

"ECONOMICS OF CULTURE, INSTITUTIONS, AND CRIME"

Hosted by **Fondazione Eni Enrico Mattei (FEEM)**

Supported by

FP6 Priority 7 "Citizens and governance in a knowledge-based society" Project: Sustainable Development in a Diverse World"(SUS.DIV) (Contract No. CIT3-CT-2005-513438)

University of Padua Research Project "Economic analysis of crime and social interactions"
(grant CPDA071899)

Fondazione Eni Enrico Mattei (FEEM)

Organized by

SUS.DIV, FEEM, University of Padua and CEPR

Milan; 20-22 January 2010

"Thou shalt not covet ...": Prohibitions, Temptations and Moral Values

Matteo Cervellati and Paolo Vanin

We are grateful to the following institutions for their financial and organizational support: SUS.DIV, FEEM, University of Padua and CEPR.

The views expressed in this paper are those of the author(s) and not those of the funding organization(s) or of CEPR, which takes no institutional policy positions.

“THOU SHALT NOT COVET ...”: PROHIBITIONS, TEMPTATIONS AND MORAL VALUES*

Matteo Cervellati
University of Bologna
IZA Bonn

Paolo Vanin
University of Bologna

January 6, 2010

Abstract

We propose a theory studying temptation in presence of both externally and internally sanctioned prohibitions. We characterize the conditions under which agents face temptations that, if not resisted with costly self-control, lead to payoff dominated choices. Moral values that (internally) sanction prohibited actions and their desires may increase utility by reducing self-control costs, thereby serving as partial commitment devices. Moral values may also involve costs in terms of foregone profitable prohibited opportunities and guilt. We apply the model to crime and study the conditions under which agents would optimally adhere to such moral values of honesty. The analysis shows that incentives to be moral are non-monotonic in the expected crime premium. Larger external punishments increase temptation and the demand for morality so that external and internal sanctions are complements. The model helps rationalizing stylized facts that proved difficult to explain with available theories.

Keywords: Prohibitions, Temptation, Self-Control, Moral Values, Crime.

*We would like to thank Giacomo Corneo, Ernesto dal Bó, Davide Dragone and seminar participants at ASSET Florence, EEA-ESEM Barcelona, BOMOPA Modena, Free University of Berlin, SIDE-ISLE Bologna, Universities of Pavia and Padova. Contacts: m.cervellati@unibo.it, paolo.vanin@unibo.it.

1 Introduction

We propose a theory studying temptation and self-control in presence of both externally and internally sanctioned prohibitions.¹ A peculiar feature of external sanctions is that they are typically delayed in time and uncertain since they are contingent upon detection. We investigate the possibility that agents with self-control problems might be tempted to break a prohibition by the immediate returns yielded by the violation, although they understand that it does not pay off in the long run, once expected sanctions are taken into account. Moral values, if internalized, attach a cost of guilt to violation of moral prescriptions. Unlike external punishment, moral costs are suffered immediately, upon act or even desire.² Most cultures (and religions) explicitly attach moral sanctions to many prohibited, and often formally illegal, actions. For instance, the Bible features commandments against theft and adultery, and analogous moral prescriptions can be found in different cultures. The Bible does not only prohibit theft and adultery, however, but also the *desire* of other people's goods and wife.³ This suggests that moral values are often a response to temptation problems (why otherwise prohibiting desires?). At the same time, the association between prohibitions and temptation appears a relevant, or even pervasive, problem. The link between prohibitions, temptations and moral values has, however, essentially been overlooked in economics.⁴

Technically, we study the interaction between external prohibitions, temptation and moral values by considering rational individuals with self-control problems. Temptation results from immediate urges at the moment of action, which may induce payoff dominated choices and can be resisted by exerting costly self-control. Agents would be better off if they could exclude all tempting options from their choice set. In practice, however, this may not be feasible, but tempting options can be made more costly. We propose the view that this is precisely what moral values do, so that they serve as a 'partial' commitment device.⁵ After characterizing exposure to temptation in presence of heterogenous returns to the prohibited actions, we study individual incentives to adhere to a

¹We discriminate between external sanctions, whether legal or social, and internal sanctions, which are not inflicted by any external party and which we relate to moral punishments associated to the emotion of guilt.

²We discuss below the possible interaction between moral and social norms.

³The commandment against actions is: 'You shall not commit adultery. You shall not steal.' (Exodus, 20:12). The prohibition of desire is: 'YOU SHALL NOT COVET your neighbor's house; you shall not covet your neighbor's wife, nor his male servant, nor his female servant, nor his ox, nor his donkey, nor anything that is your neighbor's.' (Exodus, 20:13). The title of this paper comes from the ancient English version of this commandment.

⁴We discuss the existing evidence and related theoretical literature in Section 2.

⁵While changing own moral values to avoid experiencing guilt appears possible, this is generally a long, and often hard, process. Moral values can therefore effectively reduce self-control problems, since they are stable over the shorter time horizon of the urges of temptation.

moral code sanctioning those actions and their desire. Specifically, we frame the problem in terms of individual choices between a legal and an illegal (or criminal) action.⁶ While this is relevant *per se* and helps drive intuition, the applicability of the theory is wider.

With heterogeneous returns, if crime premia are either very low (possibly negative) or very high, then legality and crime, respectively, are the unequivocally optimal choices and no temptation problems arise. For intermediate levels of crime premium agents are tempted and face a trade-off between the costs of self-control and the payoff costs of giving in to temptation. By attaching guilt costs to crime, moral values of legality modify this trade-off and can allow to resist temptation, thereby increasing utility. To this end, they must sanction both actions *and* desires. Morality also imposes a trade-off, however. It saves on self-control costs, but it may also prevent to take advantage of high return illegal opportunities and it exposes agents to the cost of guilt in case they do not resist temptations and give in to illegal and immoral behaviors. Adhering to a system of moral values (or receiving a moral education) is therefore optimal only for those expecting to face temptation and, more importantly, expecting to resist it in equilibrium. The model therefore provides an answer to the question of why individuals would voluntarily adhere to a morality that may involve utility costs. The model also highlights two potential sources of temptation: attraction by immediate returns, which can operate even in the absence of any morality, and morality itself. Subtly, moral prohibitions have therefore both the potential to alleviate temptation and to induce in temptation.

Despite its simplicity the proposed framework delivers a rich set of predictions. First, individuals with intermediate crime premia are those who benefit the most from moral values.⁷ Second, while legal enforcement reduces crime, it also raises self-control problems and correspondingly increases the scope for moral values of honesty.⁸ So legal and moral enforcement turn out to be complements. Third, for most tempted individuals it is optimal to adhere to moral values if and only if these are sufficiently strict. The reason is that a weak morality would not be sufficient to resist temptation, and it would only add the cost of guilt to that of forgone payoffs (in this sense morality is a risky asset). Finally, in presence of random returns, temptation and morality may lead to two types of payoff dominated choices: agents may give in to (payoff dominated) tempting choices, but they may also abstain from pay-off dominant (but morally sanctioned) opportunities. The proposed framework

⁶We use henceforth crime and illegality as synonyms, without any further distinction.

⁷If crime premia were decreasing in socio-economic status, we would have that poorest people reject moral codes of legality, while richest people are amoral. The middle class would be the one that mostly gains from morality, with the lower-middle class being morally stricter than the upper-middle class.

⁸This result hinges upon the idea that temptation depends on the difference between static and dynamic optimality, and that, when future punishment is contingent upon detection, a higher detection probability widens this difference.

therefore helps reconciling two puzzling empirical regularities in crime economics: there are crimes which are not committed, although their expected economic payoff is clearly positive, and at the same time there are crimes that are committed, although their expected return is clearly negative.⁹

The remainder of the paper is organized as follows. In Section 2 we discuss the contributions of the paper to the related literature. In Section 3 we present our model, which is analyzed in Section 4. In Section 5 we discuss the role of our simplifying assumptions, the implications of the analysis and some possible extensions.

2 Related Literature

This paper contributes to the literature on temptation, crime and moral values. Our modeling of temptation follows the approach of Gul and Pesendorfer (2001), where the cost of self-control is correctly anticipated and individual choices are time consistent.¹⁰

The economic approach to crime, pioneered by Becker (1968), looks at it as a rational choice, based on expected costs and benefits. In an early and influential paper, Ehrlich (1973) proposes to frame crime as a risky occupational choice, essentially driven by long run concerns. In this perspective temptation does not appear to be a crucial issue. Accordingly, the subsequent crime economics literature based on this approach essentially disregards it. Nonetheless, most illegal actions and crimes are not performed by professional criminals and, as discusses next, temptation appears a relevant and widespread problem in a large set of illegal and criminal behavior.

In sharp contrast to economists, criminologists have long considered temptation (and thus short run concerns) as one of the deep roots of crime. Gottfredson and Hirschi (1990) argue that ‘Nearly all crimes are mundane, simple, trivial, easy acts aimed at satisfying desires of the moment [...] When [offenders’] desires conflict with [their] long term interests, those lacking self-control opt for the desires of the moment, whereas those with greater self-control are governed by the restraints imposed by the consequences of acts displeasing to family, friends and the law’ (p. xv).¹¹

⁹The literature has pointed out striking regularities of both types of behavior which cannot be easily rationalized within available economic theories of crime as discussed in Section 2.

¹⁰See Dekel et al. (2009) and Stovall (2009) for the most recent developments of this approach. Internal conflicts may also be represented by time inconsistent preferences, as in Strotz’(1955) seminal paper and in Laibson’s (1997) analysis of hyperbolic discounting. Benhabib and Bisin (2005) offer a neuroeconomics approach to temptation.

¹¹As we do here, they identify the roots of temptation in the fact that, ‘Whereas the pleasure attained by the [criminal] act is direct, obvious and immediate, the pains risked by it are not obvious, or direct, and are in any event at greater remove from it’ (p. 95).

Since Becker (1968), most of crime economists' attention has been devoted to external sanction mechanisms, whether legal (such as fines and jail) or social (such as stigma). Relatively little attention has been paid, by contrast, to moral enforcement mechanisms. Polinsky and Shavell (2000) review the economic theory of public enforcement of law. Glaeser et al. (1996), Rasmusen (1996), Funk (2004) and Weibull and Villa (2005), among others, provide crime theories based on peer pressure, stigmatization and internalized social norms, which emphasize the potential for multiple equilibria. Conley and Wang (2006) treat moral values as exogenous preference traits.¹²

To our knowledge, there are no economic studies of the link between temptation, moral values and crime. While the traditional economic approach to crime has greatly improved our understanding of the phenomenon, its neglect of the role of temptation and morality may represent a limit for the empirical relevance of positive theories (and, hence, for their normative implications).¹³ This paper contributes to the literature a simple rational model of crime, which incorporates temptation-driven rational choices coupled with an analysis of role of moral values as self-control devices. The predictions of the model appear empirically relevant and help reconcile rational economic theories of crime with empirical stylized facts, which are puzzling for available theories. In particular, they offer a natural explanation for the stylized fact that several crimes are not committed despite having clearly positive expected returns, while many crimes are committed despite being clearly payoff dominated.¹⁴ Although the focus of the present contribution is not normative side, our model also offers interesting predictions on the interactions between internal and external enforcement.

Beyond the specific crime application, our paper also contributes to the literature studying the

¹²Some works find evidence for the role of social pressure and moral/social norms for criminal choices. For instance, Chamlin and Cochran (1997) and Buonanno, Montolio and Vanin (2009) find a crime reduction effect of altruistic norms, proxied by charitable donations and by blood donations, respectively. Buonanno, Pasini and Vanin (2008) find evidence that the strength of social sanctions, proxied by social interaction density, significantly reduces crime.

¹³Dills et al. (2008) conclude their critical assessment of the empirical literature arguing that economists still 'know little about the empirically relevant determinants of crime'. Approaches based on multiple equilibria, such as those proposed by the literature on social determinants of crime, may provide part of the explanation. Temptation and moral values may offer another, possibly complementary part.

¹⁴Calculating the expected return and punishment of tax evasion, Becker and Posner (2007) argue in their blog, available at <http://www.becker-posner-blog.com/archives/2007/11>, that 'if taxpayers responded only to the expected cost of evading taxes, evasion would be far more widespread'. They argue that risk aversion does not appear sufficient to explain these observations, which may rather be imputed to the fact that 'most people do not believe it is moral to steal money even when there is little chance they will be found out'. Coricelli et al. (2007) provide experimental evidence that induced shame on evaders does, indeed, favor compliance. Concerning the observation of payoff dominated crimes, we refer to Gottfredson and Hirschi (1990) for an extensive discussion of several documented instances of seemingly irrational crimes committed by individuals facing temptation.

role of moral values, and the interaction between external and internal (moral) enforcement devices. Moral values are modeled as working through self-regulative emotions of guilt, which reduce utility in case of deviation from morally appropriate standards of behavior.¹⁵ The works of Lindbeck, Nyberg, and Weibull (1999) and Weibull and Villa (2005) study the role of internalized social norms that change behavior by inducing guilt (and/or shame). These papers consider individuals exogenously endowed with internalized social norms, where the utility cost of stigma depends on the share of individuals behaving in line with the norm.¹⁶ Concerning the rationale for their existence, moral values may increase well being by improving cooperation, as in the theories by Frank (1987) and Tabellini (2008), among others. Dal Bó and Dal Bó (2009) provides experimental evidence for the effects of moral suasion on cooperation. Moral values may also be welfare improving in presence of externalities, as studied by Kaplow and Shavell (2007), among others.¹⁷

Differently from these works, our temptation framework allows to study individual incentives to endogenously adhere to a system of moral values, even when morality may be costly to its carrier (in terms of guilt or of forgone payoffs).¹⁸ Accordingly, moral values can be self-sustaining even in the absence of external enforcement, such as shame induced by social pressure. Despite the different focus, we share with Corneo and Jeanne (2009) the idea that adhering to a value system may be a choice.¹⁹ Dal Bó and Terviö (2008) provide a dynamic analysis of the endogenous evolution of moral capital in terms of self-image about own moral “type”. While we focus on morality as a partial commitment device rather than on the evolution of the self-image in terms of morality, we share with this paper the focus on prohibitions and moral values.

Some recent works investigate the emergence and/or transmission of different ethics or moral

¹⁵This approach essentially follows the literature in Social Psychology. According to Crocker and Park (2003) moral values are effective in changing individual actions by inducing the immediate experience of self-regulatory emotions such as guilt or shame (which tend to reduce self-esteem) for behaviors which are not in line with moral standards.

¹⁶Similarly to these works in our model the moral cost enters the utility function additively (the consideration of a non-additive utility formulation leaves the main predictions unchanged and it is relegated to the Appendix). Cervellati, Esteban and Kranich (2006) presents a theory where the comparison between moral standards (work ethics) and own behavior also induces guilt and affects self-esteem. In that paper, however, it is crucial that self-esteem enters the utility function non-additively.

¹⁷Cervellati, Jansen and Sunde (2008) argue that moral (and more specifically religious) values may also help improving utility in presence of uncertainty and market incompleteness.

¹⁸In a related way, Rotemberg’s (1994) asks why individuals would choose to be altruistic towards colleagues, when altruism is potentially costly. While his answer is based on strategic complementarity, ours is rooted in an individual decision problem, with temptation and self-control.

¹⁹Differently from them we are interested in temptation and moral values rather than symbolic values and tolerance. As in their work, however, our costs of guilt can be interpreted as a reduction in self-esteem.

norms. Lindbeck and Nyberg (2006) explicitly study why parents should seek to instill in their children work values which are sustained by guilt. We share with their analysis the focus on moral values as commitment devices, although our case is about self-commitment rather parents/children commitment problems and we focus on temptation rather than time inconsistency. Tabellini (2008) studies the adoption and transmission of values of pro-social behavior and values of generalized morality in a cultural transmission framework *à la* Bisin and Verdier (2001). Adriani and Sonderegger (2009) show that parents may want to socialize their children to pro-social behavior even if they do not derive direct utility from their children values. While we do not explicitly study the issue of moral setters, our model provides predictions on the differential incentives to adhere to moral codes of honesty for individuals (or families) facing different expected returns from legal and illegal options.²⁰

Finally, the paper contributes to the literature studying the relationship between different types of enforcement of prohibitions. Several works, such as Posner (1997) and Aldashev et al. (2009), study the interactions between different types of external sanctions: legal and social. Our model, in turn, studies the link between external and internal sanctions. It predicts that stronger external punishment may increase the demand for moral values, since they are likely to increase individual exposure to temptation. External sanctions may therefore also increase the demand for “stricter” moral values of honesty.

3 The model

Self-control preferences are represented in terms of an internal conflict between a ‘commitment’ ranking, reflecting preferences under the possibility of commitment, and a ‘temptation’ ranking, reflecting immediate urges at the moment of action which may induce payoff dominated choices (according to the ‘commitment’ ranking). Temptation may be resisted, but only at a cost of self-control. The decision process is represented in terms of a planning and an action stage. When evaluating things ‘coldly’ at the planning stage an individual knows that his interests are best served by a given choice, namely the one that maximally satisfies his ‘commitment’ ranking, but that he might be tempted by a different choice, the one that is best for his ‘temptation’ ranking. At the planning stage he may therefore look for commitment devices, that is, restrictions in his action stage choice set, which exclude or at least make more costly tempting options.

²⁰We discuss on this point in more details in Section 5.

3.1 Self control preferences

Formally, consider a rational individual with self-control preferences, as axiomatically characterized by Gul and Pesendorfer (2001).²¹ Such individual is otherwise standard, but for the fact that he may be subject to temptation. Let \mathcal{A} be the compact set of partial commitment devices. Each element $A \in \mathcal{A}$ is assumed to be a compact set of alternative options. Preferences are defined on the set of compact subsets of lotteries Δ .²² The relationship between individual choices and outcomes is defined by a (probabilistic) technology, represented by a function mapping choices into lotteries, $F : \mathbb{A} \rightarrow \Delta$. Individuals take F as given.²³ The decision process is represented in terms of a planning and an action stage. At the planning stage the agent chooses an element $A \in \mathcal{A}$; at the action stage he chooses an element $a \in A$.

Self-control preferences reflect the conflict between the commitment and the temptation ranking. Such rankings are represented by two different continuous and linear (expected) utility functions, representing commitment and temptation utility, and denoted $u : \Delta \rightarrow \mathbb{R}$ and $v : \Delta \rightarrow \mathbb{R}$, respectively. Given a compact subset $C \subseteq \Delta$, the cost of self control associated to one of its element, $x \in C$, is assumed to be equal to the distance between x and the most tempting alternative (in C), in the temptation utility space: $[\max_{y \in C} v(y) - v(x)]$. This implies that the costs of self control at the action stage increase with the utility loss which is experience by not choosing the preferred action according to the temptation ranking. The (expected) utility associated to x is therefore simply given by the difference between the commitment utility and the costs of self-control:

$$u(x) - \left[\max_{y \in C} v(y) - v(x) \right] \quad (1)$$

For future reference we refer to (1) simply as expected utility to distinguish it from the commitment and the temptation (expected) utility. Given $A \in \mathcal{A}$, the solution of the action stage problem involves the choice of the optimal action and can be stated as,

$$U(F(A)) = \max_{a \in A} \left\{ u(F(a)) - \left[\max_{b \in A} v(F(b)) - v(F(a)) \right] \right\}. \quad (2)$$

²¹These are preferences over sets of alternatives, satisfying a number of standard axioms (completeness, transitivity, continuity and independence), plus a ‘set betweenness’ axiom. Gul and Pesendorfer (2001) show that these axioms are characterized by the representation in terms of commitment and temptation ranking. Agents facing, and resisting, temptation sustain self-control costs and restricting the choice set in by eliminating suboptimal options make them strictly better off. Self-control preferences therefore naturally generate a demand for commitment devices.

²²For future reference, let $\mathbb{A} = \bigcup_{A \in \mathcal{A}} A$ be the set of all possible options, Z be the set of all possible outcomes and Δ be the set of all possible lotteries over Z . A lottery is a probability distribution over outcomes.

²³We assume that, for any $A \in \mathcal{A}$, its image $F(A) \subseteq \Delta$ is compact. We say that a lottery $x \in \Delta$ is feasible if there exists $a \in \mathbb{A}$ such that $x = F(a)$.

Finally, the individual problem at the planning stage concerns the choice of a compact subset of alternative choices:²⁴

$$\max_{A \in \mathcal{A}} U(F(A)). \quad (3)$$

where $U(F(A))$ is the indirect expected utility obtained at the action stage, given the partial commitment device A .²⁵

3.2 Prohibitions, Temptation and Moral Values

Consider an individual who at the action stage has to choose between two actions. These actions differ in terms of their returns as well as for the fact that one is prohibited and externally sanctioned. For illustration, we call the first action legal and the second one illegal (or criminal).²⁶ We investigate the role of prohibitions as source of temptation, and of moral values as a partial commitment device. Moral values are assumed to involve internally, rather than externally, imposed sanctions in the form of utility losses (due to e.g. to the experience of guilt or reduction in self-esteem for mis-behavior).

At the planning stage, the agent has to choose between adhering and rejecting a (received) moral code. The (action stage) choice sets in these two cases are denoted by A_μ and A_0 , respectively. At the action stage an agent has to choose between committing crime or behaving legally. Choosing to commit crime with probability $a \in [0, 1]$ in set A_μ and A_0 are framed as different options since the two choices map into different lotteries and materialize in different levels of expected utility.²⁷ The set of outcomes includes returns to each action, together with external and moral sanctions. As a benchmark we consider a linear additive separable utility formulation.²⁸ The commitment expected

²⁴Gul and Pesendorfer (2001) propose the example of a vegetarian restaurant chosen to avoid temptation to eat meat. The choice of the restaurant in this example works as a partial commitment device for the agent by restricting his choice set. Since a menu contains more than one possible meal, such commitment is only partial.

²⁵As in Gul and Pesendorfer (2001), it is assumed that preferences over sets of alternatives are determined by the best element in each set (where ‘best’ is meant in terms of expected utility). The expected utility associated to the subset $C \subseteq \Delta$ is given by the continuous and linear real-valued function $U(C) = \max_{x \in C} \{u(x) - [\max_{y \in C} v(y) - v(x)]\}$ defined on the set of compact subsets of Δ . This formulation is equivalent to (2). We find it more convenient to frame our application directly in terms of action choices rather than lotteries.

²⁶More generally the framework applies to any problem where some action involve some potential expected utility loss which is externally imposed on the individual., be it due to monetary or non-monetary sanctions.

²⁷As will become clear later, it is immaterial to allow for mixed strategies. It simply saves notation.

²⁸The separable formulation is mainly chosen to simplify the exposition. In Section 6 we show that the main predictions do not depend on the specific assumptions on the utility formulation. In Section 5 we discuss the role of possible interaction between external and internal sanctions.

utility associated to choosing crime with probability a is given by²⁹

$$u(a) = r(a) + \pi(a) + \mu(a) \text{ for any } a \in A_0 \cup A_\mu. \quad (4)$$

Denote the returns for the legal and the illegal activity as r_w and r_c , respectively so that the expected return is $r(a) = ar_c + (1 - a)r_w$ for any $a \in A_0 \cup A_\mu$. External sanctions are conditional upon having committed a crime. We denote by π the expected punishment for a criminal activity so that $\pi(a) = a\pi$ for any $a \in A_0 \cup A_\mu$.³⁰ Finally, moral sanctions are also conditional upon having committed a crime, but differ from externally enforced sanctions in that they do not require external detection. For those who adhere to the received moral code (i.e. if $a \in A_\mu$), the moral sanctions attach a utility cost $\mu > 0$ to the choice of crime; for those who rejected the received morality (i.e. $a \in A_0$), such cost is absent. This implies that $\mu(a) = a\mu$ if $a \in A_\mu$ while $\mu(a) = 0$ if $a \in A_0$.

The temptation ranking differs from the commitment ranking for an underestimation of the disutility from external punishment,

$$v(a) = r(a) + \lambda\pi(a) + \mu(a) \text{ for any } a \in A_0 \cup A_\mu \quad (5)$$

where $\lambda \in [0, 1]$, parameterizes the exposure to self-control problems. The lower is λ the lower the weight attached to punishment in the temptation ranking and the more serious is the underestimation of the costs of crime.

The main difference between external and moral punishment for mis-behavior is assumed to rely on the asymmetric under-evaluation of the two types of punishment. This modeling strategy appears reasonable also in view of the fact that external punishment are typically deferred in time. By contrast, the moral feeling of guilt may be experienced immediately upon act, and even upon desire.³¹ The assumption that moral costs enter in both the commitment and temptation ranking is crucial for our argument. The fact that they enter exactly in the same way is not important, but it allows a clear understanding of their effects even under the assumption of separable utility. We discuss in Section 5 the implications of assuming that moral values only concerns actions and not to desire, or the other way around.

²⁹Notice that with a slight abuse of notation we write $u(a)$ to indicate $u(F(a))$.

³⁰This can also be thought as $\pi = dp$, that is as resulting from the probability that a crime is detected, d , and the associated punishment p .

³¹Indeed, the existence of moral values prohibiting desires, together with the prohibition of action, seems coherent with this view as discussed below.

4 Equilibrium

Let us start from a standard case, in which there are neither temptation nor moral values. Let $r \equiv r_c - r_w \geq 0$ be the (gross) utility premium. The net premium is therefore $(r - \pi)$. In the absence of temptation the optimal choice involves the maximization of the commitment utility so that legality is chosen whenever $r < \pi$. Crime is optimal only when the expected punishment more than compensates the crime premium. Notice that, absent temptation, nobody would have an incentive to adhere to a sanctioning moral code, since this simply shifts the utility frontier inwards.

4.1 Action Choices in the Presence and Absence of Morality.

Let us first consider temptation, in the absence of moral values that is for any $a \in A_0$. Recall that a denotes the probability to commit a crime and that for $a \in A_0$, the commitment and temptation utility are given by $u(a) = (1 - a)r_w + a[r_c - \pi]$ and $v(a) = (1 - a)r_w + a[r_c - \lambda\pi]$, respectively.

The existence of temptation and costs of self-control imply that the indirect utility changes non monotonically in the crime premium. The following Figure depicts the indirect utility $U(A_0)$ with a bold line (where for illustration we normalize $r_w = 0$, so that $u(0) = v(0) = 0$ and $r = r_c$). The commitment and temptation utility from crime $u(1)$ and $v(1)$ are also depicted. Notice that, since the only difference between u and v is an under-evaluation of the external punishment, it always holds that $v(1) \geq u(1)$. For low crime premium, $r < \lambda\pi$, legality represents the best option according to both rankings (since $v(1) < v(0)$ and $u(1) < u(0)$) so that agents do not suffer from temptation and their indirect utility is independent of the crime premium. In the range $r \in [\lambda\pi, \pi]$ the agents suffer from temptation since $u(1) < u(0) = v(0) < v(1)$. Crime does not represent the best choice according to the commitment ranking but is the best choice according to the temptation ranking. The cost of self control is given by the difference between between $v(0)$ and $v(1)$. In this "temptation range", $r \in [\lambda\pi, \pi]$, the indirect utility is initially decreasing in r since the temptation is resisted but at an increasing cost in terms of self control. Eventually (when $r = (1 + \lambda)\pi/2$) this self-control cost is too large and for the agent giving in to temptation is be best option although it implies a suboptimal choice from the point of view of the commitment ranking. An even larger r is associated to an increasing indirect utility due to the increase in $u(1)$. Finally, when the premium is large enough, $r > \pi$, then crime is the rational choice also according to the commitment ranking and the indirect utility keeps increasing with r .

These action stage choices and temptation problems in the absence of moral values are illustrated in the following Figures, as a function of the crime premium.

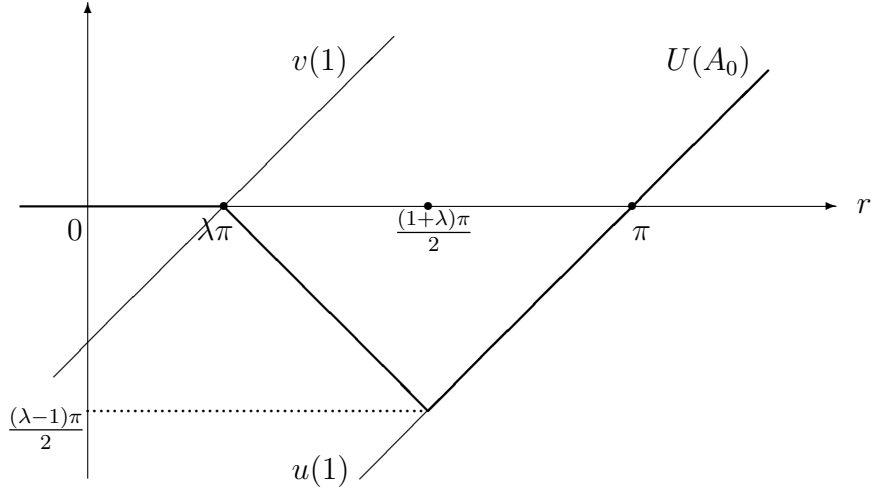


Figure 1: Indirect Utility as Function of Crime Premium

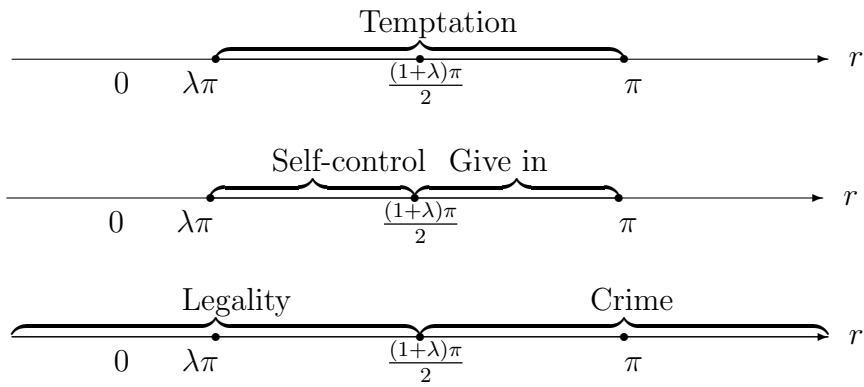


Figure 2: Ranges of Temptation, Self-Control and Legality as function of the crime premium

Let us now consider an individual who, at the action stage, is endowed with moral values of legality, that is has to choose $a \in A_\mu$. Recall that μ represents the moral punishment associated to the illegal action. The commitment and temptation utility are given by $u(a) = (1-a)r_w + a[r_c - \pi - \mu]$ and $v(a) = (1-a)r_w + a[r_c - \lambda\pi - \mu]$, respectively. Notice that in both rankings the expected utility from crime is reduced by the moral punishment. In fact the analysis in the absence of moral values is formally identical to that the one with moral values for $\mu = 0$. This implies, in particular, that the indirect utility displays a non-monotonicity in both cases. Notice, however, the thresholds level of r leading to temptation and optimal choices is shifted by μ . This discussion is summarized in,

Proposition 1. (*Temptation and self-control*) *Let $\mu > 0$ if $a \in A_\mu$ and $\mu = 0$ if $a \in A_0$.*

- *Legality is optimal for both u and v and the agent is not subject to temptation if $r < \lambda\pi + \mu$,*
- *Legality delivers the largest commitment utility $u(0) > u(1)$, but crime is tempting, $v(0) < v(1)$, if $r \in [\lambda\pi + \mu, \pi + \mu]$. In this range:*
 - *Temptation is resisted but at a self-control cost $(r - \lambda\pi - \mu)$ if $r \in [\lambda\pi + \mu, (1 + \lambda)\pi/2 + \mu]$;*
 - *Temptation is overwhelming and the option with the lower commitment utility is chosen if $r \in ((1 + \lambda)\pi/2 + \mu, \pi + \mu]$.³²*
- *Crime is optimal for both u and v and the agent is not subject to temptation if $r > \pi + \mu$.*

Proof. Recall that the action stage problem is $\max_{a \in [0,1]} \{u(a) - [\max_{b \in [0,1]} v(b) - v(a)]\}$, whose solution is equivalent to $\max_{a \in [0,1]} \{u(a) + v(a)\}$ which implies (leaving aside indifference) that:

$$a^* = 1 \iff u(1) + v(1) > u(0) + v(0)$$

Notice that $u(1) > u(0) \iff r > \pi + \mu$ and $v(1) > v(0) \iff r > \lambda\pi + \mu$. Moreover, if $r \in [\lambda\pi + \mu, \pi + \mu]$, then $u(1) + v(1) > u(0) + v(0) \iff r > (1 + \lambda)\pi/2 + \mu$. \square

Agents face temptation for intermediate crime premium r . This temptation is resisted in equilibrium only when the cost of self-control $(r - \lambda\pi)$ is not too high.

From Proposition 1 it immediately follow that crime is chosen for a large enough premium as illustrated in,

³²If $r = \frac{(1+\lambda)\pi}{2} + \mu$, then crime and legality are indifferent.

Corollary 1. (*Equilibrium choices at the Action Stage*) *An individual with a (gross) crime premiums r , chooses crime if and only if $r > (1 + \lambda)\pi/2 + \mu$ (with $\mu = 0$ if $a \in A_0$) and chooses legality otherwise.*

The size of the external punishment π has effects on both the equilibrium level of crime, the temptation range and the indirect utility of agents. First, from Corollary 1, an increase in the external punishment π reduces the range in which agents choose crime. Second, from Proposition 1, since the difference between the commitment and temptation rankings is due to an under-evaluation of the external punishment, an increase in π increases the length of the temptation range, $(1 - \lambda)\pi$.³³ Third, a marginal increase in π reduces the indirect utility for all agents choosing crime. It raises, however, the indirect utility of those resisting temptation since the cost of self-control given by $v(1) - v(0) = (r - \lambda\pi - \mu)$ is smaller. Finally, the inefficiencies associated to temptation gets worse for low λ . A reduction in λ increases both the ranges of temptation and crime as well as the costs of self control.

4.2 Planning stage: Moral values as Partial Commitment

We now turn to the analysis of the role of moral values for crime, temptation and individual indirect utility. A moral cost attached to crime makes it a less attractive option thereby shifting the range of temptation to a higher interval of crime premia. In particular, from Corollary 1, morality triggers legality for $r \in ((1 + \lambda)\pi/2, (1 + \lambda)\pi/2 + \mu)$ since for crime premia in this range, an individual without moral values would choose crime, whereas one with moral values would choose legality. In fact, for any crime premium, there exists a sufficiently large μ , such that legality would be chosen (by a moral individual). The presence of temptation and moral values may induce choices of both legal and illegal activities which are payoff dominated according to the commitment ranking. From Proposition 1, all agents in the range $r \in ((1 + \lambda)\pi/2 + \mu, \pi + \mu]$ give in to temptation and choose illegal activities which are payoff dominated according to the ranking $u(\cdot)$. The existence of a moral cost μ reduces the range under which this payoff dominated choices are made. But if the level of μ is large enough even agents for whom it would be optimal to choose crime in the absence of moral values may choose legal activities which are payoff dominated according to $u(\cdot)$. Consequently, being moral does not necessarily represent the best option at the planning stage since morality involves both costs and benefits. This implies that while the effect of μ on the choice of crime is monotonic, the effect of μ on the indirect utility may be non monotonic. The indirect utility $U(A_\mu)$ for an

³³This could be due to either a higher detection probability d or to a higher punishment p .

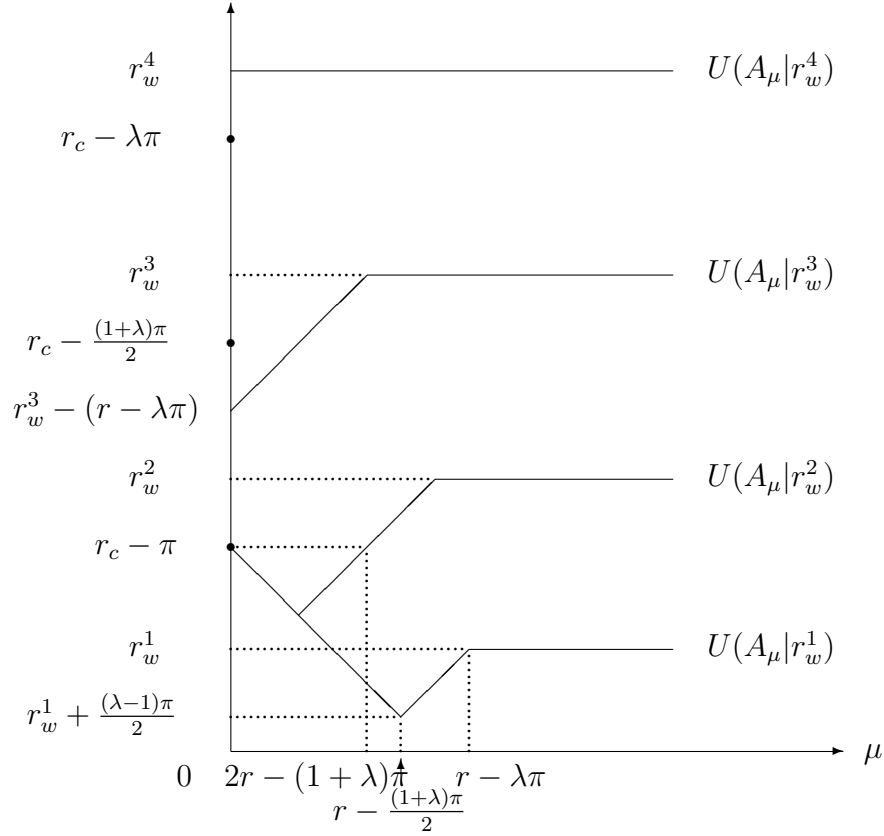


Figure 3: Indirect Utility as function of μ for individuals with different crime premium r

individual who embraced moral values is given by,

$$U(A_\mu) = \begin{cases} r_w & \text{if } r < \lambda\pi + \mu \\ r_w - [r_c - r_w - \lambda\pi - \mu] & \text{if } r \in (\lambda\pi + \mu, (1 + \lambda)\pi/2 + \mu) \\ r_c - \pi - \mu & \text{if } r > (1 + \lambda)\pi/2 + \mu \end{cases}$$

To illustrate the role of moral values, for given levels of r_c, π and λ , consider moral individuals facing different levels of legal returns. The following Figure depicts the indirect utility $U(A_\mu)$ as a function of μ for individuals with different levels of legal returns r_w .

Fix a given r_c and consider an agent with $r_w = r_w^1$, so that $r > \pi$. In the absence of moral values, or $\mu = 0$, this individual faces no temptation and chooses crime. As μ increases, he starts feeling guilty but as long as $\mu < r - \pi$, the moral cost does not change the optimality of crime for both rankings. If $\mu > r - \pi$ then crime is not optimal for the commitment ranking $u(\cdot)$ but as long as $\mu \in (r - \pi, r - \lambda\pi)$ it is still optimal for the $v(\cdot)$ ranking so that individual faces temptation. Notice that for this agent it is precisely the moral prohibition of crime that creates temptation. Nonetheless, stricter moral values progressively alleviate self-control problems since the cost of self-

control ($r - \lambda\pi - \mu$) is decreasing in μ but for any $\mu \in (r - \pi, r - (1 + \lambda)\pi/2)$, self-control is still too costly, and for the individual gives in to temptation and keeps choosing crime. Accordingly, in this range an increase in μ leads to a reduction in indirect utility. For $\mu \in (r - (1 + \lambda)\pi/2, r - \lambda\pi)$, the cost of self-control is low enough and the individual resists temptation and chooses legality, the cost of guilt disappears and cost of self-control decreases, so that in this range $U(A_\mu)$ is increasing in μ . Finally, for $\mu > r - \lambda\pi$, crime stops being tempting and the self-control cost disappears, leaving $U(A_\mu) = r_w$. For agents with $r_w = r_w^2$ so that $r \in ((1 + \lambda)\pi/2, \pi)$ the indirect utility is qualitatively analogous but with $\mu = 0$ such agent is giving in to temptation. If $r_w = r_w^3$, so that $r \in (\lambda\pi, (1 + \lambda)\pi/2)$ then with $\mu = 0$ the agent exerts self control and resists temptation, so that his indirect utility is monotonically increasing. Finally, an agent with $r_w = r_w^4$, so that $r < \lambda\pi$, always chooses the legal option and never faces temptation and the indirect utility is independent of μ .

The non monotonicity of the indirect utility in μ drives the planning stage choice between embracing and rejecting a received moral code. Notice that moral values involve two types of costs. In the first place, if moral values are weak (if μ is low), they may be insufficient to induce an agent to resist temptations. In this case, not only the agent commits crimes which are sub-optimal from the point of view of the commitment ranking $u(\cdot)$, but he also pays a disutility cost for behaving against his morality. A second type of cost arises if a moral agent refrains from illegal actions which would be otherwise optimal for the commitment ranking. Morality makes such agents forgo payoffs. These costs are traded off with the increase in expected utility due to the reduction in the costs of self control. As a result not all agents necessarily gain from the adherence to a system of moral values which punish dishonesty and temptation. From the above analysis it is immediate to characterize the planning stage choice between A_0 and A_μ .³⁴

Proposition 2. (*Planning stage choice: moral values*) *For any given $\mu > 0$, the following holds. An individual:*

- *is indifferent between A_μ and A_0 , if $r \leq \lambda\pi$;*
- *prefers A_μ if $r \in (\lambda\pi, (1 + \lambda)\pi/2)$;*
- *given $r \in ((1 + \lambda)\pi/2, \pi)$, prefers A_μ if $\mu > 2r - (1 + \lambda)\pi$ and A_0 if $\mu < 2r - (1 + \lambda)\pi$;*
- *prefers A_0 if $r > \pi$.*

³⁴In what follows we focus on generic cases in the parameter space and avoid referring explicitly the agents which, by definition, are indifferent at the thresholds.

To fix ideas, we call individuals moral, amoral or immoral if $U(A_\mu)$ is higher, equal or lower than $U(A_0)$, respectively. The adherence to a system of moral values represents a partial commitment device with expected costs and benefits. Agents with a low enough (gross) crime premium $r \leq \lambda\pi$ are amoral. For these agents both the expected costs and benefits of morality are zero since, irrespective of μ , they choose legality and do not face temptation. At the other extreme for those agents with a large crime premium $r > \pi$ crime is the optimal choice in the absence of morality. They are therefore immoral since morality only involves expected costs of either guilt or forgone crime payoffs. For an intermediate crime premium, $r \in (\lambda\pi, \pi)$, agents face temptation in the absence of morality. Adhering to a moral code of honesty is beneficial only if it allows to resist temptation in equilibrium. Notice that all those who would resist temptation even in the absence of morality strictly gain from any $\mu > 0$ thanks to lower costs of self control. This group of agents, $r \in (\lambda\pi, (1 + \lambda)\pi/2)$, is therefore moral for any $\mu > 0$. This is not true for those who, absent morality, would give in to temptation. For them morality is beneficial if and only if it induces to choose legality. Weak moral values, $\mu < 2r - (1 + \lambda)\pi$, which are insufficient to induce an agent to resist temptations only reduce utility. This implies that the agents facing stronger temptations would adhere to moral values if and only if these are sufficiently strict, $\mu > 2r - (1 + \lambda)\pi$.

Let us conclude by briefly commenting on role of μ , λ and π for the choice at the planning stage. From Proposition 2, the set of values of crime premium for which $U(A_\mu) > U(A_0)$, henceforth called morality range, is given by,

$$r \in \left(\lambda\pi, \min \left\{ \pi, \frac{\mu + (1 + \lambda)\pi}{2} \right\} \right) \quad (6)$$

Morality increases individual utility only if it helps resist temptations. Consequently, the set of r for which morality is preferred strictly increases in μ up to the point in which nobody is tempted which is case for $\mu = (1 - \lambda)\pi$ (since for larger μ the range (6) is given by $r \in (\lambda\pi, \pi)$).³⁵ An increase in temptation due to either a reduction in λ or an increase in π , expands the morality range. Consequently, even if external and moral punishment affect utility additively, an increase in external punishment increases the demand for morality. This result may appear counterintuitive at first sight. The intuition is, however, straightforward once one considers the role of increased external punishment on temptation. Increased external punishment has two main effects. On the one hand they reduce the returns to illegal activities and crime in equilibrium. On the other hand they increase the range of temptation and this makes moral values of honesty even more valuable from an individual perspective.

³⁵The threshold for which all agents facing temptation, i.e. with $r \in ((1 + \lambda)\pi/2, \pi)$, prefer A_μ is given by $\mu = (1 - \lambda)\pi$.

5 Discussion

We briefly discuss the implication of the modeling assumptions and some possible implications and extensions of the analysis.

Actions and Desires. In our formulation the moral cost enters both the commitment and temptation ranking sanctioning both the action ("Thou Shalt not Steal") and the "desire" ("Thou Shalt not Covet"). If moral costs enter only the commitment ranking, i.e. morality sanctions actions but not desires, then it involves only costs (for either guilt or foregone payoffs) but no benefits, since it does not help reducing the costs of self-control.³⁶ In this case A_μ would never be preferred to A_0 . Conversely if the moral cost enters only the temptation ranking it is still true that agents prefer A_μ to A_0 if, and only if, morality helps resisting temptation.³⁷ Yet, punishing only desires reduces the efficacy of morality since a larger μ is required to change behavior. Therefore while the qualitative predictions do not depend on the assumption that the moral punishment is the same in both ranking they do require some moral sanction of temptation.

Temptation, external and internal punishments. A key assumption is that external and internal punishments have a different exposure to problems of temptation. We have assumed that internal punishment is not subject to under-evaluation in the heat of the moment of choice. This extreme assumption is of no consequence for the results which nonetheless require some differential exposure to temptation. This assumption appears reasonable in view of the main differential features of external and internal punishment: External sanctions are conditional upon detection, may be delayed in time and are inherently uncertain while the experience of guilt does not require detection and can be immediate upon committing the immoral act. Furthermore, internalized moral values tend to be persistent in time and are not easy to manipulate in the short run.

Formulation of the utility function. We have considered a simple set up with separable utility in returns, external and moral sanctions. The actual formulation of the utility function is of no consequence for the main qualitative predictions (the analysis with a more general utility function is reported in the Appendix). In particular, the prediction that the agents face temptation for intermediate crime premia, and that agents with large (respectively low) crime premia are immoral (respectively amoral) are confirmed. The actual utility formulation affects the levels of the thresholds

³⁶Graphically, the indirect utility represented in the above Figure would have no increasing range.

³⁷Furthermore, if μ enters only in v , individuals with high crime premia may have to exert self-control to resist the temptation of becoming legal. This cannot happen when μ also enters in u since guilt induces legality when crime is still tempting.

and the slope of the indirect utility as a function of μ (which is not piece-wise linear). All comparative statics predictions on the equilibrium choices at both the action and planning stages are confirmed as well as those on the effects of λ on the temptation range. The predictions on the role of punishments (π and μ) and returns to crime r_c on the size of the temptation range are qualified, since changes in any of these parameters affect marginal utility at both indifference thresholds (given by $u(1) = u(0)$ and $v(1) = v(0)$).³⁸ Relaxing the assumption of separability allows to consider interaction effects. For instance, if external punishment is in the form of jail, which arguably reduces the marginal utility of money, then an increase in r_c expands the temptation range. The reason is that the effect of a higher r_c is higher for the temptation than for the commitment ranking. The opposite holds in the case of monetary fines, which, by making an individual poorer, increase the marginal utility of money. Similarly, the effect of μ on the size of the temptation range depend on how external and internal punishment interact with one another. If they reinforce each other, then an increase in μ expands the temptation range, because the it makes the under-evaluation of external punishment more severe. For instance, this happens if external punishment increases the feeling of guilt. The opposite holds if one form of punishment weakens the other, in the sense that they are perceived as substitutes.

Punishments and premia. We have restricted attention to external punishments for mis-behavior but the logic of the analysis identically applies also to problems involving external rewards which, being related to external interventions and being delayed in time, can suffer from temptation problems as well. Considering moral rewards rather than punishments would qualify the predictions in two main dimensions. First, since the indirect utility would be increasing in μ for those choosing the legal action, a large enough μ would induce some agents which in the current setting are immoral to be moral. Second, there would be no amoral individuals (since those choosing legality strictly benefit from the moral premium).³⁹

Heterogenous Exposure to Temptation. We framed our analysis with reference to a single agent. The analysis can be interpreted, however, in terms of the equilibrium choice of different individuals facing e.g. different expected returns r or different levels of temptation λ . For example in the simple

³⁸In the linear case this range is given by $r \in (\lambda\pi, \pi)$ but in general it depends also on the utility formulation. Recall that the temptation range is given by the constellation of parameters for which $u(1) \leq u(0)$ and $v(1) \geq v(0)$. Hence the comparative statics effects of each parameter depends on the change induced on both thresholds.

³⁹This case appears of limited interest, however, since in this formulation there are no utility costs of morality which is therefore a "free lunch". In particular, and trivially, if $\mu = \infty$ then all agents would like to adhere to a system of morality.

application to crime one could consider that agents with unequal expected legal returns r_w (or market wages) that, for a given r_c , imply different expected crime premia $r_c - r_w$. Interpreting the results in terms of socio-economic classes the model predicts that the rich would generally be amoral (since they expect to face less temptations) while the poor would generally be immoral (since illegality represents for them the payoff dominant option). Middle class individuals have the largest exposure to temptation, since they face the largest risk of indulging in payoff dominated illegal actions, and therefore, have the largest incentives to adhere to moral values of honesty.⁴⁰ In the same logic, since the model also predicts that the stronger the temptation the larger the level of μ required to embrace morality, the lower middle classes are expected to display a stricter morality. Similarly, different agents may face different exposure to temptation problems in terms of λ . For example if teen-agers face larger temptation problems than adults, they would give in more frequently to payoff dominated illegal actions. They would also benefit more from a system of strict moral values of honesty compared to adults.

Social interactions. To focus attention on the role of moral values as an individual self-control device we have developed the analysis in a partial equilibrium setting. The framework can be extended to study the role of morality in presence of social interactions which are likely to influence individual choices in terms of crime premia, external punishment and moral sanctions. In fact, in most of the studies cited in Section 2, the rationale for morality relates to general equilibrium effects or externalities. In general equilibrium, one likely effect of having moral agents is the emergence of multiple equilibria due to strategic complementarities. Recall that in equilibrium a larger μ increases the morality range. If there are agents facing different expected returns, then the morality range determines the share of moral individuals. If, in turn, the moral pressure μ increases with the share of moral individuals then we can observe multiple equilibria.⁴¹ This possibility is illustrated in the Figure below. The Figure plots the equilibrium share of moral agents denoted by $\sigma(\mu, \pi_0)$ which is an increasing function of μ , for a given π_0 . It also plots, a (reduced form) function, denoted by $\mu(\sigma)$,

⁴⁰This prediction is, however, conditional on both the distribution of legal and illegal returns. It may also be the case that r_w and r_c are positively correlated if the rich have access to better illegal opportunities. In this case also the rich may face temptation and have larger incentives to be moral. It remains true that it is individuals with intermediate crime premia who face temptation.

⁴¹This amounts to consider also a component of shame related to social norms. As in Weibull and Villa (2005) the model yields a unique equilibrium if the utility cost associated to moral values (the degree to which they generate guilt) is exogenous, whereas multiple equilibria are possible when it is endogenously related to the share of (moral) compliers. In such case, social norms effectiveness and crime rates are jointly determined in equilibrium.

that links the strength of moral values to the share of moral individuals.⁴²

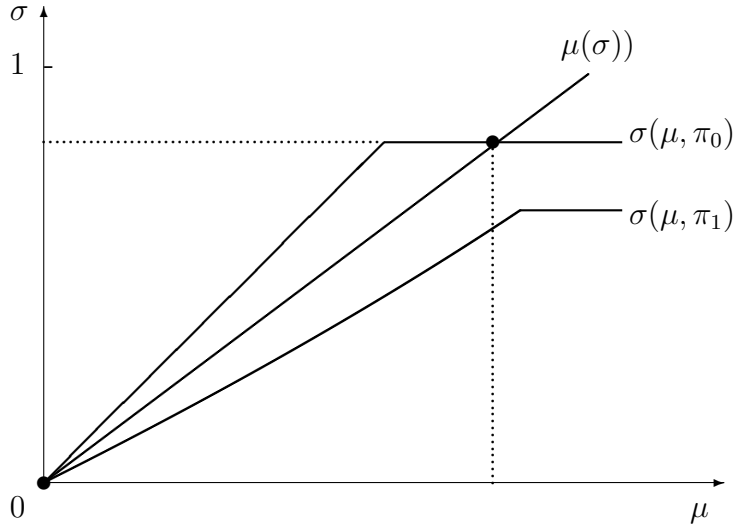


Figure 4: Social Interactions and Multiple Equilibria

An interesting implication of the existence of multiple equilibria is that the interaction between external and internal punishments may lead to the disappearance of equilibria with high morality and little equilibrium crime. From the previous analysis we know that a reduction in external punishment $\pi_1 < \pi_0$ tends to reduce the incentives to acquire morality. This is represented in the above Figure by a downwards shift in the locus representing the equilibrium share of moral individuals, $\sigma(\mu, \pi_1)$. Notice that if the reduction in π is large enough the multiplicity of equilibria may vanish and the only equilibrium is characterized by $\mu = 0$ and $\sigma = 0$. This stylized example illustrates the more general point that in the presence of multiple equilibria induced by social interactions even small variation in external punishment may lead to relatively large changes in equilibrium level of crime and morality.⁴³

Instilling Moral Values. In our analysis we have limited attention to study the effect of morality, and its strength, on individual utility. We did not explicitly investigate the emergence or transmission of moral values. As in the papers discussed in Section 2, moral education is generally performed by parents and other norms setters (as well as social interactions). Our analysis provides a rationale for

⁴²As shown above the share of moral individuals in equilibrium increases until the previously characterized threshold and stabilize afterwards. Concerning the link from σ to μ , we simply postulate a linear relationship in our illustration which delivers two corner equilibria (with shares of moral individuals equal to zero or the maximum). More generally this framework may deliver multiple interior equilibrium in presence of a non linear "social stigma" function.

⁴³Similarly one could consider general equilibrium effects in terms of external detection which are likely to be related to the share of criminals, as in Sah (1991).

altruistic parents to instill moral values in their children in the face of expected future temptations. In this respect, the theory predicts that the optimal level of moral education should depend on parents' expectation about the level of temptation that will be faced by their children. For example if parents expect their children to face some temptation they should optimally inculcate moral values. The model points out, however, that the choice of morality may be a difficult one for those parents that expect their children to face large temptations. They must inculcate very strict moral values in order to successfully change their children's behavior. This is a risky strategy, however, since failure to resist temptation exposes children to the costs of both foregone payoffs and moral guilt. The choice by parents may also involve costs for instilling moral values which, presumably, are increasing with the targeted level of their "strictness". This would lead to large discontinuities in term of choices of moral education since in the face of increasing temptations the optimal moral education requires either instilling a stricter morality or a no morality at all. This also implies that increasing exposure to temptation (which could also be induced by stricter external enforcement) may lead to more "morally polarized" societies: for the individuals it is optimal to either completely give up on morality or to adhere to even stricter moral codes. The consideration of these issues, which is for the moment left for future research, may deliver interesting further insights.

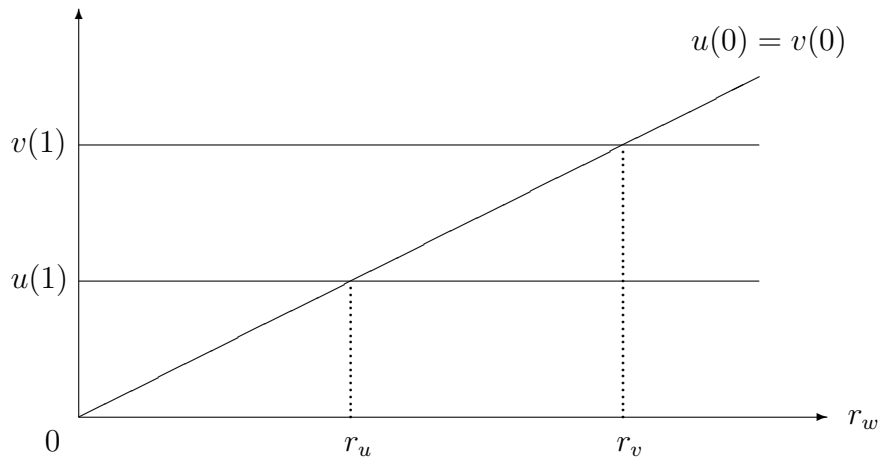
6 Appendix

Formulation of the utility function. Let $Z = Z_1 \times Z_2 \times Z_3$ be the set of outcomes, whose elements are vectors of monetary payments, external punishments and intensity of moral misconduct. The commitment utility function, $u : Z \rightarrow \mathbb{R}$, is defined on such set. Temptation utility, $v : Z \rightarrow \mathbb{R}$, is assumed to differ from u just for an under-evaluation of external punishment, so that, letting $\lambda \in [0, 1]$, it holds that $v(z_1, z_2, z_3) = u(z_1, \lambda z_2, z_3)$. We slightly abuse of notation and denote $u : \Delta \rightarrow \mathbb{R}$ and $v : \Delta \rightarrow \mathbb{R}$ the expected utility functions, where Δ is the set of all probability distributions over Z . In our model $Z_1 = \{r_w, r_c\}$, $Z_2 = \{0, p\}$, $Z_3 = \{0, \mu\}$ and there is a detection probability of crime d . If crime is committed with probability $a \in A_0 \cup A_\mu$, then the probability distribution over Z , denoted $F(a)$, is as follows: $Pr(z_1 = r_c) = a$, $Pr(z_1 = r_w) = 1 - a$, $Pr(z_2 = p) = ad$, $Pr(z_2 = 0) = 1 - ad$; moreover, if $a \in A_0$, then $Pr(z_3 = 0) = 1$ and $Pr(z_3 = \mu) = 0$, whereas if $a \in A_\mu$, then $Pr(z_3 = 0) = 1 - a$ and $Pr(z_3 = \mu) = a$. To simplify notation, we also denote expected utility by $u(a)$ instead of $u(F(a))$. The case studied in the text amounts to assuming that u is additively separable and linear in its arguments, with $\pi = dp$. Here we generalize the analysis and only assume that utility is twice differentiable and its partial derivatives satisfy $u_1 > 0$, $u_2 < 0$ and $u_3 < 0$.

Action stage. With this notation in mind, it is immediate to write the full expression of expected utilities when either crime or legality are chosen in pure strategies and thus solve the action stage problem. Independently of the presence or absence of moral values, legality yields the same commitment and temptation utility $u(0) = u(r_w, 0, 0) = v(0)$. In presence of moral values, crime yields expected utilities $u(1) = u(r_c, 0, \mu)(1 - d) + u(r_c, p, \mu)d$ and $v(1) = u(r_c, 0, \mu)(1 - d) + u(r_c, \lambda p, \mu)d$. Utilities in the absence of moral values are formally identical to these ones, with $\mu = 0$, so we simply state results for a general μ . By inverting $u(0)$ with respect to its first argument, so that $u(r_w, 0, 0) = y \iff r_w = g(y)$, we can generalize Proposition 1 and Corollary 1. Define $r_u \equiv g[u(1)]$, $r_v \equiv g[v(1)]$ and $r_\ell \equiv g\left[\frac{u(1)+v(1)}{2}\right]$.

Proposition 3. *The three thresholds r_u, r_ℓ, r_v satisfy $r_u \leq r_\ell \leq r_v$ and such that (i) individuals with $r_w \in (r_u, r_v)$ are tempted by crime and split between those with $r_w \in (r_u, r_\ell)$, who give in to temptation, and those with $r_w \in (r_\ell, r_v)$, who resist temptation; hence (ii) crime is chosen whenever $r_w < r_\ell$ and legality whenever $r_w > r_\ell$.*

Proof. Since $\lambda \in [0, 1]$ and $u_2 < 0$, it holds that $v(1) \geq u(1)$, and since $u_1 > 0$, this implies that $r_u \leq r_\ell \leq r_v$. Then $u(1) > u(0) \iff r_w < r_u$; $v(1) > v(0) \iff r_w < r_v$; and $u(1) + v(1) > u(0) + v(0) \iff r_w < r_\ell$. \square



The above figure gives a visual representation of the implied thresholds for the case in which utility is linear in monetary payments (that is, when the second derivative in the first argument satisfies $u_{11} = 0$). It is immediate to see how the picture changes if $u(0)$ is a concave function of r_w . In terms of comparative statics, we have the following results. Both external punishment p and detection probability d reduce crime: $\frac{\partial r_\ell}{\partial p} < 0$, $\frac{\partial r_\ell}{\partial d} < 0$. They also move down the range of temptation, in the sense that they reduce both r_u and r_v . If utility is linear in monetary payments (if $u_{11} = 0$), increases in p and d also expand the length of the temptation range, $(r_v - r_u)$.⁴⁴ An increase in temptation problems due to a reduction in λ increases both crime and the length of the temptation range. An increase in crime's monetary payoff r_c increases crime and moves up the temptation range (it raises r_ℓ, r_u and r_v). It also expands the temptation range if, as it is plausible, external punishments decrease the marginal utility of money (if the cross-derivative satisfies $u_{12} < 0$), since in this case the difference between the commitment and the temptation ranking makes $0 < \frac{\partial u(1)}{\partial r_c} < \frac{\partial v(1)}{\partial r_c}$.⁴⁵ Finally, an increase in the perceived seriousness of moral violation (an increase in μ) reduces crime and moves down the temptation range (it reduces r_ℓ, r_u and r_v). In the case in which utility is linear in monetary payments, if external and internal punishments reinforce each other (if $u_{23} < 0$), then an increase in μ also expands the range of temptation (it raises $r_v - r_u$), whereas the opposite is true if external and internal punishment weaken each other (if $u_{23} > 0$).⁴⁶

Planning stage. We can now study planning stage choices and generalize Proposition 2. Given $\mu > 0$, planning stage choices depend on the comparison between $U(A_\mu)$ and $U(A_0)$. Since formally (and trivially) $U(A_0) = U(A_\mu)$ for $\mu = 0$, we can initially use the general notation $U(A_\mu)$ for $\mu \geq 0$ and later distinguish between A_0 and A_μ when $\mu > 0$. Denote $u(1|A_\mu)$ the commitment utility function for a moral individual choosing crime and use an analogous notation for v . Notice that $u(0)$ and $v(0)$ are equal and independent of μ , so we need not condition on a legal individual's morality. To make it explicit that the relevant thresholds depend on μ , we write them as $r_u(\mu) \equiv g[u(1|A_\mu)]$, $r_v(\mu) \equiv g[v(1|A_\mu)]$ and $r_\ell(\mu) \equiv g\left[\frac{u(1|A_\mu)+v(1|A_\mu)}{2}\right]$. From the above analysis we then have $U(A_\mu) =$

$$\begin{aligned}
&= \begin{cases} u(0) & \text{if } r_w > r_v(\mu) \\ u(0) - [v(1|A_\mu) - v(0)] & \text{if } r_w \in (r_\ell(\mu), r_v(\mu)) \\ u(1|A_\mu) & \text{if } r_w < r_\ell(\mu) \end{cases} \\
&= \begin{cases} u(r_w, 0, 0) & \text{if } r_w > r_v(\mu) \\ 2u(r_w, 0, 0) - [u(r_c, 0, \mu)(1-d) + u(r_c, \lambda p, \mu)d] & \text{if } r_w \in (r_\ell(\mu), r_v(\mu)) \\ u(r_c, 0, \mu)(1-d) + u(r_c, p, \mu)d & \text{if } r_w < r_\ell(\mu) \end{cases}
\end{aligned}$$

The graphical representation of $U(A_\mu)$ is qualitatively analogous to the one shown in the main text, apart for the possible non linearities. Let $h(\cdot)$ be the (decreasing) inverse function of $v(1|A_\mu)$ with respect to μ , so that $v(1|A_\mu) = [u(r_c, 0, \mu)(1-d) + u(r_c, \lambda p, \mu)d] = y \iff \mu = h(y)$, and define $\bar{\mu} \equiv h(2u(0) - u(1|A_0))$. Given $\mu > 0$, define the (decreasing) function $r_m(\mu) \equiv g\left[\frac{u(1|A_0)+v(1|A_\mu)}{2}\right]$ and notice that $r_w > r_m(\mu) \iff \mu > \bar{\mu}$. With this notation, we can extend the results in the main text.

⁴⁴This is due to the fact that $u_2 < 0$ and $\frac{\partial u(1)}{\partial x} < \frac{\partial v(1)}{\partial x} = \lambda \frac{\partial u(1)}{\partial x} < 0$, for $x = p, d$. The expansion of the temptation range is not granted if utility is concave in monetary payments (if $u_{11} < 0$), since concavity would reduce $(r_v - r_u)$ if $u(1)$ and $v(1)$ were to decrease at the same speed, and this effect counterbalances the effect of the faster decrease of $u(1)$.

⁴⁵Notice that this result, which holds both under $u_{11} = 0$ and under $u_{11} < 0$, differs from the case analyzed in the text, which was derived under separability, so that $u_{12} = 0$, in which case r_c has no effect on the length of the temptation range.

⁴⁶In the case of concavity (if $u_{11} < 0$), the former effect is counterbalanced and the latter is reinforced, so that the effect of an increase in μ on $(r_v - r_u)$ is ambiguous if $u_{23} < 0$ and is negative if $u_{23} > 0$.

Proposition 4. For any given $\mu > 0$, the following holds. An individual:

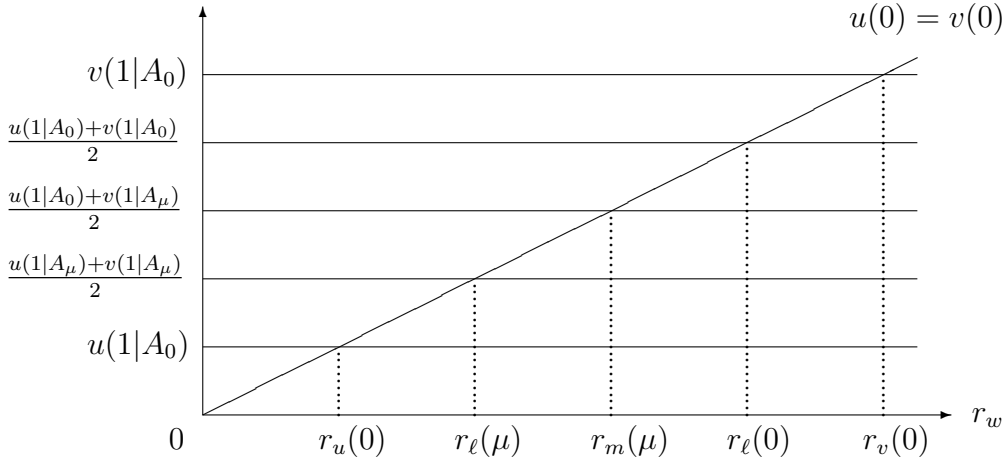
- is indifferent between A_μ and A_0 if $r_w \geq r_v(0)$;
- prefers A_μ if $r_w \in (r_\ell(0), r_v(0))$;
- given $r_w \in (r_u(0), r_\ell(0))$, prefers A_μ if $\mu > \bar{\mu}$ and A_0 if $\mu < \bar{\mu}$;
- prefers A_0 if $r_w < r_u(0)$.

Proof. First, if $r_w \geq r_v(0)$, then independently of the moral choice legality is chosen and no self-control costs are paid, making the individual indifferent between A_μ and A_0 (amoral). Second, if $r_w \in (r_\ell(0), r_v(0))$, then legality is chosen in any case, but morality reduces (and eventually eliminates) self-control costs, which would be paid in its absence, making the individual prefer A_μ (moral): formally, $u(1|A_0) < u(0) < v(1|A_0)$ and $u(0) + v(0) > u(1|A_0) + v(1|A_0)$, implying $U(A_\mu) > U(A_0)$ for any $\mu > 0$. Third, if $r_w \in (r_u(0), r_\ell(0))$, then in the absence of morality legality is optimal for the commitment ranking but the individual gives in to temptation and chooses crime. Formally, as above $u(1|A_0) < u(0) < v(1|A_0)$, but now $u(0) + v(0) < u(1|A_0) + v(1|A_0)$. In this case, $U(A_\mu) > U(A_0) \iff 2u(0) - v(1|A_\mu) > u(1|A_0) \iff \mu > \bar{\mu}$, so that the individual adheres to a strict morality but rejects a mild one. Forth, if $r_w < r_u(0)$, then in the absence of morality crime is optimal even for the commitment ranking. In this case, $U(A_\mu) < U(A_0)$ for any $\mu > 0$. \square

Corollary 2. Given $\mu > 0$, an individual strictly gains from A_μ if and only if

$$r_w \in (\max\{r_u(0), r_m(\mu)\}, r_v(0)).$$

We refer to this interval as to the morality range and call M its length, so that $M \equiv r_v(0) - \max\{r_u(0), r_m(\mu)\}$. The following figure gives a visual representation of the various thresholds and helps understand the comparative statics on M .⁴⁷



A stricter moral code (an increase in the intensity of moral misconduct μ) expands the morality range until it reaches its maximum length: for $\mu \rightarrow 0$ we have $M \rightarrow r_v(0) - r_\ell(0)$; as μ increases above zero, we have $\frac{\partial M}{\partial \mu} > 0$ as long as $r_m(\mu) > r_u(0)$; since $r_m(\mu)$ is decreasing, there is a threshold of μ above which this inequality does not hold any more and we have $M = r_v(0) - r_u(0)$, with

⁴⁷It is drawn for the case in which utility is linear in monetary payments ($u_{11} = 0$), but it is immediate to see how it changes if $u(0)$ is a concave function of r_w ($u_{11} < 0$). The equal spacing between lines and thresholds is a simplification adopted just for graphical clarity.

$\frac{\partial M}{\partial \mu} = 0$.⁴⁸ Let us now focus on the case of $r_m(\mu) > r_u(0)$. It is easy to find sufficient conditions, under which the comparative statics results on expected external punishment extend. For instance, an increase in either detection probability d or external punishment p expands the morality range if utility is linear in monetary payments and concave in external punishment (convex cost), and if external and moral sanctions reinforce each other: formally, if $u_{11} = 0$, $u_{22} < 0$ and $u_{23} < 0$. In this case higher expected external sanctions make temptation problems more severe and morality a more effective commitment device, thus raising its demand. Outside of these sufficient conditions, however, these results have to be qualified.^{49,50} Although not necessarily granted, the result that an increase in temptation due to a reduction of λ expands the morality range also extends to plausible configurations.⁵¹ Finally, an increase in monetary returns to crime, which has no effect on the length of the morality range when utility is additively separable, now has the interesting effect of expanding such range, at least under the plausible condition that both external and moral punishments lower the marginal utility of money.⁵²

⁴⁸Notice that $\frac{\partial M}{\partial \mu} > 0 \iff \frac{\partial}{\partial \mu} \left[\frac{u(1|A_0) + v(1|A_\mu)}{2} \right] < \frac{\partial v(1|A_0)}{\partial \mu}$ and $r_m(\mu) > r_u(0)$. Notice as well that $u(1|A_0)$ and $v(1|A_0)$ are not responsive to μ .

⁴⁹Notice that $\frac{\partial M}{\partial p} > 0 \iff \frac{\partial r_m(\mu)}{\partial p} < \frac{\partial r_v(0)}{\partial p}$. If $u_{11} = 0$, this holds if and only if $\frac{1}{2} \frac{\partial u(1|A_0)}{\partial p} + \frac{1}{2} \frac{\partial v(1|A_\mu)}{\partial p} < \frac{\partial v(1|A_0)}{\partial p}$. Since $\frac{\partial u(1|A_0)}{\partial p} < \frac{\partial v(1|A_0)}{\partial p} \iff u_{22} < 0$ and $\frac{\partial u(1|A_0)}{\partial p} < \frac{\partial v(1|A_0)}{\partial p} \iff u_{23} < 0$, we have that $u_{11} = 0$, $u_{22} < 0$ and $u_{23} < 0$ are sufficient conditions for $\frac{\partial M}{\partial p} > 0$. This last result is not granted outside these conditions. For instance, if $u_{11} = 0$ and both $u_{22} > 0$ and $u_{23} > 0$, then $\frac{\partial M}{\partial p} < 0$.

⁵⁰Analogously, $\frac{\partial M}{\partial d} > 0 \iff \frac{\partial r_m(\mu)}{\partial d} < \frac{\partial r_v(0)}{\partial d}$. If $u_{11} = 0$, this holds if and only if $\frac{1}{2} \frac{\partial u(1|A_0)}{\partial d} + \frac{1}{2} \frac{\partial v(1|A_\mu)}{\partial d} < \frac{\partial v(1|A_0)}{\partial d}$. Since $\frac{\partial u(1|A_0)}{\partial d} < \frac{\partial v(1|A_0)}{\partial d}$ always holds and $\frac{\partial u(1|A_0)}{\partial d} < \frac{\partial v(1|A_0)}{\partial d} \iff u_{23} < 0$, we have that $u_{11} = 0$ and $u_{23} < 0$ are sufficient conditions for $\frac{\partial M}{\partial d} > 0$.

⁵¹Formally, $\frac{\partial M}{\partial \lambda} < 0 \iff 0 > \frac{\partial r_m(\mu)}{\partial \lambda} > \frac{\partial r_v(0)}{\partial \lambda}$. If $u_{11} = 0$, this holds if and only if $0 > \frac{1}{2} \frac{\partial v(1|A_\mu)}{\partial \lambda} > \frac{\partial v(1|A_0)}{\partial \lambda}$. This is granted if $u_{23} \geq 0$; while it is not for $u_{23} < 0$, the effect of the cross-derivative on temptation utility should be twice as strong as the direct effect of λ in order to revert the result. Moreover, if $u_{11} < 0$, it is even easier that $\frac{\partial M}{\partial \lambda} < 0$.

⁵²Formally, $\frac{\partial M}{\partial r_c} > 0 \iff \frac{\partial r_m(\mu)}{\partial r_c} < \frac{\partial r_v(0)}{\partial r_c}$. If $u_{11} = 0$, this holds if and only if $\frac{1}{2} \frac{\partial u(1|A_0)}{\partial r_c} + \frac{1}{2} \frac{\partial v(1|A_\mu)}{\partial r_c} < \frac{\partial v(1|A_0)}{\partial r_c}$. Since $\frac{\partial u(1|A_0)}{\partial r_c} < \frac{\partial v(1|A_0)}{\partial r_c} \iff u_{12} < 0$ and $\frac{\partial u(1|A_0)}{\partial r_c} < \frac{\partial v(1|A_0)}{\partial r_c} \iff u_{13} < 0$, we have that $u_{11} = 0$, $u_{12} < 0$ and $u_{13} < 0$ are sufficient conditions for $\frac{\partial M}{\partial r_c} > 0$. The fact that $\frac{\partial M}{\partial r_c} > 0$ is all more likely if $u_{11} < 0$.

References

- ADRIANI, F., AND S. SONDEREGGER (2009): “Why do Parents Socialize their Children to Behave Pro-Socially? An Information-Based Theory,” *Journal of Public Economics*, 93(11-12), 1119–1124.
- ALDASHEV, G., I. CHAARA, J.-P. PLATTEAU, AND Z. WAHHAJ (2009): “Using the Law to change the Custom,” mimeo.
- BECKER, G. S. (1968): “Crime and punishment: an economic approach,” *Journal of Political Economy*, 76(2), 169–217.
- BENHABIB, J., AND A. BISIN (2005): “Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption-saving decisions,” *Games and Economic Behavior*, 52(2), 460–492.
- BISIN, A., AND T. VERDIER (2001): “The Economics of Cultural Transmission and the Evolution of Preferences,” *Journal of Economic Theory*, 97(2), 298–319.
- BUONANNO, P., D. MONTOLIO, AND P. VANIN (2009): “Does Social Capital Reduce Crime?,” *Journal of Law & Economics*.
- BUONANNO, P., G. PASINI, AND P. VANIN (2008): “Crime and social sanction,” Marco Fanno Working Paper 71, Department of Economics, University of Padua.
- CERVELLATI, M., J. ESTEBAN, AND L. KRANICH (2006): “Work Values, Endogenous Sentiments and Redistribution,” *IZA DP*.
- CERVELLATI, M., M. JANSEN, AND U. SUNDE (2008): “Insurance, Human Capital and Technological Change: A Theory on the Economic Role of Religious Norms,” *mimeo*.
- CHAMLIN, M. B., AND J. K. COCHRAN (1997): “Social altruism and crime,” *Criminology*, 35(2), 203–228.
- CONLEY, J. P., AND P. WANG (2006): “Crime and ethics,” *Journal of Urban Economics*, 60(1), 107 – 123.
- CORICELLI, G., M. JOFFILY, C. MONTMARQUETTE, AND M.-C. VILLEVAL (2007): “Tax Evasion: Cheating Rationally or Deciding Emotionally?,” IZA Discussion Papers 3103, Institute for the Study of Labor (IZA).
- CORNEO, G., AND O. JEANNE (2009): “A theory of tolerance,” *Journal of Public Economics*, 93(5-6), 691–702.
- CROCKER, J., AND L. E. PARK (2003): “Seeking Self-Esteem: Construction, Maintenance and Protection of Self-Worth,” *Handbook of Self and Identity*, M. Leary and J. Tangney Eds NY Guilford, pp. 291–313.
- DAL BÓ, E., AND P. DAL BÓ (2009): “”Do the Right Thing”: The Effects of Moral Suasion on Cooperation,” Nber working papers, National Bureau of Economic Research, Inc.
- DAL BÓ, E., AND M. TERVIÖ (2008): “Self-Esteem, Moral Capital, and Wrongdoing,” NBER Working Papers 14508.
- DEKEL, E., B. L. LIPMAN, AND A. RUSTICHINI (2009): “Temptation-Driven Preferences,” *Review of Economic Studies*, 76(3), 937–971.

- DILLS, A. K., J. A. MIRON, AND G. SUMMERS (2008): "What Do Economists Know About Crime?," NBER Working Papers 13759.
- EHRlich, I. (1973): "Participation in illegitimate activities: a theoretical and empirical investigation," *Journal of Political Economy*, 81(3), 521–565.
- FRANK, R. H. (1987): "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?," *American Economic Review*, 77(4), 593–604.
- FUNK, P. (2004): "On the effective use of stigma as a crime deterrent," *European Economic Review*, 48(4), 715–728.
- GLAESER, E. L., B. SACERDOTE, AND J. A. SCHEINKMAN (1996): "Crime and social interactions," *Quarterly Journal of Economics*, 111(2), 507–548.
- GOTTFREDSON, M. R., AND T. HIRSCHI (1990): *A General Theory of Crime*. Stanford University Press.
- GUL, F., AND W. PESENDORFER (2001): "Temptation and Self-Control," *Econometrica*, 69(6), 1403–1435.
- KAPLOW, L., AND S. SHAVELL (2007): "Moral Rules, the Moral Sentiments, and Behavior: Toward a Theory of an Optimal Moral System," *Journal of Political Economy*, 115(3), 494–514.
- LAIBSON, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 62(May), 443–477.
- LINDBECK, A., AND S. NYBERG (2006): "Raising Children to Work Hard: Altruism, Work Norms and Social Insurance," *Quarterly Journal of Economics*, 121(4), 1473–1503.
- POLINSKY, M. A., AND S. SHAVELL (2000): "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature*, 38(1), 45–76.
- POSNER, R. A. (1997): "Social Norms and the Law: An Economic Approach," *American Economic Review*, 87(2), 365–369.
- RASMUSEN, E. (1996): "Stigma and self-fulfilling expectations of criminality," *Journal of Law & Economics*, 39(2), 519–544.
- ROTEMBERG, J. J. (1994): "Human Relations in the Workplace," *Journal of Political Economy*, 102(4), 684–718.
- SAH, R. K. (1991): "Social Osmosis and Patterns of Crime," *Journal of Political Economy*, 99(6), 1272–95.
- STOVALL, J. E. (2009): "Multiple Temptations," *Econometrica*, *Forthcoming*.
- STROTZ, R. H. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization," *The Review of Economic Studies*, 23(3), 165–180.
- TABELLINI, G. (2008): "The Scope of Cooperation: Values and Incentives," *The Quarterly Journal of Economics*, 123.
- WEIBULL, J., AND E. VILLA (2005): "Crime, punishment and social norms," *SSE/EFI Working Paper*, (610).