European University Institute

MAX WEBER PROGRAMME

EUI Working Papers

MWP 2011/10
MAX WEBER PROGRAMME

SANCTION AS A VIABLE TOOL FOR PROMOTING COOPERATION: A COGNITIVE AND SIMULATION MODEL

Giulia Andrighetto and Daniel Villatoro

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
MAX WEBER PROGRAMME

Sanction as a Viable Tool for Promoting Cooperation: A Cognitive and Simulation Model

GIULIA ANDRIGHETTO AND DANIEL VILLATORO

EUI Working Paper **MWP** 2011/10

ISSN 1830-7728

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract
Punishment plays a crucial role in achieving and maintaining norm compliance. Several works have shown that cooperation greatly increases when punishment opportunities are allowed. However, these studies have mainly looked at punishment from the classical economic perspective, as a way of changing people's conduct by increasing the cost of undesired behaviour. In this paper, we distinguish between two enforcing mechanisms, punishment and sanction, focusing on the specific ways in which they promote and maintain cooperation. In particular, by punishment we refer to a practice that works only by imposing a cost, while by sanction we indicate a practice that in addition to that also signals the existence of a norm and that its violation is not condoned. To achieve this, we have developed a normative agent able both to punish and sanction offenders and to be affected by these enforcing mechanisms itself.The results obtained through agent-based simulation show us that sanction is more effective and makes the population more resilient to sudden changes than mere punishment.

Keywords
Punishment, sanction, cooperation, social norms, cognitive modelling, agent-based simulation

A special thanks to Samuel Bowles, Cristiano Castelfranchi, Rosaria Conte, Francesca Giardini, Sven Steinmo, Yane Svetiev, and Luca Tummolini for their helpful comments and suggestions.

Giulia Andrighetto
Max Weber Fellow, 2010-2012

Daniel Villatoro
IIIA - Artificial Intelligence Research Institute, CSIC - Spanish Scientific Research Council

Introduction

Punishment plays a crucial role in promoting and maintaining norm compliance. Several experimental and theoretical studies have shown that cooperation is favoured when punishment opportunities are allowed (Fehr & Gachter, 2002; Boyd & Richerson, 1992; Boyd, Gintis, & Bowles, 2010; Herrmann, Thoni, & Gachter, 2008, Sigmund, 2007). Although these studies have provided key insights to the understanding of punishment, they have largely looked at this mechanism from the classical economic perspective. Namely, it is assumed that individuals obey or break the norm depending on the *price* of violaton – that is, the severity of punishment discounted by the probability that it will be imposed (Becker, 1968).

This *Beckerian* approach to punishment is at odds with some recent experimental evidence that show that (a) in some circumstances punishment has a detrimental effect (Gneezy & Rustichini, 2000); (b) the way in which punishment is implemented affects its effectiveness in promoting norm obedience (Bicchieri & Xiao, 2009), and that (c) when perceived as legitimate punishment is much more powerful (Faillo, Grieco and Zarri, 2010).

In this paper, we suggest that looking at punishment only as a carrot and stick mechanism is an incomplete view and argue that a more insightful understanding of this practice is available once its *norm-signalling* nature is identified (Sunstein, 1996; Masclet et al. 2003; Galbiati and Vertova, 2008; Xiao and Houser, 2009, Giardini, Andrighetto, & Conte, 2010). We suggest that punishment is a powerful means to convey normative messages and normative requests that have the effect of eliciting people's compliance.

As in previous work (Giardini, Andrighetto, & Conte, 2010; Andrighetto, Villatoro, & Conte, 2010), we use the term *punishment* to refer to a practice that works only by imposing a cost; while we will use *sanction* to indicate a practice that in addition to this also communicates the existence of a norm and that its violation is not condoned, thus exploiting the motivating power of norms.

As proposed by several psychologists (Cialdini, Reno, & Kallgren, 1990), philosophers (Bicchieri, 2006), and economists (Houser & Xiao, 2010), the norm focusing effect of sanction plays an important role in promoting norm compliance. When the norm is made explicit, the situation is framed in such a way that both motivations to avoid costs and normative ones are elicited. With *normative motivation*, we refer to the fact that people are disposed to obey the norm even when there is little possibility of instrumental gain, future reciprocation, and when the surveillance rate is very small.

Thus, sanction promotes norm obedience and discourages misconduct by combining the motivating power of social norms with the driving force of the individual's expectations about the price of non-compliance.

In this paper, we explore the hypothesis that the use of sanctions has two main advantages over mere punishment. When enforcing and maintaining cooperation, sanctioning (1) leads to a *higher* level of cooperation, and (2) is *less* costly at a societal level since fewer instances of such enforcing actions are actually needed. The tandem work of cost-avoidance and normative motivations enables a higher and more durable cooperation with respect to an enforcement strategy that relies on cost-avoidance motivations only.

To test this theoretical intuition, we employ an agent-based simulation approach. The modelling focus lies in the effect that the normative information conveyed by sanction has in influencing agents' conduct. What is important is the explicit description of the agents' beliefs and goals and how those are modified by the normative information available in their social context.
Simulation experiments allow us to isolate punishment and sanction and to explore their relative effects on cooperation. Moreover, these experiments enable us to perform what-if analyses relevant for policy design issues.

The article is organized as follows: the section *Punishment and Sanction: A Cognitive Perspective* outlines the theoretical and empirical research background; the section *Agent Architecture*

defines the agent-based model and the internal architecture of the agent; finally, we present and discuss some agent-based simulation results aimed to compare the effectiveness of punishment and sanction and their relative costs. Future work and conclusions follow.

Punishment and Sanction: A Cognitive Perspective

As already stated, we distinguish between two different enforcing strategies, punishment and sanction. We use punishment to indicate a practice that consists in imposing a cost on the offender, with the aim of deterring him from future offenses. Deterrence is achieved by modifying the relative costs and benefits of the situation, so that wrongdoing becomes a less attractive option. The effect of punishment is achieved by shaping the individual's payoffs (Kreps, Milgrom, Roberts, & Wilson, 1982).

This approach to punishment is in line with the economic model of crime, also known as the rational choice theory of crime (Becker, 1968). In this prospect the deterrent effect of punishment is obtained by increasing individuals' expectations about the price of non-compliance. A rational comparison of the expected costs and benefits guides criminal behaviours and this produces a disincentive to engage in criminal activities.

This view of punishment has been attacked by several scholars. In particular Hirschman (1984) states that it considers "citizens just as consumers with unchanging or arbitrarily changing tastes in matters civic as well as commodity-related behaviour".

These researchers criticize the idea that human behaviour is driven only by the motivation to avoid costs. Moreover, this idea is also put into question by considerable empirical evidence showing that punishment can increase cooperation also if it is purely *symbolic* and merely expresses social disapproval, without any material consequences for the punished individual (Noussair & Tucker, 2005).

To make the contrast with punishment vivid, we use *sanction* to indicate the enforcing strategy that, apart from imposing a cost for the wrongdoing, is also intentionally aimed at *signalling* to the offender (and possibly to the audience) that his conduct is not approved of because it has violated a social norm (Giardini, Andrighetto, & Conte, 2010; Houser & Xiao, 2010; Galbiati & D'Antoni, 2007; Masclet, Noussair, Tucker, & Villeval, 2003)[1].

The sanctioner uses scolding to reign in wrongdoers, or expresses indignation or blame, or simply mentions that the targeted behaviour violated a norm. Through these actions, he focuses people's attention on different normative aspects, such as: (a) the fact that the targeted conduct is not approved of because it violates a social norm; (b) the high rate of norm surveillance; (c) the causal link between violation and sanction: "you are being sanctioned because you violated that norm"; (d) the fact that the sanctioner is acting as a norm defender and not for reasons that deprive sanction form its normative content. All these normative messages have a key effect in producing norm compliance and favouring social control as well.

Works in psychology suggest that the influence of a norm is crucially related to the degree to which individuals' attention is focused on the norm. Even a strong personal commitment to a norm does not predict behaviour if that norm is not activated or a focus of attention (Bicchieri, 2006; Xiao & Houser, 2005; Cialdini et al., 1990). Furthermore, the more these norms are made *salient*, the more they will elicit a normative conduct.

We refer to salience as the measure indicating to an individual how operative and relevant a norm is within a group and a given context (Andrighetto et al. 2010; Bicchieri, 2006; Cialdini et al., 1991; Xiao and Houser, 2010). It is a complex function, depending on several contextual, social and individual factors. On the one hand, the actions of others provide information about how important a norm is within that social group. The level of compliance (Cialdini et al. 1991), the surveillance rate, the probability and intensity of punishment, the enforcement typology (private or public, 2nd and 3rd party, punishment or sanction, etc.) (Masclet et al. 2003; Galbiati and Vertova, 2008; Houser and Xiao, 2010), the efforts and costs sustained in educating the population to form a certain norm, the visibility and explicitness of the norm, the credibility and legitimacy of the normative source (Faillo et

[1] Clearly, also punishment can have a norm-signallig effect as by-product, but only sanctions are aimed to achieve this effect.

al. 2010) are all signs through which people infer how important and active a social norm is in a specific context[2]. Individuals with their actions (intentionally or not) signal that there is a norm governing a certain situation and that they want and (explicitly or implicitly) ask that others comply with it. On the other hand, every single agent evaluates how salient a norm is for itself, depending on how much it is consistent with beliefs, goals, values and previously internalized norms of the agent (Deci and Ryan, 2000).

Sanction endows the offender with new normative knowledge that possibly will elicit normative conduct. In other words, the normative information and request conveyed by sanction have the effect of framing the situation in such a way that not only motivations to avoid costs are activated, but normative motivations as well [3].

If successful, sanction drives the wrongdoer to change his conduct not just to avoid the penalty, but because he recognizes that there is a norm and because he wants to respect that norm. Thus sanction has a strong *pedagogical* function, i.e. informing the offender that there is a norm stating that a certain action is prohibited or obligatory and indicating the consequences associated to its violation (Csibra & Gergely, 2009).

The norm-signaling power of sanction allows social norms to be activated and to spread more and more quickly in the population than if it were governed only by mere punishment. This normative elicitation has the effect of activating people's normative motivations to cooperate thereby increasing pro-social behaviours and consequently cooperation within the population.

Thus sanction mixes together material and normative aspects: it is aimed at changing the future behaviour of individuals by acting both on their *cost-avoidance* and *normative* motivations. In order to decide how to behave, the individual will be driven by a combination of cost-avoidance and normative goals[4].

We claim that both punishment and sanction favor the increment of cooperation in social systems, but sanction achieves cooperation in a more stable way and at a lower cost. We expect cooperation to be more robust if agents' decisions are driven not only by cost-avoidance considerations, but are also based on normative ones. Moreover, an individual that cooperates for normative reasons – and not just to avoid punishment – is also more willing to exercise a special form of social control as well: i.e. he will reproach transgressors and remind would-be violators that they are doing something wrong.

In the following sections, we present an agent based simulation aimed to test these hypotheses and discuss some results.

Simulation model

In order to capture the specific dynamics of punishment and sanction and to test their effects in promoting and maintaining cooperation, we have developed a simulation model.

In this model, agents play a variation of the classic Prisoner's Dilemma (PD), in which an extra stage has been incorporated into the game: after deciding whether to cooperate or not, agents can also punish/sanction the opponents who defected. Agents act according to mixed strategies. Unlike a pure strategy, mixed strategies have a probability with which a certain action will be chosen. We designed the simulation experiments in such a way that agents can use *either* punishment *or* sanction, but these two mechanisms cannot coexist in the same experiment. In this way the effects of these two enforcing practices can be isolated and evaluated separately.

[3] The hypothesis that people follow norms as ultimate ends is controversial. But there are several interesting models, such as for example Gintis (2003), that show how normative preferences can be included in the utility function of individuals and how these preferences interact with other preferences of the individual.

[4] In this paper, we assume that the two goals are comparable. On the basis of the relative values of these goals, the individuals will decide which one they want to satisfy, but this is a controversial point that needs a more detailed analysis.

We assume that agents are located in a social network, which determines a fixed interaction topology[5]. Each time-step of the simulation is structured in four phases, that are repeated for a fixed number of time-steps. More specifically, these phases consist in:

Partner Selection: Agents are paired with other agents randomly chosen from their neighbours.
First Stage: Agents play a PD game, with the following payoffs: P(C,C) = 3 , 3; P(C,D) = 0 , 5; P(D,C) = 5 , 0; P(D,D) = 1 , 1.
Second Stage: Agents decide whether to punish/sanction the opponents who have defected. The damage to the offender of both punishment and sanction is 5; while the cost sustained by the punisher/sanctioner is 5/3.

On the one hand, punishment works by imposing a cost to the defector, in this way affecting its payoffs. Apart from imposing a cost, sanction is performed in such a way that a message informing the target that the performed action has violated a social norm is transmitted, thus having an impact both on agents' payoffs and on the process of norm recognition and norm salience. This message can also be listened to by the punished neighbours.

Strategy Update: As agents act according to mixed strategies, these strategies are updated on the basis of the payoffs agents obtained in that round and of the social and normative information acquired.

In the section *Decision Making and Strategy Update*, a description of how agents update their decision making is provided.

Agent Architecture

Unlike the vast majority of simulation models in which heterogeneous agents interact according to simple local rules, in the present model all the agents are endowed with a normative architecture, allowing them: (a) to recognize norms; (b) to generate new normative representations and according to their salience to act on their ground; (c) to observe the behaviours of their neighbours; (d) to influence other agents by direct communication and by the use of punishment or sanction. We base our architecture on a simplified version of EMIL-I-A (Andrighetto et al., 2010).

Our normative architecture has two important components: the norm *recognition module* and the *salience meter*.

The norm recognition module allows agents to interpret a social input as a norm. To recognize the existence of a norm, agents have to listen at least *two* normative messages, such as you should not take advantage of your group members by shirking, and observe *ten* normative actions compliant with the norm or aimed to defend it (i.e. cooperation, punishment and sanction, observed or received). When these conditions are fulfilled, the agents generate a normative belief that will activate a normative motivation (see the normative drive, in following section) to comply with the norm.

The salience meter indicates to the agent how salient a certain norm is and it directly affects the normative drive value. This measure is updated (interaction after interaction) according to both the personal decisions taken by the agents (individual norm-salience) and the normative information that they infer from interacting with their neighbours (social norm-salience).

Information	Weight
Self Norm Compliance/Violation	(+/-)0.99
Observed Norm Compliance	(+) 0.33 $\times n$
Non Punished Defectors	(-) 0.66 $\times n$
Punishment Observed/Given/Received	(+) 0.33 $\times n$
Sanction Observed/Given/Received	(+) 0.99 $\times n$
Norm Invocation Listened/Received	(+) 0.99 $\times n$

Table 1: Norm Salience Meter: Cues and Weights. *n* represents the registered proportional quantity of those events with respect to their neighbour size.

[5] Agents can only observe and interact with their direct neighbours.

Each of these cues (see Table 1) are aggregated with different weights, and a higher weight is given to those that are interpreted as normative actions[6]. For example, all the behaviours that explicitly mention the norm, such as norm invocations or sanctions, have a stronger impact on norm salience than actions in which the normative request is not as explicit, such as punishment. In contrast, observing non punished/sanctioned defectors makes norm salience decrease.

The resulting salience measure (salience [0-1], 0 representing minimum salience and 1 maximum salience) is subjective for each agent. This norm salience meter enables the agents to *dynamically* monitor the normative scene and to adapt according to it.

For example, in an unstable social environment, if a specific norm decays, our agents are able to detect this change, ceasing to comply with it and adapting to the new state of affairs. Instead, if norm enforcement suddenly decreases, agents having highly salient norms are less inclined to violate them. A highly salient norm is a reason for an agent to continue complying with it even in the absence of punishment. It guarantees a sort of *inertia*, making agents less prone to change their strategy to a more favourable one.

Decision Making and Strategy Update

In this model, agents have to take two decisions at two different stages: to cooperate or defect and to punish/sanction or not, and both of them are probabilistic. These decisions are influenced by an aggregation of cost-avoidance and normative considerations. More specifically, the decision to cooperate or defect is affected by the following drives:

(1) Self-Interested Drive: it motivates agents to maximize their individual utility independently of what the norm asks. The self-interested drive is updated according to (a) the calculation of the marginal reward obtained during the last time-step, and (b) the actual action taken. A proportional and normalized value of the marginal reward obtained indicates how the agent's cooperation probability will change. For example, if by defecting an agent finds its payoff of three units improved with respect to the last time-step, its probability of cooperating will decrease with intensity relative to 3[7].

(2) Normative Drive: once the cooperation norm is recognized, agents' decisions are also influenced by the normative drive. The normative drive is affected by the norm salience: the more salient the norm is, the higher the motivation to cooperate.

The agents who cooperated during the first stage of the game can decide to punish/sanction defectors. The punisher and the sanctioner are driven by different motivations. The former punishes in order to induce the future cooperation of others thus expecting a future pecuniary benefit from its acts (Kreps et al., 1982). On the other hand, the sanctioner is driven by a normative motivation: he sanctions to defend the norm, thus favouring the generation and spreading of norms within the population. Given these differences, the probabilities governing the decision of punishing or sanctioning are modified by different factors and change in the following way:

(1) Punishment Drive: Agents change their tendency to punish on the basis of the relative number of defectors with respect to the last round. If the number of defectors has increased, agents' motivation to punish will decrease accordingly.

(2) Sanction Drive: Agents change their tendency to sanction on the basis of the norm salience. The more salient the norm is, the higher the probability that agents will sanction defectors.

---

[6] These values have been extracted from Cialdini et al. (1990).

[7] If the marginal reward is 0 (this and last time-step reward are the same), agents would change their strategy with an inertial value in the same direction it last changed its probability.

Experimental Design

To explore the specific effects of punishment and sanction on the achievement of cooperation and their relative costs for maintaining it, we designed the simulation experiments in such a way that agents can use either punishment or sanction, but these two mechanisms cannot coexist in the same population. In this way the effects of these two enforcing practices can be isolated and evaluated separately.

To reduce the search space (and computational costs), some parameters have been fixed in advance[8]. In all the simulations the population was composed of 100 agents, located in a fully connected network[9].

In this work, we are not interested in analyzing the emergence of norms, therefore some agents already endowed with the cooperation norm are initially loaded into the simulation (in the experiments presented in this paper, these agents number 50): we refer to them as holders of norms[10]

Simulation Results

The first experiment focuses on the relative effects of punishment and sanction on the achievement of cooperation.

In Figure 1, the different levels of cooperation obtained by imposing punishment or sanction and in the no punishment condition are shown. The x-axis represents the time-steps of the simulation, the y-axis the cooperation rate.

On the one hand, in the no punishment condition, the cooperation level abruptly decreases. The incentive schemes are structured in such a way that non contributing is the dominant strategy for payoff-maximizers. On the other hand, both types of enforcing strategies – punishment and sanction – increase the cooperation level, with respect to the non punishment condition. This result is expected because once a cost is imposed on the non-cooperative action, contribution becomes the dominant strategy for payoff-maximizers.

However, sanction leads to a quicker and higher cooperation level than punishment. As stated in the previous section, the agents' probability to cooperate is driven by a combination of self-interested and normative motivations. Due to its norm-signalling power, sanction has a stronger effect on the agents' normative motivation than mere punishment. The tandem work of the self-interested and normative motivations allows cooperation to be achieved more quickly and in a more durable way (see the section *What happens when punishing/sanctioning is interrupted?*).

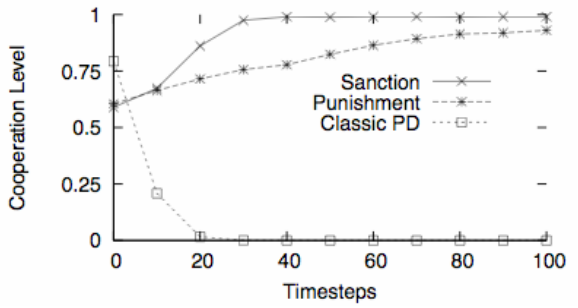

Figure 1: Effects of no Punishment, Punishment and Sanction on the achievement of Cooperation

[8] The initial cooperation probability for all agents is 0,8 and a punishment probability of 0,5.

[9] Different social networks of interaction would definitely produce different dynamics in the system and this will be explored in future works.

[10] The number of agents holding norms varies in each simulation, and they are specified in each figure.

	N° of punishing/sanctioning acts	Global Costs
Punishment	31.221	51.515
Sanction	37.757	62.300

Table 2: Relative Costs of Punishment and Sanction

The simulation experiments shown in Table 2 also provide us with some data on the specific costs of punishment and sanction.

To obtain the levels of cooperation shown in Figure 1, using of sanctions is 20,93% less costly for the system as compared to punishment. In other words, when using sanction, the number of sanctioning acts and consequently the associated costs are reduced by 1/5 (see Table 2). This is an interesting result that confirms our idea that sanctioning combines high efficacy in discouraging defectors with lower costs for society as compared to punishment.

What happens when punishing/sanctioning is interrupted?

This experiment is aimed at testing the hypothesis that sanction makes the population more resilient to change than if it were enforced only by mere punishment. Our hypothesis is that if defection becomes an attractive option, for example because it is very unlikely that defectors are discovered or because enforcement is suddenly interrupted, defectors will take longer to invade the population in which sanction has been used. In this population a larger number of agents have recognized that there is a cooperation norm with respect to the population enforced by mere punishment, and this normative elicitation has the power to activate and make stronger their normative motivation. This happens because of a refraining effect on the decision to abandon the cooperative strategy when it is no longer an attractive option.

To recreate a situation with no enforcement, after the time-step 600 of the simulation, the possibility to punish/sanction defectors has been deactivated.

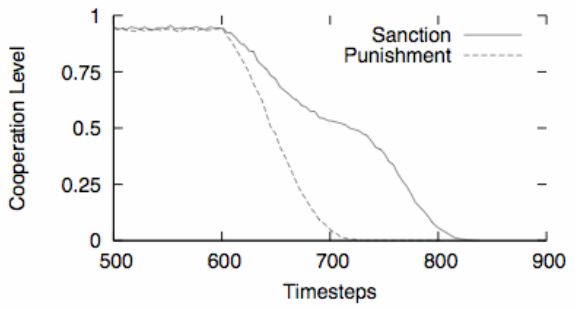

Figure 2: No punishment and sanction after timestep 600

Figure 2 indicates that, when enforcement is suddenly interrupted, agents enforced by sanction continue to comply with the norm for a longer period compared to agents enforced only by punishment. The explanation of this phenomenon is again in the close relationship between sanctions (executed, observed and received) and their impact on the norm's salience. Agents having in mind highly salient norms of cooperation continue to cooperate for a while even in the absence of deterrent penalties. One of the main advantages of this inertial effect of sanction is that policy makers and system designers can take advantage of this delay in order to reestablish the state of the system.

Conclusions and future work

In this paper we have presented a cognitive model that contributes to the understanding of enforcing strategies. In particular, we have distinguished between punishment and sanction, pointing out the different ways in which these strategies aim to influence people's conduct. On the one hand, when it is purely material, punishment only elicits people's motivations to avoid costs; on the other hand when a normative request is also expressed, the motivation to follow the norm as an end in itself is also elicited. We then described a normative agent architecture, whose behaviour is driven by a combination of cost-avoidance and normative motivations. Finally, we presented some simulation results aimed at comparing and clarifying the specific ways in which punishment and sanction affect the achievement and maintenance of cooperation. In particular, those results seem to support our hypothesis that sanction is more effective than punishment in (a) promoting cooperation, (b) reducing the costs for cooperation to be achieved and maintained and (c) making the population resilient to environmental change - e.g. an abrupt interruption of the enforcement mechanism.

To our knowledge, the work presented here is the first simulation study that focuses specifically on this topic. Clearly, further experimental research is necessary to fine-tune some of the values set in the simulation model – such as those related to norm salience. Furthermore, it would be desirable to compare our simulation results with natural and experimental data. This is be part of a larger cross-methodological project on social norms and punishment that we are currently developing. Finally, an interesting venue for future research would be to include an evolutionary mechanism allowing agents to dynamically calculate which is the optimal amount of punishment or sanction to impose in order to obtain compliance.

References

Andrighetto, G., Villatoro, D., & Conte, R. (2010). Norm internalization in artificial societies. AI Communications. Volume 23 Issue 4, pp. 325-339.

Becker, G. S. (1968). Crime and punishment: An economic approach. The Journal of Political Economy, 76(2), 169–217.

Bicchieri, C. (2006). The grammar of society: The nature and dynamics of social norms. New York: Cambridge University Press.

Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. Journal of Behavioral Decision Making, 22(2), 191–208.

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. Science, 328(5978), 617–620.

Boyd, R., & Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethology and Sociobiology, 13(3), 171–195.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. Journal of Personality andSocial Psychology, 58(6), 1015-1026.

Csibra, G. & Gergely, G. (2009). Natural pedagogy. Trends in Cognitive Sciences, 13, 148-153.

Deci, E. L., & Ryan, M., R. (2000). The ”what” and ”why” of of goal pursuits: Human needs and the self-determination of behaviour. Psychological Inquiry. Fehr, E., & Gachter, S. (2002). Altruistic punishment in humans. Nature, 415, 137–140.

Faillo, M, Greco, D. & Zarri, L., (2010) Legitimate Punishment, Feedback, and the Enforcement of Cooperation, Working Paper N. 16/2010, Economics Department, University of Verona.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. Nature 415: 137-140.

Galbiati, R., & D'Antoni, M. (2007). A signalling theory of nonmonetary sanctions. International Review of Law and Economics, 27, 204-218.

Galbiati, R., & Vertova, P. (2008) Obligations and Cooperative Behaviour in Public Good Games. Games and Economic Behavior. 64(1), pp.146-170.

Giardini, F., Andrighetto, G., & Conte, R. (2010). A cognitive model of punishment. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* Austin, TX: Cognitive Science Society, pp. 1282-1288.

Gintis, H. (2003). The Hitchhiker’s Guide to Altruism: Gene-culture Coevolution, and the Internalization of Norms. Journal of Theoretical Biology, 220(4), 407–418.

Gneezy, U., & Rustichini, A. (2000). Pay Enough or Don’t Pay at All. The Quarterly Journal of Economics, 115(3), 791–810.

Herrmann, B., Thoni, C., & Gachter, S. (2008). Antisocial Punishment Across Societies. Science, 319(5868), 1362–1367.

Hirschman, A. O. (1984). Against parsimony: Three easy ways of complicating some categories of economic discourse. American Economic Review, 74(2), 89-96.

Houser, D., & Xiao, E. (2010). Understanding context effects. Journal of Economic Behavior & Organization, 73(1), 58–61.

Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. Journal of Economic Theory, 27(2), 245 - 252.

Masclet, D., Noussair, C., Tucker, S., & Villeval, M.-C. (2003, March). Monetary and nonmone- tary punishment in the voluntary contributions mechanism. American Economic Review, 93(1), 366-380.

Noussair, C., & Tucker, S. (2005). Combining monetary and social sanctions to promote cooperation. Tilburg University.

Sigmund, K. (2007). Punish or perish? Retaliation and collaboration among humans. Trends in Ecology & Evolution, 22(11), 593–600.

Sunstein, C. R. (1996). Social norms and social roles. Columbia Law Review, 96(4), 903–968.

Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. Proc Natl Acad Sci U S A, 102(20), 7398–7401.


European
University
Institute
European University Press