Pyramid*. Cheltenham: Edward Elgar.

BRAITHWAITE, John and PETTIT, Philip (1990). *Not Just Deserts. A Republican Theory of Criminal Justice.* Oxford: Clarendon Press.

BRAITHWAITE, Valerie MURPHY, Kristina and REINHART, Monika (2007). «Threat, Motivational Postures and Responsive Regulation». *Law and Policy*, 29, (1), 137-158.

DRAHOS, Peter (2004). «Intellectual Property and Pharmaceutical Markets: a Nodal Governance Approach». *Temple Law Review*, (77), 401-424.

GRABOSKY, Peter N. (1995). «Counterproductive Regulation». *International Journal of the Sociology of Law*, (23), 347-369.

GRABOSKY, Peter and BRAITHWAITE, John (1986). *Of Manners Gentle*. Oxford: Oxford University Press.

GUNNINGHAM, Neil and GRABOSKY, Peter (1998). *Smart Regulation. Designing Environmental Policy*. Oxford: Clarendon Press.

HAWKINS, Keith (1984). *Environment and Enforcement: Regulation and the Social Definition of Pollution*. New York: Oxford University Press.

MCBARNET, Doreen and WHELAN, Christopher (1999). *Creative Accounting and the Cross-eyedJjavelin Thrower*. Chichester [etc.]: John Wiley & Sons.

NOGUERA, José A. (2009). «Normative Uses of Social Theory». Paper presented at the conference 'The Purpose of Social Theory', Institute for Social Research (YTI), University of Tampere, 24 April 2009.

PETTIT, Philip (1993). *The Common Mind: an Essay on Psychology, Society and Politics*. New York: Oxford University Press.

— (1996). «Institutional Design and Rational Choice». In: GOODIN, Robert (ed.). *The Theory of Institutional Design*. Cambridge: Cambridge University Press.

— (1997). *Republicanism: a Theory of Freedom and Government*. Oxford: Clarendon.

— (2002). «Rational Choice Regulation: Two Strategies». In: PETTIT, Philip (2002). *Rules, Reasons and Norms: Selected Papers*. Oxford: Oxford University Press.

REISS, Albert (1984). «Selecting Strategies of Social Control Over Organisational Life». In: HAWKINS, Keith and THOMAS, John (ed.). *Enforcing Regulation*. Boston: Kluwer-Nijhoff.

SHERMAN, Lawrence W. (1993). «Defiance, Deterrence and Irrelevance: A Theory of the Criminal Sanction». *Journal of Research in Crime and Delinquency*, (30), 445-473.

— (2003). «Reason for Emotion: Reinventing Justice with Theories, Innovations and Research: The American Society of Criminology, 2002 Presidential Adress». *Criminology*, 41, 1-38.

SUNSTEIN, Cass (1990): «Paradoxes of Regulatory State». *University of Chicago Law Review*, (57), 407-441.

— (1990b). *After the Rights Revolution. Preconceiving the Regulatory State*. Harvard: Harvard University Press.

TENA SÁNCHEZ, Jordi (2010). «El Pluralismo Motivacional en la Especie Humana. Aportaciones recientes de la ciencia social experimental». *Papers. Revista de Sociologia*. vol. 95, núm. 2, 421-439.