DISCUSSION PAPER SERIES

IZA DP No. 7334

"Thou Shalt Not Covet ...": Prohibitions, Temptation and Moral Values

Matteo Cervellati
Paolo Vanin

April 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

“Thou Shalt Not Covet ...”: Prohibitions, Temptation and Moral Values

Matteo Cervellati
University of Bologna
and IZA

Paolo Vanin
University of Bologna

Discussion Paper No. 7334
April 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 7334
April 2013

ABSTRACT

"Thou Shalt Not Covet ...": Prohibitions, Temptation and Moral Values*

This paper proposes a theory of the relationship between prohibitions and temptation. In presence of self-control problems, moral values may increase individual material welfare (and utility) by serving as a self-commitment device. The model investigates the relationship between morality and temptation, the individual gains from morality, the interaction between external sanctions and moral self-punishment and the spread and strength of individually optimal moral values. The empirical analysis, based on survey data for a large set of countries, documents a hump-shaped pattern of morality in social class, which supports the theoretical predictions of the model.

JEL Classification: D03, K42, Z13

Keywords: prohibitions, temptation, self-control, moral values, crime

Corresponding author:

Paolo Vanin
University of Bologna
Department of Economics
Piazza Scaravilli 2
40126 Bologna
Italy
E-mail: paolo.vanin@unibo.it

* We are particularly grateful to four anonymous referees, to Hanming Fang and to Dan Silverman for very constructive and insightful comments and to Simona Valmori for excellent research assistantship. We would also like to thank Alberto Bisin, Giacomo Calzolari, Luca Corazzini, Giacomo Corneo, Ernesto dal Bó, Davide Dragone, David Levine, Sten Nyberg, conference participants at ASSET Florence, EEA-ESEM Barcelona, BOMOPA Modena, SIDE-ISLE Bologna, Social Economics Bologna, Grass Modena, SIE Catania, Economics of Culture, Institutions and Crime Milan, Institutions, Individual Behavior and Economic Outcomes Alghero, and seminar participants at Free University of Berlin, San Andrés University of Buenos Aires and Universities of Bologna, Pavia and Padova. We gratefully acknowledge financial support from the University of Padua (research grant CPDA071899) and EIEF Rome.

1 Introduction

Many activities are characterized by externally sanctioned prohibitions. These include not only illegal or criminal behaviors sanctioned by law, but also norms of conduct enforced through social sanctions. Most cultures (and religions) also explicitly attach moral sanctions to prohibited, and often formally illegal, actions. For instance, the Bible features commandments against theft and adultery, and analogous prescriptions can be found in different cultures. Interestingly, the commandments do not only prohibit the acts, but also explicitly censure the *desire* of other people's goods and wife. This is suggestive of the perceived relevance of temptation problems and of the possible role of moral values in reducing them.[1] This paper proposes a theory of prohibitions and temptation, in which internal (moral) sanctions may serve the role of self-commitment device in presence of externally sanctioned prohibitions and self-control problems.[2]

We study a theoretical set-up along the lines of the standard economic approach pioneered by Becker (1968), in which breaking prohibitions (or acting illegally) is a rational choice based on expected costs and benefits. We consider the choice between a prohibited (or illegal) and a legal choice. A main difference with respect to the existing literature in crime economics is the explicit consideration of self-control problems. Individuals can be tempted to break a prohibition by its immediate returns, although they understand that this action will not pay off in the long run, once expected sanctions are fully taken into account.[3] As a result, the theory allows for the existence of temptation-driven crimes in equilibrium. This approach combines the modeling of rational individual behavior, which is typical of the economic approach, with the view that prohibited (or

[1]The commandment against actions reads: '*You shall not commit adultery. You shall not steal.*' (Exodus, 20:12). The prohibition of desire reads: '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 is inspired by the ancient version of this commandment.

[2]While the existence of temptation problems in association to prohibited actions appears a relevant issue, the link between prohibitions, temptations and moral values has essentially been neglected in economics. Dal Bó and Terviö (2012) study the accumulation of "moral capital" as an introspective reputation for goodness in presence of temptation. Fershtman, Gneezy and Hoffman (2011) investigate individual adherence to taboos, which associate a cost not only to the act of breaking a prohibition but also to the thought of doing it.

[3]Preferences are framed along the lines of Gul and Pesendorfer (2001). Specifically, self-control problems are modeled as a (temporary) increase in the subjective discount rate at the moment of actual choices. This is line with the view that under the effects of a visceral urge individuals' mind may temporarily relegate the possibility of future punishments to the background.

illegal) actions are also driven by short term returns, with temptation playing a crucial role.[4]

The baseline framework is extended to the consideration of internalized (moral) punishments, which can affect individual behavior by triggering the experience of 'self-regulative' emotions like guilt.[5] Moral punishments have two well studied features that are particularly interesting for the present purposes. The first, defining, feature is that internal enforcement of prohibitions relies on self-detection (or, better, self-awareness) of misbehavior. Guilt for breaking moral prescriptions may be felt immediately upon act (and in fact even upon desire). Prohibitions enforced through moral sanctions can thus involve a lower exposure to temptation than those enforced through external punishments. Individuals with self-control problems who are exposed to external punishments, may have a higher expected utility, or even a higher material welfare, if they are endowed with (potentially costly) moral values, which serve the role of self-commitment devices.[6]

The second distinctive feature of moral punishments is that they require an active process of education (or inculcation) by some moral setter. Contrary to external sanctions, which can be effectively imposed upon individuals irrespective of whether they find them legitimate or not, moral prohibitions only work if individuals accept or 'internalize' them and if they are self-aware of misbehavior. This leads to the natural questions about the incentives to provide moral education and about its optimal strictness. We tackle such questions from the perspective of perfectly

[4]While, following Ehrlich (1973), economists have modeled crime as a long term occupational choice problem, criminologists emphasize its short term drivers and the fact that most crimes are not performed by professional criminals. For instance, 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). 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). The modeling of temptation proposed here is in line with this view.

[5]Following the tradition in social psychology, moral punishments result from the experience of 'self-regulative' emotions. See, e.g., Crocker and Park (2003) for an extensive survey and in particular Tangney and Tracy (2012) for a recent overview of the literature on 'self-conscious' emotions. The role of guilt is modeled along the lines of Lindbeck, Nyberg and Weibull (1999) and, more specifically, Lindbeck and Nyberg (2006) and Cervellati, Esteban and Kranich (2010). This literature and the (neuro) experimental evidence on the working of moral punishments is discussed in further detail in Section 3.

[6]While individuals with self-control problems would be better off if they could exclude all tempting options from their choice set, this may not be feasible if temptation arises from externally imposed punishments. Yet, by attaching self-inflicted costs of guilt, moral values may make the prohibited but tempting options less desirable.

altruistic moral setters, like parents.[7] Inculcating moral values need not always be beneficial, since it involves both expected costs and benefits. Moral values can raise material welfare (or utility) only if they are strong enough to allow temptation to be resisted: a weak morality, which does not prevent from giving in to temptation, only adds the costs of guilt to those of forgone payoffs (for giving in to tempting but suboptimal choices). Too strict values need not be optimal either, however, since they may prevent from taking advantage of payoff-maximizing (although prohibited) opportunities.

The theory contributes to the literature a positive analysis of the emergence of (individually costly) moral values.[8] The idea that moral values (or a 'conscience') may benefit their carrier by serving as a commitment device is not new. Frank (1987) forcefully made the point that a credible preference for honesty might help improving individual welfare by allowing to sustain cooperation with others.[9] In this paper, the commitment role of moral values is towards oneself rather than towards others, which implies that agents have no incentive to mimic moral preferences and that moral values can emerge and be sustained irrespective of the actual observability and credibility of moral preferences.[10]

The model delivers some relevant qualifications of standard policy implications and sheds light on the interactions between externally and internally enforced prohibitions.[11] It suggests that,

[7]We assume perfect altruism, rather than imperfect empathy, as in Bisin and Verdier (2001). This essentially implies looking at values transmission from the perspective of the receiver's well being, rather than from that of the value setter. Perfectly altruistic parents may either care for their children's material welfare, as in Adriani and Sonderegger (2009), or for their overall utility, as in Corneo and Jeanne (2009). While the former paper looks at imperfect information and the latter at the emergence of values of tolerance in the face of parents' uncertainty, we are rather interested in moral values in presence of self-control problems.

[8]Kaplow and Shavell (2007) analyze optimal moral systems in presence of externalities and imperfect moral inculcation. We focus on costs and benefits from morality at the individual rather than at the social level. The model provides a simple positive rationale for moral values, even in the absence (or on top) of externalities, multiple equilibria and coordination failures, which motivate moral values from a normative perspective.

[9]Rotemberg (1994) gives a related account, based on strategic complementarities in cooperation, of (potentially costly) altruism towards co-workers. Tabellini (2008) provides a recent investigation of the endogenous emergence of cooperative values.

[10]If some observable behavior signals morality and induces other people's cooperation, individuals with opportunistic preferences have an incentive to mimic that behavior, posing a serious threat to the credibility of morality signals. See Robson and Samuelson (2011) for a recent overview of the literature on the evolution of preferences, including a discussion of the credibility of morality signals.

[11]The literature has studied the interaction between different types of external enforcement of prohibitions,

even in the absence of morality, the timing of detection and punishment, and not only the overall expected punishment by itself, are crucial for reducing crimes. A new insight is that early detection plays a crucial role in discouraging temptation-driven crimes. Raising external punishments, like for instance criminal justice penalties, increases the incentive for benevolent moral setters to provide (some) moral education, but it crowds out the intensity of morality by inducing them to instill weaker moral values.

Finally, the analysis provides insights on the empirical analysis of crime and individual moral values. With random returns, individuals may make 'mistakes' (from the point of view of material welfare) by either committing 'stupid' crimes, which do not pay off in material terms, or by refraining from profitable ones. The model therefore offers a simple explanation, rooted in temptation and morality, for these seemingly irrational 'mistakes' whose joint observation has been considered difficult to rationalize within standard frameworks.[12] The theory also predicts an interesting non-monotonicity: individuals with intermediate relative returns to legality should benefit the most from moral values.[13] We provide empirical support for this novel prediction using data from the World Value Survey on individual moral values and socio-economic class. In line with the theory, individuals belonging to the middle classes give the largest importance to moral values.[14]

The paper is organized as follows. Section 2 lays down the baseline model. Section 3 introduces moral values, studies their implications and the incentives to provide moral education. Section 4 discusses the empirical relevance and Section 5 concludes. All proofs are relegated to the Appendix.

whether legal or social. Posner (1997) discusses the interplay between social norms and the law. Glaeser et al. (1996), Rasmusen (1996) and Aldashev et al. (2010), among others, provide crime theories based on peer pressure and stigmatization. See Polinsky and Shavell (2000) for a review of the economic theory of public enforcement of law. The focus in this paper is rather on the study of the interaction between self-punishment, third-party punishments and the emergence of moral values.

[12]This issue is discussed in more details in Section 4. Documenting the empirical role of morality for crime has proved difficult (see Dills, Miron and Summers (2008) for a critical assessment of the literature). Buonanno, Montolio and Vanin (2009), among others, provide evidence that morally motivated activities reduce crime.

[13]For very low or very high crime premia, legality or crime are the optimal choices and no temptation problems arise. In those cases, having moral values is either irrelevant of harmful.

[14]The empirical analysis exploits information for about 170,000 individuals living in 79 countries for the period 1994-2008. The results are derived in regressions with country-year fixed effects (thereby dealing with both country-specific, time-invariant unobserved characteristics and with changes in morality over time) and are robust to the inclusion of standard controls and to several checks.

2 A Simple Model of Prohibition and Temptation

2.1 Set-Up

Consider an individual who has to choose between a permitted and a prohibited action. To fix ideas, let us call these actions legality (ℓ) and crime (c), respectively. Both yield strictly positive returns, denoted r_ℓ and r_c. The main difference between the two activities is that the prohibited action, once and if detected, is punished with a sanction $p > 0$. As in Davis (1988), detection is stochastic and may take place at different moments in time: either in period 1, that is when the action is chosen and takes place (*in flagrante delicto*), or at a later stage, which we call period 2. The probability of detection in period 1 is denoted d_1. The probability of being detected for the first time in period 2 is $(1 - d_1)d_2$, where d_2 is the probability of being detected in period 2 conditional on not being detected in period 1. Denote by $\mathbf{d} = (d_1, d_2) \in [0, 1]^2$ the vector of probabilities. The expected punishment, discounted at the moment of action, is given by

$$\pi(\mathbf{d}, p, \beta) \equiv [d_1 + \beta(1 - d_1)d_2]p, \tag{1}$$

where $\beta \in (0, 1)$ is the subjective discount factor.[15]

As benchmark we consider an additively separable and linear utility function.[16] The expected utility yielded by legality and by crime is given by

$$u(\ell) = r_\ell \tag{2}$$

and

$$u(c) = r_c - \pi(\mathbf{d}, p, \beta), \tag{3}$$

respectively. In the absence of temptations problems, crime is an (individually) optimal action if, and only if, the crime premium $r \equiv r_c - r_\ell$ is higher than the expected discounted punishment: $u(c) \geq u(\ell) \iff r \geq \pi(\mathbf{d}, p, \beta)$.

Let us now depart from the standard model and introduce temptation and self-control problems, arising from an increase in impatience at the moment of choice. Specifically, suppose the individual

[15] To make the problem interesting, we assume that $\mathbf{d} \neq (0, 0)$, so that there is some positive detection probability (either upon act, or in the future, or both). Considering a two-period model is without loss of relevant generality. The model can be easily extended to a true multi-period framework since, as clarified below, what matters is the probability of detection upon act relative to detection in the future.

[16] As discussed below, the main results do not depend on the actual formulation of the utility function. The linear additive formulation is chosen as benchmark since it allows a very simple illustration of the main implications.

faces a tension between the way in which he ranks actions when he evaluates them 'coldly' (as if he could abstractedly commit to the best choice and therefore had no self-control problems) and the way in which he does 'in the heat of the moment' (when faced with the actual choice). Let this tension be captured by two different rankings (or utility functions), u (defined above) and v, differing from u only for a lower discount factor, $\beta_v \in (0, \beta)$. According to the last ranking, the utility derived from legality and crime are therefore given by

$$v(\ell) = r_\ell \tag{4}$$

and

$$v(c) = r_c - \pi(\mathbf{d}, p, \beta_v). \tag{5}$$

This simple set-up captures the idea that, when faced with actual choices, individuals may underestimate the potential future utility costs resulting from punishment of prohibited actions. In other words, temptation is modeled as a (temporary) increase in impatience (or myopia), which is due to the urge of experiencing immediate utility.

To save on notation, let $\pi = \pi(\mathbf{d}, p, \beta)$ and $\pi_v = \pi(\mathbf{d}, p, \beta_v)$. The commitment and temptation utility from crime can therefore be expressed as $u(c) = r_c - \pi$ and $v(c) = r_c - \pi_v$, respectively. The prohibited action maximizes the temptation ranking if, and only if, $r \geq \pi_v$. Self-control problems are present if for crime the temptation utility is larger than the commitment utility, that is if

$$v(c) - u(c) = \pi - \pi_v = (1 - d_1)d_2(\beta - \beta_v)p > 0, \tag{6}$$

which, given $p > 0$ and $\beta_v < \beta$, is always true, unless either $d_1 = 1$ or $d_2 = 0$, in which case $\pi_v = \pi$ and hence $v(c) = u(c)$. If detection is only possible upon act ('in flagrante delicto'), then a (temporary) increase in impatience does not cause self-control problems; otherwise, it may create an internal conflict by opening a wedge between the two utility rankings.

If $\pi_v < \pi$, then the individually optimal choice according to u may be different from the optimal choice according to v. We have to specify how individuals solve this conflict. To this purpose, we follow Gul and Pesendorfer (2001).[17] A choice is "tempting" if it maximizes the temptation but not

[17]Internal conflicts may give rise to time inconsistent preferences, as in Strotz'(1955) seminal paper and in Laibson's (1997) analysis of hyperbolic discounting. As shown by Benabou and Pycia (2002), Gul and Pesendorfer's (2001) menu approach is coherent with the view that self-control problems arise from an intrapersonal conflict between different 'selves' (Planner and Doer, or Ego and Id, or prefrontal cortex and limbic system), along the lines of Thaler and Shefrin (1981) and of the psychological literature (see Fudenberg and Levine (2006) for a more recent

the commitment utility. Individuals can resist temptation only by exerting self-control, whose cost is measured by forgone temptation utility: making a choice $x \in \{\ell, c\}$, which does not maximize the temptation ranking v, imposes a cost of self-control equal to the difference between the highest attainable temptation utility and that derived from the chosen action, $\left[\max_{y\in\{\ell,c\}} v(y) - v(x)\right]$.

The optimal choice for tempted individuals involves a trade-off between the loss of commitment utility incurred for giving in to actions that do not maximize u, and the cost of self-control needed to resist temptation. Facing this trade-off, individuals optimally choose the action $x \in \{\ell, c\}$ that maximizes the commitment utility net of the cost of self-control,[18]

$$u(x) - \left[\max_{y\in\{\ell,c\}} v(y) - v(x)\right]. \tag{7}$$

2.2 Optimal Choices

In the standard case, in which agents do not suffer from self-control problems, legality is chosen whenever crime does not pay off in expectation: $r < \pi$. As discussed above, in presence of self-control problems, $\pi_v < \pi$, individuals are tempted whenever one action is best for u and the other one is best for v. For low levels of crime premium, $r < \pi_v$, legality represents the best option according to both rankings, so agents do not suffer from temptation. Symmetrically, when the crime premium is large enough, $r > \pi$, individuals are not tempted either, since crime is optimal for both rankings. In turn, the prohibited action is tempting for intermediate crime premia, for $r \in [\pi_v, \pi]$, since legality is optimal for the commitment ranking but not for the temptation ranking. Let us call this set of crime premia the "temptation range".

Outside the temptation range the optimal choices are straightforward: as discussed above, legality is chosen for $r < \pi_v$ and crime for $r > \pi$. Within the temptation range, for $r \in [\pi_v, \pi]$, the optimal choice is determined by the comparison between the utility loss from giving in to temptation, $u(\ell) - u(c) = \pi - r$, and the cost of self-control incurred to resist it, $v(c) - v(\ell) = r - \pi_v$. While the former is decreasing, the latter is increasing in r. Moreover, for $r = \pi_v$, the cost of giving in is positive but the cost of self-control is zero, whereas for $r = \pi$ the reverse is true. The two

version of the planner-doer approach), but allows to avoid time inconsistency. Relative to Gul and Pesendorfer (2001) we slightly abuse notation, since they define u and v on lotteries, that is, on probability distributions over outcomes, whereas we define utility over actions to simplify illustration. The two approaches are equivalent in this simple model since the choice of an action uniquely maps into the choice of a probability distribution over outcomes.

[18] Individuals are not tempted when the same choice maximizes both rankings, since in that case maximizing (7) amounts to maximizing $u(x)$.

costs are equal for $r = (\pi_v + \pi)/2$. As a result, for any given π and π_v, this threshold discriminates between crime and legality. For $r \in [\pi_v, (\pi_v + \pi)/2)$, agents face temptation and the cost of self-control is lower than that of giving in, so that temptation is resisted and legality is chosen. We call this the "range of self-control". For $r \in ((\pi_v + \pi)/2, \pi]$, the cost of self-control is too large to be worth bearing, temptation is overwhelming and the optimal individual choice is to give in to it and choose crime. We call this the "range of temptation-driven crimes" or "range of giving in". The set of crime premia for which agents choose crime (whether because they give in to temptation-driven crimes or because crime maximizes commitment utility), which is given by $r \in ((\pi_v + \pi)/2, \infty)$, is called "crime range". The complement set, $r \in (-\infty, (\pi_v + \pi)/2)$, is called "legality range".

The previous discussion is summarized in

Proposition 1. (Optimal Choices of Legality and Crime.) *For any* $\{r, \pi, \pi_v\}$,

- *If the crime premium is sufficiently low,* $r < \pi_v$*, the legal action is optimal for both the commitment and the temptation utility,* u *and* v*, and it is chosen without experiencing temptation;*

- *For intermediate crime premia,*

$$r \in [\pi_v, \pi], \tag{8}$$

  the legal action delivers the largest commitment utility, $u(\ell) \geq u(c)$*, but crime is tempting,* $v(\ell) \leq v(c)$*. In this "temptation range":*

  - *If* $r \in [\pi_v, (\pi_v + \pi)/2)$*, temptation is resisted, at a self-control cost equal to* $(r - \pi_v)$*;*
  - *If* $r \in ((\pi_v + \pi)/2, \pi]$*, temptation is overwhelming and crime is chosen, at a commitment utility cost equal to* $(\pi - r)$*.*[19]

- *If the crime premium is sufficiently large,* $r > \pi$*, crime is optimal for both the commitment and the temptation ranking,* u *and* v*, and it is chosen without experiencing temptation.*

Proposition 1 characterizes the range of crime premia for which agents are exposed to temptation and their equilibrium crime choices. The "temptation range" and the "crime range" are illustrated in Figure 1.

The cost of self-control increases in the crime premium r, since this makes crime more tempting. The cost of giving in to temptation decreases in r, however, since choosing crime is a 'mistake'

[19] We leave aside indifference between crime and legality, which is the case for $r = (\pi_v + \pi)/2$.

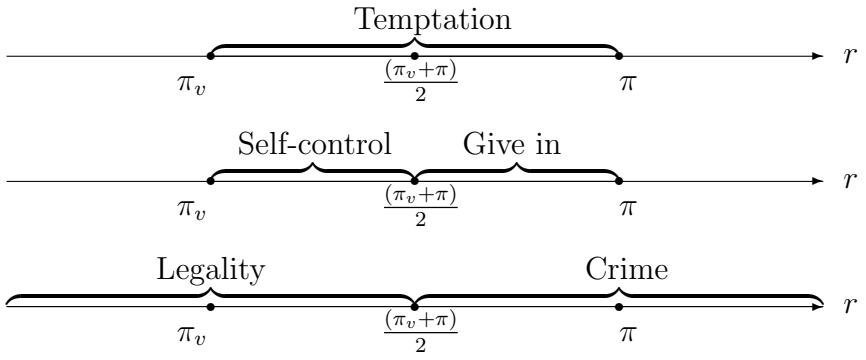


Figure 1: Graphical illustration of the ranges of "Temptation", "Self-control" and "Give-in" (or "Temptation driven-crimes"), "Legality" and "Crime" as a function of the crime premium, r.

in terms of commitment utility, but the larger the crime premium the smaller the mistake. The opposite effect of r on the cost of self-control and on the loss of commitment utility implies that indirect utility is non-monotonic in the returns to crime. For given legal return, within the temptation range $r \in [\pi_v, \pi]$ the indirect utility, given by expression (7) evaluated at the optimal choice, decreases in r as long as temptation is resisted, but increases in r once crime is chosen. Figure 2 depicts the commitment and temptation utility of crime, $u(c)$ and $v(c)$, as well as the indirect utility (the bold line), as functions of r.

2.3 Self-control, Punishment and Detection

Before turning to the analysis of the interactions between different types of prohibition and, in particular, to the role of internalized moral sanctions, it is useful to briefly study the effects of self-control problems, punishment intensity and detection probabilities.

Lemma 1. (Self-control problems.) *A marginal increase in impatience at the moment of choice (that is, a reduction of β_v)*

1. *Has no effect on either temptation or individual choices, if detection is only possible upon act ("in flagrante delicto"), that is, if either $d_1 = 1$ or $d_2 = 0$;*

2. *It broadens both the temptation range, the range of temptation-driven crimes and the crime range, if detection may also take place in the future, that is, if $d_1 < 1$ and $d_2 > 0$;*

Lemma 1 highlights how the effects of self-control problems depend on the time structure of

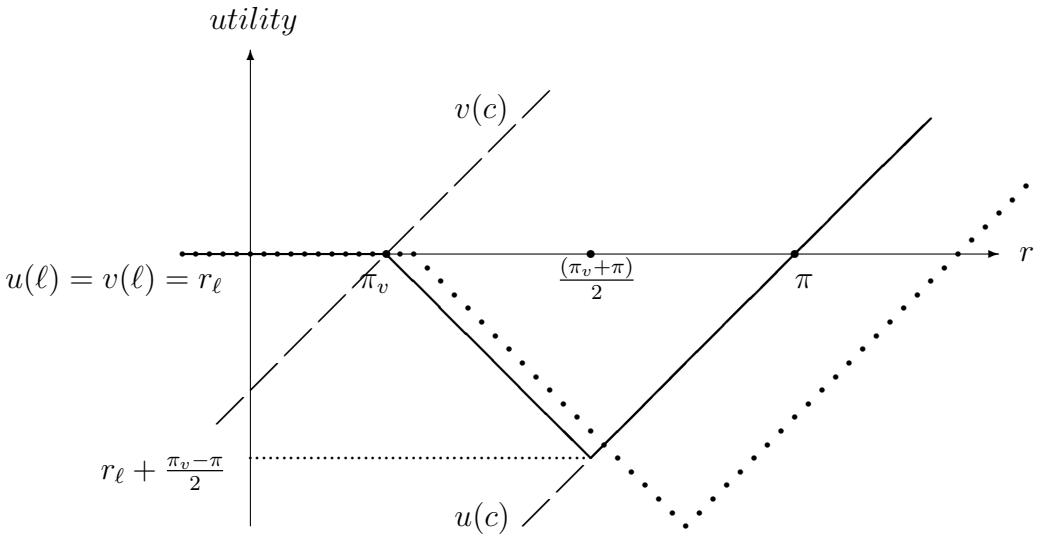


Figure 2: The upward sloping dashed lines depict commitment and temptation utility, $u(c)$ and $v(c)$, as function of the crime premium r (for given r_ℓ, π_v and π). The solid bold line depicts indirect utility: it is flat at r_ℓ for $r < \pi_v$, since legality is chosen without experiencing temptation; decreasing for $r \in (\pi_v, (\pi_v + \pi)/2)$, since temptation is resisted but the cost of self-control increases with r; increasing for $r > (\pi_v + \pi)/2$, since crime is chosen and (given r_ℓ) its returns increase with r. The dotted line illustrates the shift in indirect utility following a rise in the extent of punishment, p, and the consequent rightward shift and expansion of the temptation range.

detection. More serious self-control problems, leading to a temporary under-evaluation of future detection and punishment, expand the ranges of both temptation and temptation-driven crimes.

In the remainder of the paper we study the case in which self-control problems are present, so $\pi_v < \pi$. The effects of punishment intensity p are given in

Lemma 2. (Punishment.) *Consider individuals with self-control problems, so $\pi_v < \pi$. A marginal increase in punishment intensity, p, shrinks the crime range, but expands the ranges of temptation and of temptation-driven crimes.*

Lemma 2 highlights the contrasting effects of an increase in punishment intensity p. By raising expected punishment, it discourages crime. Yet, since temptation arises from an under-evaluation of expected punishment, a rise in p makes crime tempting for a wider range of crime premia, thereby increasing the range of temptation-driven crimes. From (6), the wedge between the evaluation of crime in the temptation and in the commitment ranking, which determines the length of the temptation range, is proportional to punishment intensity.

Particularly interesting for the purposes of this paper is the role of detection probabilities.

Lemma 3. (Detection.) *Consider individuals with self-control problems, so $\pi_v < \pi$. A marginal increase in the probability of*

1. immediate *detection, d_1, shrinks the ranges of crime, temptation and temptation-driven crimes;*

2. future *detection, d_2, shrinks the crime range, but expands the ranges of temptation and of temptation-driven crimes.*

Besides the intuitive result that an increase of any detection probability reduces crime, Lemma 3 highlights the opposite effects of immediate and future detection probability on the ranges of temptation and temptation-driven crimes. An increase in the probability of detection *in flagrante delicto* shrinks them, whereas an increase in future detection probability expands them.

If, as criminologists argue, many crimes are driven by temptation, then the observations in Lemmas 2 and 3 may provide a base for rationalizing the investment of public resources to raise immediate detection probability, like, for instance, increasing police presence on the streets.[20] Such rationale may be contrasted with the standard argument in crime economics, according to which high punishment threats would be preferable to high detection probability because, while both raise expected punishment, increases in detection probability are costlier to obtain. If temptation-driven crimes are relevant, then the cost of raising immediate detection probability might be worth bearing.[21]

3 Moral Values

Self-control problems are modeled as resulting from a (temporary) increase in impatience at the moment of actual choices. The results in the previous section illustrate that the prohibition of a

[20] This policy implication provides a further rationale for the arguments proposed by Davis (1988) based on the existence of high subjective discount rates.

[21] As we focus on an individual decision problem, we concentrate attention on the characterization of the relevant (e.g. temptation and giving-in) ranges. Providing empirically rooted policy implications appears beyond the scope of this paper. In fact, in order to map the theoretical predictions into policy prescriptions, one would need to know the actual (empirical) distribution of the population in the ranges of r and not only the expected change in the boundary of these ranges.

given action may create different temptation problems, depending on the way in which violations are detected and punished. As noted in the introduction, many actions are prohibited both by external codes, such as the law, and by internalized codes, such as moral values. A main difference is that detection and punishment are implemented by a third party (e.g., the state or the peers) in the former case, whereas in the latter case infringement triggers a cost of guilt based on self-awareness of misbehavior.[22] As baseline framework, we consider fully internalized moral values, which lead to the experience of self-conscious emotions even in the absence of any third-party detection and punishment.[23]

Section 3.1 extends the model to the consideration of two punishments with different time profiles of detection, which therefore generate different exposure to temptation. Since, as recalled above, self-detection is at the heart of internalized moral sanctions, it appears reasonable to assume that third-party sanctions rely on a more delayed (and possibly more uncertain) detection of infringements of prohibitions. A second relevant distinctive feature of moral sanctions is that, contrary to external punishment, they require an active (endogenous) process of education (or inculcation) to be effective in changing individual behavior.[24] The second part of the analysis, in Section 3.2, investigates the incentive for benevolent moral setters (like parents) to instill in their children moral values, which attach a cost of guilt to the infringement of moral prohibitions.

[22]Individuals' immediate emotional response when facing moral dilemmas (or even when contemplating breaking moral prohibitions) is widely documented by the experimental literature, using a variety of methodologies, from magnetic resonance (Green et al., 2001) to skin conductance (Coricelli, 2010). The recent neuro-experimental evidence by Koenigs et al. (2007) also shows the crucial role of emotions in the generation of moral judgements.

[23]The literature has extensively documented that guilt can be felt even in the absence of external detection. Guilt and shame (which is triggered by exposure to third-party punishments) can be related in non trivial ways, however (see Tangney and Dearing, 2002). While considering interactions between moral and social punishments (or between the experience of guilt and shame) does not change the main results, it allows to qualify some of the side predictions, as discussed below, obtaining results which are in line with those of Weibull and Villa (2005).

[24]Differently from other self-regulative emotions (such as, e.g., fear), moral emotions (like guilt) are not innate, but need to be inculcated (and in fact, contrary to fear, they cannot be experienced before the age of about three). The literature in social psychology documents that moral education has long-lasting effects (with the early phases in young age being the crucial ones) and that it changes behavior more effectively the higher is the achieved level of self-awareness of misconduct. Self-awareness limits the possibility of 'self-cheating' and thus, for instance, of telling oneself that an immoral action is not so wrong after all. See Katchadourian (2009) for an extensive analysis of the the role of moral education and guilt in enforcing internal punishments. Dal Bó and Dal Bó (2009) present experimental evidence on the effects of moral suasion, which awakes awareness of wrong-doing.

3.1 The Role of Moral Values

Consider a prohibited action (crime), which is subject to both external and internal (moral) sanctions, which are experienced upon detection and whose intensity is denoted by $p > 0$ and $\mu \geq 0$, respectively. If individuals are fully self-aware, then self-detection is immediate upon act, so that, denoting d_1^s the probability of immediate self-detection, we have $d_1^s = 1$. From equation (1) and Lemma 1, we then have that the expected moral sanction at the moment of choice is simply equal to μ, both in the commitment and in the temptation ranking.[25] By contrast, external detection of misbehavior need not take place upon act, but may also be delayed in time (if it takes place at all) so that, slightly abusing notation and letting from now on d_1 and d_2 denote external detection probabilities, we have $d_1 < 1$ and $d_2 > 0$. The commitment and temptation utility associated to the legal action are unchanged: $u(\ell) = v(\ell) = r_\ell$. The commitment and temptation utility from crime are given by $u(c) = r_c - \pi - \mu$ and $v(c) = r_c - \pi_v - \mu$ where, from now on, π and π_v denote the expected external punishment in the two rankings.

The only difference with respect to the model studied in Section 2 is therefore the presence of a moral punishment μ, which enters both the commitment and the temptation ranking (if $\mu = 0$ the model is identical the one studied above). Accordingly, the characterization of the equilibrium is analogous. In terms of equilibrium choices, the main difference is that the existence of a moral cost μ makes the prohibited action less attractive, so that individuals who would be exposed (and potentially would give in) to temptation may more easily resist it. As a result, the existence of a moral punishment shifts both the temptation and the crime range to higher intervals of crime premia. This implies that the predictions of Proposition 1 still hold, with the only modification that all the thresholds depicted in Figure 1 are shifted up by μ. The results reported in Lemma 1, Lemma 2 and Lemma 3 hold true as well, with p, d_1 and d_2 now referred to external punishment.

Morality may alleviate temptation problems, since the moral punishment μ reduces the cost of self-control. At the same time, however, the presence of moral values may be itself a source of temptation. Consider an individual who, absent moral values, would optimally chose crime without experiencing any temptation. Such an individual can be induced in temptation precisely by the existence of a moral prohibition. Moral values involve both expected costs and benefits for their carrier. They unambiguously reduce the crime range, but do not necessarily increase individual expected utility. Furthermore, the effect of a stronger moral punishment on indirect utility may

[25] From equation (1), $\mathbf{d^s} = (1, d_2^s)$ implies $\pi(\mathbf{d^s}, \mu, \beta) = \mu$, for any $\beta \in (0, 1)$ and $d_2 \in [0, 1]$.

be non-monotonic.

Proposition 2. (Effects of Moral Values.) *An increase in moral punishment μ:*

- *Shifts both the temptation range and the crime range to higher levels of crime premia;*[26]
- *Creates temptation if $r > \pi$ and μ rises from $\mu_0 < r - \pi$ to $\mu_1 \in (r - \pi, r - \pi_v)$;*
- *May have non-monotonic effects on indirect utility, which is given by*

$$U(\mu) = \begin{cases} r_\ell & \text{if } r < \pi_v + \mu \\ r_\ell - [r_c - r_\ell - \pi_v - \mu] & \text{if } r \in (\pi_v + \mu, (\pi_v + \pi)/2 + \mu) \\ r_c - \pi - \mu & \text{if } r > (\pi_v + \pi)/2 + \mu \end{cases} \tag{9}$$

Figures 3 and 4 provide a graphical illustration of the results of Proposition 2. Figure 3 shows that moral values shifts rightward the indirect utility function, $U(\mu)$, characterized in (9), represented, as in Figure 2, as a function of the crime premium (for given values of r_ℓ, π_v and π).

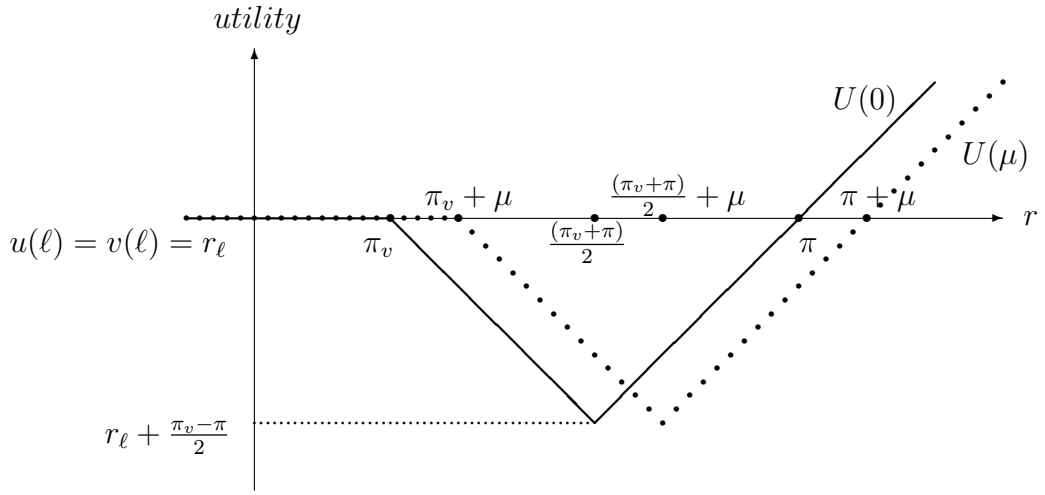


Figure 3: The solid bold line represents indirect utility in the absence of moral values, $U(0)$. The dotted line depicts indirect utility with moral values $U(\mu)$, for $\mu > 0$. As in Figure 2, indirect utility is drawn as a function of r, for given r_ℓ, π_v and π. Moral values shift all thresholds.

We next depict $U(\mu)$ in Figure 4 as a function of μ, for individuals with different legal returns r_ℓ (but the same values of r_c, π_v and π). Consider an individual whose legal returns, r_ℓ^L, are so low that his crime premium is larger than the expected external punishment, $r > \pi$. If $\mu = 0$

[26]The temptation range is now given by $r \in [\pi_v + \mu, \pi + \mu]$, and the crime range by $r > [(\pi_v + \pi)/2] + \mu$.

(that is in the absence of moral values) this individual faces no temptation and optimally 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. Nevertheless, if the moral punishment is large enough, $\mu > r - \pi$, then crime is not optimal for the commitment ranking u, although, for $\mu \in (r - \pi, r - \pi_v)$, it is still optimal for the temptation ranking v, so that the agent now faces temptation. In this case, as reported in Proposition 2, it is precisely the existence of the moral prohibition that creates temptation.

Whether the agent chooses crime or legality in this range of moral values depends on the comparison between the cost of self-control, $v(c) - v(\ell) = r - \pi_v - \mu$, and the cost of giving in, $u(\ell) - u(c) = \pi - r + \mu$. For any $\mu \in (r - \pi, r - (\pi_v + \pi)/2)$, the cost of self-control is larger than the cost of giving in, so the individual keeps choosing crime and feels guilty for this. In this range of moral values, an increase in their strength μ leads to a reduction in indirect utility, since it only increases the cost of guilt. For even higher levels of moral values, that is for $\mu \in (r - (\pi_v + \pi)/2, r - \pi_v)$, the cost of self-control gets lower than that of giving in, so the individual resists temptation, chooses legality and, accordingly, does not feel guilty. In this range, $U(\mu)$ is increasing in μ, since stronger moral values further reduce the cost of self-control. Finally, if moral values are strong enough, for $\mu > r - \pi_v$, crime stops being tempting and the self-control cost disappears, leaving $U(\mu) = r_\ell^L$. For an agent facing low legal returns such as r_ℓ^L, the indirect utility is therefore non monotonic in μ, but, crucially, it is highest for $\mu = 0$, since in that case choosing crime does not imply moral costs and indirect utility is maximized, because $U(0) = r_c - \pi > r_\ell^L$.

Consider next an agent whose legal returns, r_ℓ^{LM}, are a bit higher, but still in a lower-middle range, so that $r \in ((\pi_v + \pi)/2, \pi)$. While the indirect utility is again non-monotonic and qualitatively analogous to the previous case, such agent would give in to temptation if $\mu = 0$. His indirect utility is highest when the strength of moral values is large enough to eliminate temptation and let him abstain from crime.

For even larger legal returns, such as r_ℓ^{UM}, which is in the upper-middle range implying $r \in (\pi_v, (\pi_v + \pi)/2)$, the individual would resist temptation even in the absence of moral punishments, $\mu = 0$, although this requires exerting costly self control. Since morality reduces the costs of self-control, the indirect utility of this agent is monotonically increasing in μ. Finally, an individual whose legal returns, r_ℓ^U, are in the upper range implying $r < \pi_v$, does not face temptation and

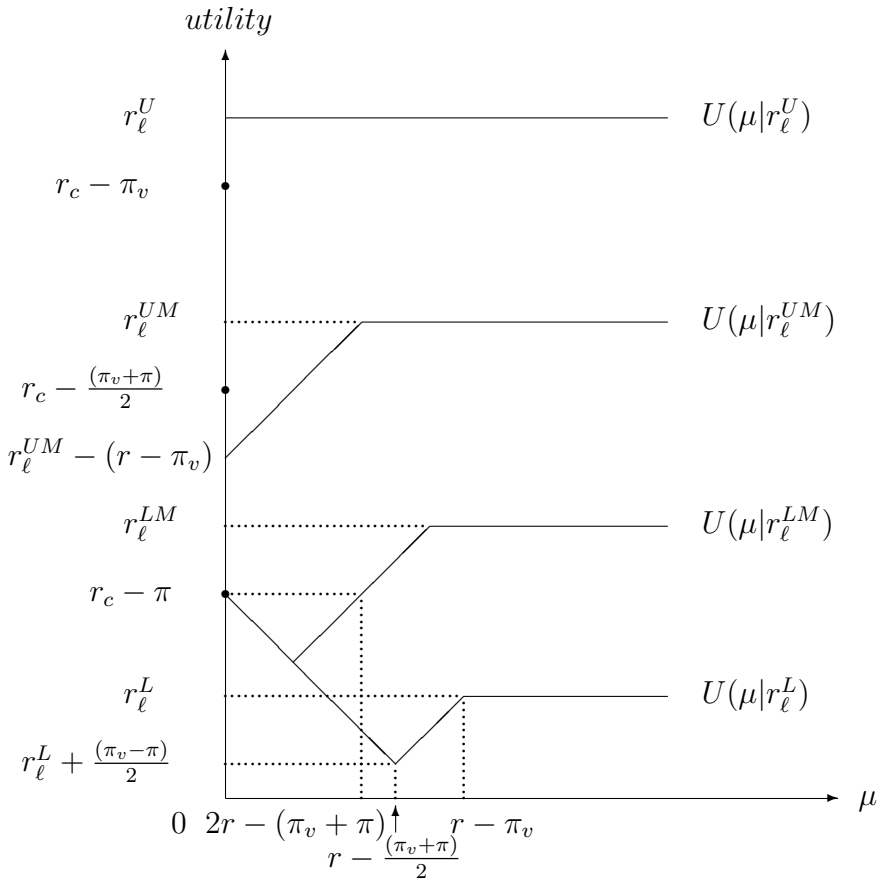


Figure 4: Indirect utility, $U(\mu)$, as a function of moral punishment μ for agents with different legal returns r_ℓ (but the same r_c, π_v and π). Absent morality (for $\mu = 0$), the crime premium falls in the four possible ranges. For either low or high legal returns, such as r_ℓ^L and r_ℓ^U, it is high ($r > \pi$) and low ($r < \pi_v$), respectively, so that crime and legality are chosen without experiencing temptation. For intermediate legal returns, such as r_ℓ^{LM} and r_ℓ^{UM}, it falls in the ranges of giving in, $r \in ((\pi_v + \pi)/2, \pi)$, and of self-control, $r \in (\pi_v, (\pi_v + \pi)/2)$, respectively. Temptation is experienced and it is resisted in the latter case but not in the former. Indirect utility is non-monotonic in μ for lower and lower-middle legal returns, whereas it is (weakly) monotonic for upper-middle and upper legal returns.

chooses the legal option, so his indirect utility is independent of μ.[27]

For those individuals who in the absence of moral values are not tempted, morality is either irrelevant or harmful. Those who choose legality without being tempted by crime are unaffected by the existence of moral values. Those who, absent morality, find crime optimal even for the

[27]Later on we will consider the implications of a heterogenous distribution of legal returns. Accordingly, the labels in Figure 4 will be interpreted as legal returns belonging to the lower, lower-middle, upper-middle and upper part of the distribution of legal returns.

commitment ranking, are harmed by morality, since it either involves a cost of guilt (if moral values are too 'weak', so that they keep choosing crime), or a cost of forgone net crime premium (if moral values are 'strong' enough to induce them to choose legality).[28]

A direct implication of Proposition 2 is that only individuals with intermediate crime premia gain, in terms of indirect utility, from being moral. This is formalized in

Corollary 1. (Who derives utility benefits from morality?) *For any given $\mu > 0$, being endowed with moral values, relative to having no morality, leaves agents*

- *Strictly better off if* $r \in \left(\pi_v, \min\left\{\pi, \frac{\mu+(\pi_v+\pi)}{2}\right\}\right)$;
- *Strictly worse off if* $r > \min\left\{\pi, \frac{\mu+(\pi_v+\pi)}{2}\right\}$;
- *Indifferent if* $r < \pi_v$.

Morality is beneficial in terms of indirect utility for all, and only, those individuals who would be tempted in its absence and who, if endowed with moral values, behave according to moral prescriptions (i.e., choose legality) at a sufficiently low self-control cost. Among those who would be tempted in the absence of morality, some would resist temptation anyway, but being moral allows them to resist at a lower cost of self-control. Their indirect utility thus increases in the strength of moral values, until the complete elimination of temptation. Other ones would give in to temptation in the absence of morality. For them, moral sanctions are valuable only if they are strong enough. A weak morality, insufficient to make them shift to legality, hurts them because it only makes them feel guilty for choosing crime, thereby adding a cost without providing any benefit. But even a morality level that is sufficient to make them resist temptation and shift to legality need not benefit them, because it may impose too high a cost of self-control. Such individuals may thus be harmed by a 'weak' morality and instead benefit from a 'strong' one.

Not all individuals who benefit from morality in terms of indirect utility also benefit from it in terms of material welfare, however. Material welfare is given by r_ℓ if the individual chooses legality and $r_c - \pi$ for crime. In equilibrium, morality leads to a strict increase in material welfare for all, and only, those individuals who by being moral change their equilibrium behavior. These are the individuals who, in the absence of internal punishments, would give in to temptation, but

[28]For any r there always exists a sufficiently large μ leading to legality. For any given value of $\mu > 0$, morality triggers legality for $r \in ((\pi_v + \pi)/2, (\pi_v + \pi)/2 + \mu)$, in the sense that crime would be chosen if $\mu = 0$. Yet moral individuals facing $r \in ((\pi_v + \pi)/2 + \mu, \pi + \mu]$ still give-in to temptation and choose crime.

choose legality once endowed with moral values.[29] The set of crime premia, for which morality is beneficial in material terms, is therefore smaller than the set for which it is beneficial in terms of indirect utility. Irrespective of whether one considers indirect utility or material welfare, it remains true that it is agents with intermediate crime premia (or intermediate legal returns, for given crime returns) who benefit from morality.

3.2 Moral Education

So far we have studied the role of exogenously given moral values. In this Section we investigate what would be the optimal moral education chosen by benevolent moral setters (henceforth called parents), who could activate a moral punishments upon misbehavior by some individuals (henceforth called children) and also influence its strength. As a benchmark we assume that parents have perfect knowledge of their children's expected net returns and self-control problems and that educational costs are negligible.[30]

If parents only face the binary choice of whether or not to instill moral values of honesty, whose strength $\mu > 0$ they regard as exogenously given, then the results of Section 3.1 directly apply. In particular, benevolent parents, no matter whether seeking to maximize their children's material welfare or overall indirect utility, will transmit a moral education to all and only those children who are expected to face intermediate levels of crime premia.[31] If parents can also choose the intensity of moral education, for instance by deciding how much time and effort to devote to it, then there is the additional dimension of the optimal strength of moral values. The main aspects of this educational process can be best appreciated by assuming that parents directly choose μ to maximize their children's material welfare. In this case, the optimal moral education

[29] Given $\mu > 0$, this set is given by $r \in \left(\frac{(\pi_v+\pi)}{2}, \min\left\{\pi, \frac{(\pi_v+\pi)}{2}+\mu\right\}\right)$. Notice that morality can be beneficial in terms of commitment utility only if it induces a shift from crime to legality. Once legality is chosen, the cost of guilt is not incurred and commitment utility coincides with legal returns, and thus with material welfare, and it is unaffected by moral values. It follows that in this model there is no difference between saying that μ provides material benefits and saying that it maximizes either material welfare or commitment utility.

[30] Assuming arbitrarily small educational costs serves the technical role of letting us focus on the minimum strength of moral values, which maximizes children's welfare. In particular, it implies that children with low crime premia, who are indifferent between receiving or not moral education, will not receive it. The implications of non negligible educational costs are discussed later.

[31] The specific range is characterized in Corollary 1 and footnote 29, for parents caring about indirect utility and material welfare, respectively.

is characterized in

Proposition 3. (Moral education.) *Benevolent parents with perfect knowledge of their children's expected net returns and self-control problems, who choose μ at negligible costs to maximize their children's expected material welfare (equivalently, their expected commitment utility), will provide some moral education ($\mu > 0$) to all, and only, children facing*

$$r \in \left(\frac{\pi_v + \pi}{2}, \pi \right). \tag{10}$$

Within this "morality range", they will provide moral values that are strong enough to induce legality, that is,

$$\mu^* = r - (\pi_v + \pi)/2. \tag{11}$$

According to Proposition 3, the strength of moral values that maximizes material welfare is non monotonic in the crime premium r or, equivalently, in relative legal returns. It is zero outside the "morality range" given in (10), and strictly positive within it. Over such range, the strength of morality, given in (11), continuously increases in the crime premium, until it discontinuously drops to zero when r is so large that crime becomes the action maximizing material welfare.

The results in Proposition 3 also imply that both larger external punishments (a higher p) and more serious self-control problems (a lower β_v) expand the morality range, (10). Nonetheless, within this range, they have opposite effects on the optimal strictness of moral education, μ^*. More serious problems of self-control (a lower β_v) induce parents to instill stricter moral values, as they are required to effectively change their children's behavior. In turn, higher external punishments, p, induce parents to transmit weaker moral values, because a higher p makes crime less attractive, so that a lower μ^* is sufficient to induce legality. This discussion is recorded in

Corollary 2. (Self Control, External Punishment and Optimal Morality.)

1. *The "morality range", given in (10), expands following both an increase in external punishments and more serious self-control problems;*

2. *Within the "morality range", the optimal level of moral education μ^*, given in (11), increases with self-control problems, but decreases with external punishments.*

Corollary 2 highlights that external punishments and moral values are complements at the extensive margin, but substitutes at the intensive one. In other words, stronger external punishments lead to a wider spread of more diluted (or weaker) moral values.

3.3 Discussion

Before moving to the empirical implications of the model in Section 4, let us briefly discuss the role of the main simplifying assumptions.

Material and utility benefits. The "morality range" is derived under the assumption that parents only care about their children's material welfare (or, equivalently, their commitment utility). This appears a natural benchmark for many reasons. Self-control costs may only reflect children's temporary difficulties in making the right choice, that is, a choice in line with their "true", long term interest, captured by the commitment ranking. Moreover, while it is easier to provide a rationale for the the transmission of potentially costly moral values on the basis of their effect on self-control costs, it is harder (and possibly more interesting) to do so on the basis of material benefits alone. From this perspective, we show that moral values naturally arise in presence of temptation, even under the restrictive assumption that their role is only to maximize individual material welfare. Finally, if potentially costly moral values provide material benefits, then their presence in a population is not only compatible with the view that they are rationally instilled by benevolent parents, but it would also be in line with a standard evolutionary (game theoretic) set-up, in which material payoffs determine fitness and evolutionary dynamics is payoff-monotonic.[32]

The qualitative results of Proposition 3 and Corollary 2 do not change, however, if parents also consider the cost of self-control that will be borne by their children and maximize their overall indirect utility, given in (9) in Proposition 2. The only difference is that in that case the "morality range" coincides with the (interior of the) "temptation range", given by (8) in Proposition 1, and therefore contains the one characterized by (10) in Proposition 3 as a strict subset. It remains true that moral education only takes place at crime premia within an intermediate range, that within such range the optimal strength of moral values is increasing in the crime premium, and that at the end of this range it discontinuously drops to zero.[33]

[32]Concentrating on material welfare also allows to clarify that the predictions cannot be imputed to circular reasoning as might be, for instance, the idea that parents instill self-harming values to maximize the utility of children with self-harming values. We thank two anonymous referees for independently suggesting this analysis.

[33]Parents maximizing indirect utility provide moral values to their tempted children not only to induce legality (which would leave out of moral education those who already resist temptation in its absence), but also to eliminate self-control costs, so now the morality range coincides with the entire temptation range. Within the temptation range, the minimum level of μ required to completely eliminate self-control costs and maximize indirect utility is

Punishments/Prizes and Costs/Benefits. As long as parents care for children's material welfare, the results on individually optimal moral education are independent of whether morality involves prizes, such as pride for appropriate behavior, or punishments, such as guilt for misbehavior.[34]

Costs of moral education. The consideration of non-negligible costs of moral education (increasing in μ) leaves the main predictions unchanged. Parents have to trade off these costs against children's expected material (or utility) benefits. The main difference is that, in presence of education costs, it may not be optimal for parents to provide moral education up to the complete elimination of either materially suboptimal behavior or of temptation (depending on whether they care for children's material welfare or overall utility). As a consequence, educational costs may shrink the morality range and induce parents to transmit weaker values, so that equilibrium moral values may only work as a partial commitment device.

Interactions between external and internal sanctions. For the sake of expositional clarity, we have so far considered an additively separable utility formulation, which rules out any direct interaction between external and internal punishments. Depending on the actual application and problem at hand, external and internal punishment might either complement or substitute each other directly. The main prediction that the morality range coincides with the range of temptation-driven crimes is robust to the consideration of direct interactions between external and internal sanctions in the utility function.[35] Yet, complementarity or substitutability would affect the optimal strength of moral values. For instance, one more year in jail may be harder if during that time the convicted also feels guilty. Similarly, if there is a component of shame to the internal punishment, ending up in jail, or being exposed to social stigma, might increase the cost of internal punishment by adding shame to guilt. In these cases, higher external sanctions make moral values a more effective commitment device, but this may reinforce the result obtained in Corollary 2, that an increase in external punishment actually weakens the optimal strength of moral values, since

$\hat{\mu} = r - \pi_v$, which is higher than the one required to simply induce legality, characterized by (11) in Proposition 3.

[34]If parents care for children's overall utility, moral prizes are a 'free lunch' for low crime premia. This would not be the case if one were to assume that increasing the pride associated to one choice automatically decreases the pride associated to the other one, as in Corneo and Jeanne (2009).

[35]We keep assuming that utility is linear in monetary returns (r_ℓ and r_c), with the marginal utility of money either not affected or reduced by higher punishments (whether external or internal). The same is true for the prediction that that the morality range coincides with the temptation range, if parents maximize children's indirect utility (including expected self-control costs), rather than their material welfare.

a weaker morality is sufficient to improve children's welfare. External and internal punishments might, nonetheless, also substitute each other, in which case the implications are symmetric to the complementarity case. For instance, being fined may reduce the feeling of guilt, since the agent may perceive that he has already paid enough for his misbehavior.[36]

4 Empirical Implications

The theoretical results, obtained so far in a deterministic model of individual decision making, have immediate empirical implications if applied to a context where, as in reality, relative legal returns are subject to randomness and are heterogenous in the population.

4.1 Random returns and *ex-post* sub-optimality of choices

The consideration of random returns allows to discuss the role of *ex-ante* and *ex-post* optimality in the choice of moral education. Suppose that returns to crime and legality are randomly extracted from two independent distributions, with expected values r_c and r_ℓ, respectively. Since parents base decisions on expectations, the analysis of moral education is identical to the one developed in the previous Section and optimal moral values are chosen to maximize expected material welfare (or utility). They are therefore *ex-ante* individually optimal or, equivalently, optimal in expectation. In the presence of randomness, they can be suboptimal *ex-post*, however. For instance, an individual may face (ex-post) an unexpectedly profitable illegal opportunity that would maximize both his commitment and temptation utility. Nonetheless, if endowed with moral values, he might refrain from breaking the prohibition. In such an instance, individual moral values, which are optimal given expected returns, turn out to be are 'too strict' given the actual realization of relative legal returns. Symmetrically, moral values may be ex-post 'too weak' if, for instance, they turn out to be insufficient to resist a temptation that is stronger than expected. In this case, as studied above, the double cost of guilt and suboptimal material benefits is paid.[37]

In presence of self-control problems and random returns, moral individuals may thus either commit crimes that do not pay off in material terms or refrain from profitable ones. Testing this

[36]This possibility is coherent with the findings by Gneezy and Rustichini (2000) that monetary fines may crowd out the perception of a certain behavior as a moral duty.

[37]A similar mismatch between the ex-ante optimality and the ex-post sub-optimality of moral education would emerge if the educational outcome μ is a random variable, which depends on parents' educational effort.

prediction is difficult, since evaluating the expected costs and benefits of a given prohibited action is generally far from easy. This, and the fact that individual preferences (like risk aversion) are not observable, makes it hard to establish whether a given behavior should be regarded as irrational in view of the standard rational theory of crime. Nevertheless, the literature in crime economics and criminology, already mentioned in Section 1, has repeatedly debated the seemingly 'puzzling' facts that individuals often commit crimes that appear clearly payoff dominated, or "stupid", but also often abstain from crimes that appear clearly "profitable".[38] While alternative explanations have been advanced to rationalize each of these behaviors within available theories of crime, to the best of our knowledge no single explanation has been put forward that could easily and naturally explain both types of "mistakes".[39]

The model provides a simple rationale for both types of seemingly irrational behavior (possibly by the same individual, if exposed to different realizations of returns) within a unique theoretical framework.

4.2 Heterogenous Returns and Morality

A specific prediction of the model, derived in Sections 3.1 and 3.2, is that moral values should be mostly beneficial, and would therefore be instilled by benevolent moral setters, to those individuals who are expected to face intermediate relative legal returns. While the model has been discussed with reference to a single individual, it can be interpreted in terms of heterogenous returns within a population. Suppose, for instance, that self-control problems, monetary returns and external punishment to crime are the same for everybody. Individuals with intermediate expected legal

[38] We refer to Gottfredson and Hirschi (1990) for an extensive discussion of examples of 'stupid' crimes. Individuals caught committing them often report that they were aware of acting stupidly but, at the same time, could not resist the temptation to do so. As an example of the role of morality for abstaining from profitable crimes, Becker argues, for instance, that '*if taxpayers responded only to the expected cost of evading taxes, evasion should be far more widespread*' and that risk aversion alone appears insufficient to rationalize observed patterns of behavior, adding: '*I believe it is necessary to recognize that most people believe they have a duty, moral or otherwise, to report their taxable income more or less honestly*' (see http://www.becker-posner-blog.com/archives/2007/11).

[39] A high risk aversion might explain why individuals abstain from seemingly profitable crimes, but makes "stupid" crimes even harder to rationalize. In turn, a high exponential discount rate (such as the one implied by Lee and McCrary's (2009) estimates on juvenile crime) might help explain why people sometimes indulge in crimes involving low payoffs and high punishments, but it makes it even harder to explain why people often refrain from easy and immediately profitable crimes.

returns should be the most moral, while those at the extremes of the distribution of expected returns should display less strict moral values.[40] Endogenous moral education should therefore result in a non monotonic pattern of morality. To our knowledge, the existing theoretical and empirically literature has not pointed out the existence of a hump-shaped pattern of morality across the distribution of expected relative legal returns.

Data. This prediction can be tested using individual-level data from the World Value Survey, WVS, which includes information for many countries and different periods of time, allowing to control for different dimensions of observed and unobserved heterogeneity.[41] The required information on the dependent and explanatory variables is available for the third, fourth and fifth wave of the WVS.[42] The base sample is composed by about 170,000 individuals, belonging to 79 countries (around 45,000 from 20 OECD countries and 125,000 from 59 non-OECD countries). Summary statistics fore the baseline sample are reported in Table 3 in the Appendix.

We use information on individual stances on "Claiming government benefits to which you are not entitled" (survey question V198), "Avoiding a fare on public transport" (question V199) and "Cheating on taxes if you have a chance" (question V200). For each question, individuals were asked "whether you think it can always be justified, never be justified or something in between", with answers ranging from 1 ("never justifiable") to 10 ("always justifiable"). To ease interpretation we reverse the scale so that higher values of a variable correspond to a stronger morality. The variables reporting the individual responses to each question are denoted *Benefits*, *Free-riding* and *Tax-cheating*. As a benchmark, we use a summary measure, called *Morality*, which is obtained by jointly considering the individual responses to each of the previous questions and summing the three variables.[43]

[40]The analysis immediately extends to the case in which crime returns r_c and self-control problems $(\pi - \pi_v)$ are heterogeneous as well, and their distribution is independent of that of legal returns r_ℓ. Legal and illegal returns might be positively correlated. If expected crime premia were increasing in legal returns, individuals with intermediate legal returns would still be the ones who receives moral education and the only difference would be that the rich, rather than the poor, would choose crime more frequently.

[41]We thank an anonymous referee for suggesting this empirical test.

[42]The five waves of the WVS report information from surveys conducted in the periods 1981-1984, 1989-1993, 1994-1999, 1999-2004 and 2005-2008. The last three waves have the largest coverage. Information on both moral values and social class is not available in the second wave, and it is available only for Argentina in the first wave (the inclusion of these data does not change the results).

[43]*Morality* summarizes, along the lines of Knack and Keefer (1997), the average individual response to the different

The database contains self-reported information on social classes and income. As benchmark explanatory variable we use information on social classes which, as discussed below, is correlated to income but should be expected to be also highly related to other relevant characteristics like financial wealth, house ownership, and social networks.[44] As explanatory variables we use *Lower*, *Working*, *Lower-middle*, *Upper-middle* and *Upper*, that are dummy indicators taking value 1 if an individual belongs to the respective social class and zero otherwise.

Results. Recall that the prediction to be tested is a non-monotonic relationship between individuals' (expected) relative legal returns and their morality.

As a first look at the data, Figure 5 depicts the average morality index for the individuals of each social class for one particular country, the US, and for the two groups of OECD and non-OECD countries.[45] In line with the hypothesis, in each sample individuals belonging to the extreme social classes tend to attribute a lower importance to civic duties. These patterns must be taken as purely suggestive, however, as they are based on row data, without controlling for any dimension of observed or unobserved heterogeneity, which might be relevant.

To provide a more systematic test, we run regressions using information on individual responses to the questions on moral behavior, individual social class and some relevant individual characteristics used as controls. The main explanatory variable of interest is social class. We run ordinary least square regressions where the reference (omitted) category is the *Working* class. The results report significance levels based on robust standard errors clustered at the country-wave-pair level.

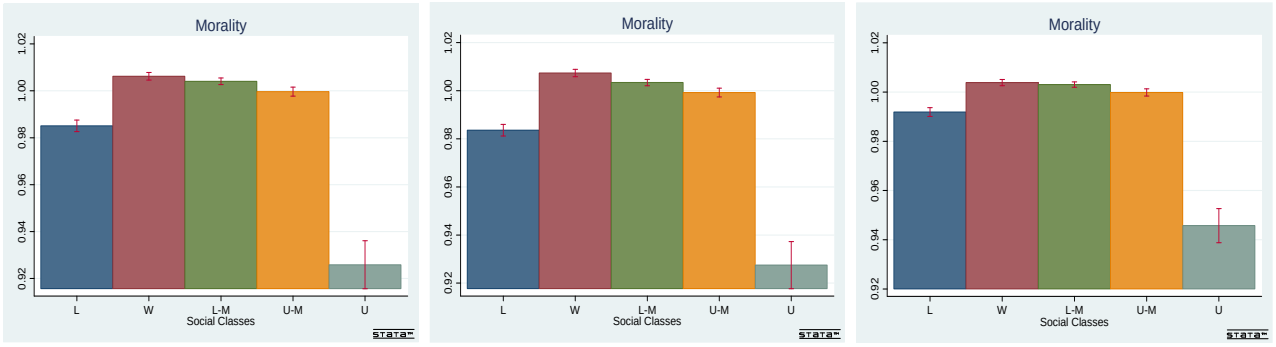
Table 1 reports benchmark results for *Morality* on the full sample of countries, using data from all waves. Column (1) includes, beyond social classes, country fixed effects to control for

questions about morally appropriate behaviors. Besides allowing to present the results more concisely, the variable *Morality* appears a reasonable empirical benchmark since the theory broadly refers to moral values and since we should expect the processes of moral education to be coherent across specific prohibitions. The results of several robustness checks on the definition of morality are discussed below.

[44]Being a broader concept, social class should be more persistent than income across different generations within a family. It should thus be a better proxy for parents' expectations on their children's future returns. As discussed in more detail below, the main qualitative pattern is similar with both social classes and income but in the latter case results are less precise and more difficult to interpret.

[45]The figure is based on all available data for the most recent decade (1999-2008) and includes the responses of about 100,000 individuals. For illustrative purposes, are normalized to get a sample mean of one in each sample. This shows that for instance, for the US, the average *Morality* of the upper class is almost 8% lower than the average in the population.

Figure 5: Social Classes and Moral Values



(a) US (b) OECD Countries (c) Non OECD Countries

Notes: Average moral values by social class (*Morality* normalized by sample mean), surrounded by a standard error above and below the mean. Source: World Value Surveys for US, OECD and Non OECD countries, 1999-2008.

country-specific unobserved heterogeneity in the importance of civic duties. Wave fixed effects are added in column (2) to control for changes over time that are common across countries. In column (3) the empirical specification further adds country-specific linear time trends to control for possible country specific changes in average morality over time. Finally, column (4) replaces the country, wave and trend controls with wave-specific country fixed effects, thereby accounting for time-varying and country-specific unobserved heterogeneity.

To interpret the results, recall that the omitted category is the working class. The estimated coefficients associated to each class therefore represent the deviation of the (average) morality of individuals in each class from the (average) morality of individuals of the working class (the reference group). In line with the theoretical predictions, the results document the existence of a significant hump-shaped pattern, which is stable across different specifications.

Table 2 extends the analysis to the consideration of individual characteristics in terms of age (and age square), gender, marital status, and employment status, which are likely to matter for individual morality. The last three variables are measured by the dummies *Male*, *Married* and *Unemployed*, where value 1 denotes the corresponding category.[46] Column (1), which refers to the full sample, shows that older and married individuals tend to attribute a larger importance to civic duties, males are less moral than females and unemployed are less moral than employed.[47]

[46]To avoid repetitions, we report only results for the specification with country-wave fixed effects, which is the most flexible and (therefore) conservative. Results for the other specifications are slightly more significant.

[47]The effect of age appears slightly concave. Notice that while the first two variables are clearly exogenous

Table 1: **Social Class and Morality**

Dependent Variable	Morality			
	(1)	(2)	(3)	(4)
Social Class:				
Lower	-0.444***	-0.442***	-0.414***	-0.408***
	[0.098]	[0.094]	[0.094]	[0.094]
Lower-Middle	-0.172**	-0.176**	-0.135**	-0.132**
	[0.064]	[0.063]	[0.061]	[0.060]
Upper-Middle	-0.252**	-0.256**	-0.262**	-0.253**
	[0.081]	[0.080]	[0.081]	[0.081]
Upper	-1.052***	-1.059***	-1.074***	-1.044***
	[0.285]	[0.286]	[0.283]	[0.283]
Country FE	yes	yes	yes	no
Wave FE	no	yes	yes	no
Country Trend	no	no	yes	no
Country-Wave FE	no	no	no	yes
Observations	170,481	170,481	170,481	170,481
R-Square	0.104	0.105	0.116	0.117

Notes: OLS regressions. The dependent variable *Morality* is the sum of the individual responses to the questions on claiming benefits, free-riding and cheating on taxes. Higher values mean higher agreement with the need to respect civic duties. The variables *Lower*, *Working*, *Lower-middle*, *Upper-middle*, and *Upper* are dummies for the respective social class, with *Working* being the omitted (reference) category. Column (1) includes as controls a constant and a full set of country dummies. Column (2) adds wave dummies. Column (3) adds country-specific linear time trends. Column (4) includes a full set of wave-specific country dummies. Significance level (***: 1%; **: 5%; *: 10%) based on robust standard errors clustered at each wave-country pair (reported in parenthesis). Data from the World Values Survey, 1994-2008. See text for further details.

The other columns adopt the same specification, but consider data from different sub-samples. Columns (2) and (3) consider the sub-samples of OECD and non-OECD countries, respectively, using data for all waves. This allows to investigate whether the non-monotonic pattern of individual morality in social class is specific to countries with lower or higher development. If something, in the developing countries the upper class tends to be relatively less moral.[48] Finally, columns (4), (5) and (6) show that the results for the single waves are very similar. The hump-shaped pattern predicted by the theory is confirmed in each sub-sample.

characteristics, employment status may be related to individual morality, for instance through work values, and we cannot exclude that people with a stronger morality tend to marry more. All the results are unaffected by the inclusion of these variables, however.

[48]While the qualitative pattern is confirmed, in some sub-samples the size of the upper class is very small (for instance, in OECD countries it is only 1.11 percent). See Table 4 in the Appendix for social class distribution in each sub-sample.

Table 2: **Social Class and Morality: Robustness to Individual Controls**

Dependent Variable	Morality					
	(1)	(2)	(3)	(4)	(5)	(6)
Social Class:						
Lower	-0.477***	-0.722***	-0.431***	-0.453***	-0.361**	-0.584**
	[0.091]	[0.127]	[0.107]	[0.124]	[0.158]	[0.179]
Lower-Middle	-0.072	-0.037	-0.091	-0.036	-0.077	-0.086
	[0.059]	[0.091]	[0.073]	[0.104]	[0.106]	[0.094]
Upper-Middle	-0.148**	-0.037	-0.207**	-0.197	-0.091	-0.142
	[0.075]	[0.106]	[0.097]	[0.141]	[0.150]	[0.098]
Upper	-0.934**	-0.436	-1.063**	-0.563	-0.754**	-1.482**
	[0.282]	[0.280]	[0.341]	[0.344]	[0.316]	[0.667]
Individual Controls:						
Age	0.051***	0.109***	0.045***	0.085***	0.033**	0.032**
	[0.008]	[0.013]	[0.009]	[0.012]	[0.013]	[0.013]
Age Squared	-0.000	-0.001***	-0.000	-0.000**	-0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Male	-0.397***	-0.659***	-0.295***	-0.556***	-0.297**	-0.311***
	[0.041]	[0.066]	[0.046]	[0.062]	[0.089]	[0.062]
Unemployed	-0.286***	-0.533**	-0.273**	-0.404**	-0.313**	-0.147
	[0.081]	[0.154]	[0.090]	[0.130]	[0.124]	[0.151]
Married	0.493***	0.600***	0.423***	0.601***	0.499***	0.420***
	[0.045]	[0.074]	[0.056]	[0.065]	[0.086]	[0.086]
Country-Wave FE	yes	yes	yes	yes	yes	yes
Observations	163,056	43,257	119,799	58,988	45,250	58,818
R-Square	0.137	0.135	0.139	0.143	0.124	0.123
Sample	Full	OECD	Non OECD	W3	W4	W5

Notes: OLS regressions of *Morality* on social class and individual controls. For *Morality* and social class variables, see note to Table 1. *Age* refers to the age of the responder. *Male*, *Married* and *Unemployed* are dummy variables taking value 1 for the respective individual characteristic and 0 otherwise. All specifications include a full set of wave-specific country dummies. Column (1) reports result for the full sample. Columns (2) and (3) replicate the analysis for the sub-samples of OECD and non-OECD countries. Columns (4), (5) and (6) replicates the analysis for the 1995, 2000 and 2005 wave, respectively, using data for all countries. Significance level (***: 1%; **: 5%; *: 10%) based on robust standard errors clustered at each wave-country pair in columns (1) to (3) and at the country level in columns (4) to (6) (reported in parenthesis). Data from the World Values Survey. See text for further details.

Robustness Checks. The results of Tables 1 and 2 are confirmed by a set of robustness checks. Before concluding, we comment on the most interesting among them.[49]

We have conducted several checks on the measure of morality. We have repeated the analysis using as dependent variable either the first principal component among the responses to the three questions on "Claiming benefits", "Free-Riding" and "Cheating on Taxes", or the responses to

[49]These results are available in the Supplementary Material.

each question separately.[50] Individual stances on different civic duties provide a consistent picture. In all cases, results are very similar and confirm both the existence of the hump-shaped pattern of morality in social class and its robustness to the inclusion of individual controls.[51]

Using information on income, we find that individuals in the second quintile attribute the largest importance to civic duties, whereas individuals at the two extremes of the income distribution are less moral. The pattern is thus identical to the one obtained with social classes, but the findings are more noisy and less statistically significant.[52] Since individual earnings are subject to life cycle, income can be a poor proxy for parents' expectations, especially in the case of older individuals. Moreover, income data refers to households and not to individuals. This renders the interpretation of the findings less straightforward since, for instance, we cannot disentangle a large household income produced by the work of two moral individuals (wife and husband), each with intermediate earnings, and one produced by a single individual with little morality. Testing the theory would require telling these two possible cases apart, but household income data do not allow it. To partially address the issues of life cycle and household composition, we have restricted attention to the sample of young and/or married individuals, finding that the non-monotonic pattern in income quintiles gets more pronounced and significant.[53] As a final check, we have also controlled for individual education. The existence of a hump-shaped morality pattern is confirmed for both social classes and income.[54]

[50]The unconditional correlation between the variable *Morality* and the variables *Benefits*, *Free-riding* and *Tax-cheating* is 0.78, 0.83 and 0.79, respectively.

[51]The hump-shaped pattern also emerges for individual stances on "Is it justified to accept a bribe in the course of one's duty?" (question V201). The upper classes tend to be least moral in terms of tax-cheating, whereas the lower classes in terms of claiming undue benefits. When including controls, the main differences is that unemployed individuals are significantly less moral in claiming benefits and free-riding, but not in cheating on taxes, and that males tend to display particularly low morality on tax-cheating.

[52]The correlation between income quintiles and social classes is 0.42. Responders are more reluctant to report information on their income than on social classes (the number of missing values is about 10 percent larger for the former than for the latter variable). This may create some selection bias.

[53]For social classes the results are essentially identical for the different subgroups. For income the results are more precise when restricting to young (below 35 years) and married responders, that is, when looking at households with at least two individuals (who presumably do not live with their parents). The result for the sample of young responders is also suggestive that the observed non-monotonic pattern is unlikely to be imputed to reverse causality, running from low morality to extremes economic returns (success and failure) later on in life.

[54]The whole analysis has been replicated using ordered probit regressions and it delivers qualitatively identical patterns. These results are available upon request.

5 Concluding Remarks

This paper proposes a simple theoretical framework studying the relationship between prohibitions, temptation and moral values. In presence of temptation, moral values may alleviate self-control problems by working as a self-commitment device. The model investigates the relationship between morality and temptation, the individual gains from morality, the interaction between external punishments and internal sanctions and the spread and strength of individually optimal moral values.

The analysis delivers novel predictions, which appear empirically relevant. Although morality may increase expected material welfare, moral individuals with self-control problems can indulge in stupid crimes or refrain from prohibited, but profitable, actions. The model may thus help reconciling, within a simple unique framework, casual observations that have been argued to be puzzling for standard theories. The theory shows that the pattern of individually optimal morality is hump-shaped. The existence of a non-monotonic pattern of morality is empirically confirmed by survey data.

In the model the main results are driven by *relative* returns. While we framed the problem in terms of crime and moral values, the analysis is more general. In particular, the results also apply to choices involving immediate costs and uncertain future benefits rather than immediate benefits and possible future costs (as in crime). In this perspective, the framework could also be fruitfully applied to the emergence of moral values of "thriftiness" and "industriousness", to the extent to which saving and hard work are subject to immediate costs and uncertain future returns in presence of self-control problems associated to the urge of experiencing immediate utility.

To focus on the interaction between prohibitions and moral values as self-commitment devices, the analysis has abstracted from several relevant issues. General equilibrium effects and social interactions are likely to be relevant for equilibrium choices, since they affect individual levels of crime premia and the strength of external (legal and social) punishments.[55] We have also abstracted from the consideration of parents' own morality and the role of social interactions for the success of moral education, which appears another relevant, non trivial issue. For instance, educating children to moral values may be less costly in communities with many moral individuals.

[55] As an example, Sah (1991) relates detection probability to the share of criminals, showing how multiple equilibria may emerge. The literature on crime determinants, discussed in Section 1, abounds of other possible sources of strategic complementarities.

In a dynamic environment, children's optimal moral values may differ from their parents' morality, possibly introducing additional educational costs and thus frictions in the evolution of morality. Moral education, social punishments and crime are therefore likely to be characterized by relevant dynamic interactions. Although addressing these issues was beyond the scope of the paper, they appear interesting avenues for future research.[56]

[56]The simple framework proposed in this paper may represent a natural basis for a dynamic investigation of moral education in a cultural transmission framework with heterogenous agents *a la* Bisin and Verdier (2001) or along the lines of Corneo and Jeanne (2010).

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 (2012): "Using the law to change the custom," *Journal of Development Economics*, 97(2), 182 – 200.

Becker, G. S. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2), 169–217.

Bénabou, R., and M. Pycia (2002): "Dynamic inconsistency and self-control: a planner-doer interpretation," *Economics Letters*, 77(3), 419–424.

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*, 52(1), 145–170.

Cervellati, M., J. Esteban, and L. Kranich (2010): "Work Values, Endogenous Sentiments and Redistribution," *Journal of Public Economics*, 94(9-10), 612 – 627.

Coricelli, G., M. Joffily, C. Montmarquette, and M. Villeval (2010): "Cheating, emotions, and rationality: an experiment on tax evasion," *Experimental Economics*, 13(2), 226–247.

Corneo, G., and O. Jeanne (2009): "A Theory of Tolerance," *Journal of Public Economics*, 93(5-6), 691–702.

——— (2010): "Symbolic values, occupational choice, and economic development," *European Economic Review*, 54(2), 237–251.

Crocker, J., and L. E. Park (2003): "Seeking Self-Esteem: Construction, Maintainance and Protection of Self-Worth," Handbook of Self and Identity (M. Leary and J. Tangney eds.), pp. 291–313. Guilford, New York.

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 (2012): "Self-Esteem, Moral Capital, and Wrongdoing," *Journal of the European Economic Association*, forthcoming.

Davis, M. L. (1988): "Time and Punishment: An Intertemporal Model of Crime," *Journal of Political Economy*, 96(2), 383–90.

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.

Fershtman, C., U. Gneezy, and M. Hoffman (2011): "Taboos and Identity: Considering the Unthinkable," *American Economic Journal: Microeconomics*, 3(2), 139–64.

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.

Fudenberg, D., and D. K. Levine (2006): "A Dual-Self Model of Impulse Control," *American Economic Review*, 96(5), 1449–1476.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996): "Crime and Social Interactions," *Quarterly Journal of Economics*, 111(2), 507–548.

Gneezy, U., and A. Rustichini (2000): "Pay Enough or Don't Pay at All," *The Quarterly Journal of Economics*, 115(3), 791–810.

Gottfredson, M. R., and T. Hirschi (1990): *A General Theory of Crime*. Stanford University Press.

Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen (2001): "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science*, 293(5537), 2105–2108.

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.

Katchadourian, H. A. (2009): *Guilt: The Bite of Conscience*. Stanford University Press, Stanford, California.

Knack, S., and P. Keefer (1997): "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *The Quarterly Journal of Economics*, 112(4), 1251–88.

Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. Damasio (2007): "Damage to the prefrontal cortex increases utilitarian moral judgements," *Nature*, 446(7138), 908–911.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 62, 443–477.

Lee, D., and J. McCrary (2009): "The Deterrence Effect of Prison: Dynamic Theory and Evidence," Mimeo, University of California at Berkeley.

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.

Lindbeck, A., S. Nyberg, and J. Weibull (1999): "Social Norms and Economic Incentives in the Welfare State," *The Quarterly Journal of Economics*, 114, 1–35.

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.

Robson, A. J., and L. Samuelson (2011): "The Evolutionary Foundations of Preferences," Handbook of Social Economics (J. Benhabib, A. Bisin and M. Jackson eds.), pp. 221 – 310. North-Holland.

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.

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.

Tangney, J. P., and R. L. Dearing (2002): *Shame and Guilt*. Guilford Press, New York.

Tangney, J. P., and J. L. Tracy (2012): "Self-Conscious Emotions," Handbook of Self and Identity, Second Edition (M. Leary and J. Tangney eds.), pp. 446–478. Guilford Press, New York.

Thaler, R. H., and H. M. Shefrin (1981): "An Economic Theory of Self-Control," *Journal of Political Economy*, 89(2), 392–406.

Weibull, J., and E. Villa (2005): "Crime, Punishment and Social Norms," *SSE/EFI Working Paper*, (610).

Appendix

Proofs

Proof of Proposition 1. The individual problem is $\max_{x\in\{\ell,c\}}\left\{u(x)-\left[\max_{y\in\{\ell,c\}}v(y)-v(x)\right]\right\}$, whose solution is $x^*=\arg\max_{x\in\{\ell,c\}}\{u(x)+v(x)\}$, which implies (leaving aside indifference) that:

$$x^*=c \iff u(c)+v(c)>u(\ell)+v(\ell)$$

Notice that $u(c)>u(\ell) \iff r>\pi$ and $v(c)>v(\ell) \iff r>\pi_v$. Moreover, if $r\in[\pi_v,\pi]$, then $u(c)+v(c)>u(\ell)+v(\ell) \iff r>(\pi_v+\pi)/2$. Hence, for $r<\pi_v$ legality is optimal for both rankings, whereas for $r>\pi$ crime is optimal for both rankings.

Proof of Lemma 1. If either $d_1=1$ or $d_2=0$, then for any $\beta_v<\beta$ and $p>0$, it holds that $\pi_v=\pi$, so the temptation range vanishes. If $d_1<1$ and $d_2>0$, then for any $\beta_v<\beta$ and $p>0$, it holds that $\pi_v<\pi$. In this case, a reduction in β_v reduces π_v, without affecting π. It thus broadens both the temptation range, $[\pi_v,\pi]$, and the range of temptation-driven crimes, $((\pi_v+\pi)/2,\pi]$, and thus also the overall crime range.

Proof of Lemma 2. For $\pi_v<\pi$ (i.e., for $d_1<1$ and $d_2>0$), a marginal increase in p broadens the ranges of both temptation and temptation-driven crimes, whose length is equal to $\pi-\pi_v=(1-d_1)d_2(\beta-\beta_v)p$ and to $\pi-[(\pi_v+\pi)/2]=[(\pi-\pi_v)/2]=(1-d_1)d_2(\beta-\beta_v)p/2$, respectively. Yet, it shrinks the crime range, which is the set of $r>(\pi_v+\pi)/2=[d_1+(1-d_1)d_2(\beta_v+\beta)]p/2$.

Proof of Lemma 3. When $\pi_v<\pi$ (i.e., for $d_1<1$ and $d_2>0$), it holds that $\partial\pi_v/\partial d_1>\partial\pi/\partial d_1>0$ and $\partial\pi/\partial d_2>\partial\pi_v/\partial d_2>0$.[57] The results on the ranges of crime, $((\pi_v+\pi)/2,\infty)$, temptation, $[\pi_v,\pi]$, and temptation-driven crimes, $((\pi_v+\pi)/2,\pi]$, are then immediate to derive.

Proof of Proposition 2. The proof of the first and third points follows directly from the analysis of Proposition 1, repeated with the new formulation of $u(c)$ and $v(c)$, which now includes μ; $U(\mu)$ in expression (9) is equal to expression (7) evaluated at the optimal choice. For the second point, notice that if $r-\pi_v>\mu>r-\pi>0$, then legality is optimal for the commitment ranking but crime is tempting. In this case, it is precisely the existence of the moral prohibition that creates

[57] To see this, recall that $\pi_v=[d_1+\beta_v(1-d_1)d_2]p$ and $\pi=[d_1+\beta(1-d_1)d_2]p$.

temptation, since, given the other parameters, with $\mu = 0$ (or any $\mu < r - \pi$) crime would be optimal according to both u and v.

Proof of Corollary 1. For $\mu > 0$, Proposition 2 implies that $U(\mu) > U(0)$ if, and only if, either $r \in (\pi_v, (\pi_v + \pi)/2]$ or both $r \in ((\pi_v + \pi)/2, \pi)$ and $\mu > 2r - (\pi_v + \pi)$. In the former case temptation is resisted in the absence of moral values; in the latter case, which amounts to $r \in \left((\pi_v + \pi)/2, \min\left\{\pi, \frac{\mu+(\pi_v+\pi)}{2}\right\}\right)$, individuals give in to temptation in the absence of morality and resist it at a sufficiently low cost of self-control when endowed with moral values. For $r > \min\left\{\pi, \frac{\mu+(\pi_v+\pi)}{2}\right\}$, it holds that $U(\mu) < U(0)$, whereas for $r < \pi_v$, $U(\mu) = U(0)$.

Proof of Proposition 3. For either $r \geq \pi$ or $r \leq [(\pi_v + \pi)]/2$, individual choices already maximize material benefits (or commitment utility), even without any moral education. For $r \in ((\pi_v + \pi)/2, \pi)$, the minimum strength of moral values necessary to induce legality is $\mu^* = r - (\pi_v + \pi)/2$.

Summary Statistics

Table 3: **Summary Statistics for the Full Sample**

Variable	Mean	Standard Deviation	Min	Max
Morality	25.74	5.61	3	30
Benefits	8.52	2.381	1	10
Free-riding	8.49	2.40	1	10
Tax-cheating	8.73	2.23	1	10
Age	40.58	15.92	15	99
Male	0.49	0.50	0	1
Unemployed	0.10	0.30	0	1
Married	0.58	0.49	0	1

Notes: Summary statistics for the baseline sample of Table 1.

Table 4: **Distribution of Social Classes by Sub-Samples**

Sample \ Social Class	Lower	Working	L-M	U-M	Upper	Total Obs.
Full	23,276 [13.65%]	47,552 [27.89%]	65,049 [38.16%]	31,905 [18.71%]	2,699 [1.58%]	170,481
OECD Countries	3,147 [6.91%]	13,460 [29.57%]	17,673 [38.83%]	10,732 [23.58%]	507 [1.11%]	45,519
Non OECD Countries	20,129 [16.11%]	34,092 [27.28%]	47,376 [37.91%]	21,173 [16.94%]	2,192 [1.75%]	124,962
Wave 3	6,948 [11.07%]	18,345 [29.23%]	25,726 [40.99%]	10,680 [17.02%]	1,058 [1.69%]	62,757
Wave 4	7,263 [15.83%]	12,188 [26.57%]	16,767 [36.55%]	8,891 [19.38%]	765 [1.67%]	45,874
Wave 5	9,065 [14.66%]	17,019 [27.52%]	22,556 [36.47%]	12,334 [19.94%]	876 [1.42%]	61,850

Notes: Absolute and relative frequency of each social class (*Lower*, *Working*, *Lower-middle*, *Upper-middle* and *Upper*) in the full sample and in each of the sub-samples considered in Table 2.