Optimal Law Enforcement with Sophisticated and Naive Offenders*

Berno Buechel†, Eberhard Feess‡, & Gerd Muehlheusser§

February 5, 2019

Abstract

Research in criminology has shown that the perceived risk of apprehension often differs substantially from the true level. To incorporate this insight, we extend the standard economic model of law enforcement (Becker, 1968) by considering two types of offenders, sophisticates and naïves. Sophisticates always fully take into account the actual enforcement effort, while naïves do so only when the effort is revealed by the authority. Otherwise, naïves rely on their fixed perceptions. When the share of naïves is high, a welfare-maximizing authority chooses a low enforcement effort, which is over-estimated by the naïves. Otherwise, it chooses a high enforcement effort, which is then revealed to all potential offenders. In three empirically relevant extensions, we allow for lower efficacy of the enforcement effort due to avoidance activities, endogenous fines, and heterogeneity with respect to naïves' perceptions.

JEL-Code: K42, D73

Keywords: Optimal Law Enforcement, Deterrence, Behavioral Law & Economics, Shrouding, Anchoring Bias, Salience

*We gratefully acknowledge comments and suggestions by Stefan Buehler, Robert Dur, Holger Herz, Johann Graf Lambsdorff, Urs Schweizer, and Avi Tabbach as well as by participants of the 2018 International Meeting in Law & Economics in Paris. Pamela Mertens and Paul Strikker provided excellent research assistance. Declarations of interest (all three authors): none.

†University of Fribourg, Department of Economics, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland. berno.buechel@unifr.ch; www.berno.info

‡Victoria University of Wellington, School of Economics and Finance, 23 Lambton Quay (Pipitea Campus), Wellington, New Zealand. eberhard.feess@vuw.ac.nz

§Corresponding author. University of Hamburg, Department of Economics, Von-Melle-Park 5, D-20146 Hamburg, Germany; and IZA and CESifo. gerd.muehlheusser@uni-hamburg.de; www.wiso.uni-hamburg.de/io

1 Introduction

The social and economic costs of crime and other illegal activities render the enforcement of the law an important task for any policy maker. In economic models of law enforcement in the tradition of Becker (1968), an illegal act is committed if and only if the offender's benefit exceeds the expected punishment. A standard assumption in these models is that every potential offender can correctly take into account the expected punishment, which relies on the probability of apprehension (see the survey by Polinsky and Shavell, 2007). However, this assumption is at odds with empirical evidence showing that the correlation between the perceived and the actual probability of apprehension is often weak (see e.g., Pogarsky and Piquero, 2003; Lochner, 2007). Some economic models incorporate this by assuming that offenders receive noisy signals on the true probability of apprehension (Bebchuk and Kaplow, 1992; Garoupa, 1999). In these studies, however, offenders are still unbiased in the sense that the mean of the belief distribution is equal to the actual detection probability.

By contrast, empirical evidence in criminology suggests that apart from the actual enforcement effort, offenders' perceived probability of apprehension might also be driven by other factors such as own experience (Matsueda et al., 2006; Anwar and Loughran, 2011) and observations in their social network (Stafford and Warr, 1993; Paternoster and Piquero, 1995; Apel and Nagin, 2011). Such factors might play the role of "anchors" in the sense of Tversky and Kahneman (1974) and as a result, perceptions about the probability of apprehension would differ from the level actually chosen (Kleck et al., 2005).

In our framework, we account for this possibility by considering two types of offenders, *sophisticates* and *naïves*.[1] Sophisticated offenders correctly take into account the authority's enforcement effort and the resulting probability of apprehension, and hence behave as in the Becker (1968) model of law enforcement. By contrast, unless the enforcement authority reveals its effort, naïve offenders have a perception about it which might lead them to over- or underestimate the actual probability of apprehension. The presence of offender naïveté raises the issue of how this affects the optimal enforcement policy.

Consequently, this paper aims at contributing to the following two research questions: First, how does the socially optimal enforcement effort and the resulting welfare depend on the share of naïve offenders in the population? Second, when does the enforcement authority want to reduce (or even eliminate) existing misperceptions by revealing its

[1] This terminology is borrowed from the literature on behavioral industrial organization, see e.g., Gabaix and Laibson (2006); Armstrong and Vickers (2012); Heidhues, Kőszegi, and Murooka (2017).

actual effort to all potential offenders?[2]

With respect to the first research question, we first analyze a baseline model where the actual (hidden) enforcement effort only affects the deterrence of sophisticates, while the deterrence of naïves is fully driven by their exogenous perception about this effort. We show that the optimal effort is then decreasing in the share of naïves. Moreover, when this share is small, the optimal enforcement effort is relatively large and exceeds the perception of naïves about it, and hence they underestimate the probability of apprehension. As a result, deterrence is stronger for sophisticates compared to naïves. Analogously, when the share of naïves is large, the optimal enforcement effort is smaller than the naïves' perception about it, i.e., they overestimate the probability of apprehension. As a consequence, the level of deterrence of sophisticates is lower compared to naïves. These properties give rise to a (maximized) social welfare function which is U-shaped in the share of naïves in the population. In equilibrium, there is either a high enforcement effort leading naïves to underestimate the probability of apprehension (when the share of naïves is low) or a low enforcement effort leading naïves to overestimate it (when this share is high). This is in line with empirical evidence (see e.g., Kleck et al., 2005), according to which individuals often underestimate the probability of apprehension for serious crimes, while they overestimate it for less serious ones.

For the second research question, we also consider a regime in which the enforcement authority can increase the salience of its enforcement policy by revealing the chosen effort to all potential offenders (see e.g., Apel, 2013; Chalfin and McCrary, 2017), thereby effectively eliminating the distinction between sophisticates and naïves.[3] We then compare the social welfare arising under each of the two enforcement regimes, which points to the following countervailing effects: With hiding, the enforcement effort only affects the sophisticated offenders, but not the naïve ones, so that the optimal effort is lower compared to the regime with revealed effort. This saves on enforcement costs, but it also leads to a relatively low deterrence of sophisticates. In addition, it induces different gain thresholds for sophisticated and naïve offenders, thereby leading to a distortion in the sense that a

[2] The second question has a trivial answer when a bias is simply added to the true enforcement effort: If the bias is positive (negative) such that the perceived effort is always higher (or lower) than the actual effort, then it is optimal not to debias (to debias) offenders. However, as soon as the perception of naïves is related to some anchor, as in our model, the chosen effort level can be both over- or under-estimated. As will become clear below, this makes the analysis considerably more intriguing.

[3] For instance, the public transport authority of the German city of Frankfurt recently announced that the number of ticket inspectors has increased by 52%; and the information was prominently reported in newspapers and reliable internet sources, see e.g., `http://www.fr.de/rhein-main/verkehr/rmv-und-vgf-schaerfere-kontrollen-gegen-schwarzfahrer-a-567360`.

given number of offenses is not committed by the offenders with the largest gains. This distortion does not arise in the regime with revealed effort.

Our analysis of both research questions highlights that the optimal policy crucially depends on the share of naïves in the population.[4] To see this, assume that naive offenders assume a probability of apprehension of 50 percent and that, if all offenders were sophisticated, the authority would minimize social costs by implementing a probability of 70 percent. If the share of naïves is large, however, then the authority prefers to choose a low effort which only leads to a probability of, say, 30 percent, and hide this fact to the naïves who are still deterred according to their perception of 50 percent. By contrast, if the share of naïves is low, then the authority implements a probability of 70% which is then revealed to all offenders. More generally, when the share of naïves in the population of offenders is small, the optimal enforcement policy stipulates a relatively high effort, which is then revealed to all offenders. Otherwise, the optimal effort is relatively low and remains hidden (and hence leads naïves to overestimate the probability of apprehension).

We extend our basic model in three empirically relevant directions. The first extension is motivated by the possibility that revealing the effort facilitates evasion activities.[5] For example, the New York City Council recently introduced a policy aimed at increasing the transparency and accountability over the use of new surveillance tools by the policy. While this policy aims at reducing the risk of information abuse, officials fear that it may also reduce the efficacy of their effort.[6] The lower efficacy of effort when revealed increases the parameter range for which hiding is optimal. Thereby, a non-monotonicity emerges as hiding the effort can be optimal for both low shares and high shares of naïves, but not in the intermediate range.

In a second extension, we assume that, in addition to the enforcement effort, the authority can also decide on the fine level. We find that the classical result that imposing

[4]In reality, the share of naïves might well depend on, among others, the type of crime, the enforcement technologies, the age of the potential offenders, or their media consumption habits. Moreover, new enforcement technologies such as, for instance, machine learning algorithms to detect violators, may not only affect the costs and benefits of optimal enforcement (for which the comparative statics are well-known), but also the share of naïves.

[5]Since our focus is on the regime comparison, we do not model such activities explicitly, but see previous studies on avoidance activities such as Langlais (2008) Nussim and Tabbach (2009), Sanchirico (2006) and Tabbach (2009). Another reason for reduced efficacy of revealing the effort are displacement effects. There is empirical evidence that revealing a high regional concentration of police reduces crime e.g. for motor vehicle thefts (Di Tella and Schargrodsky, 2004; Draca, Machin, and Witt, 2011), but part of the benefits is due to spatial displacements of crime (Donohue, Ho, and Leahy, 2013).

[6]See e.g., `https://www.huffingtonpost.com/entry/new-york-city-is-making-its-citizens-safer-by-overseeing-police-technology_us_58e23f04e4b0ba359596583b`.

the maximal fine is optimal Becker (1968) does not extend to the case with hidden effort: With the maximum fine and low effort, too many naïve offenders with high benefits from the act would be deterred as they overestimate the probability of apprehension. Hence, our analysis adds a further and novel reason why imposing the maximum fine may not be optimal.[7]

In the third extension, we allow for heterogeneity of perceptions by naïve offenders. This is motivated by a number of empirical studies which show that the accuracy of perceptions varies considerably among subjects (Lochner, 2007; Apel, 2013). We first demonstrate the robustness of our basic model as, for any given enforcement policy and any model with heterogeneous perceptions, there is one with homogeneous perceptions leading to the same welfare level. However, heterogeneity reinforces the inefficiency with hidden effort that some offenders with high private benefits may be deterred, while others with low benefits commit the act. As a result, this extension provides an additional argument in favor of revealing the effort.

Last, but not least, in the baseline model, the deterrence of naïve offenders is independent of the actual enforcement effort and, hence, solely driven by their exogenous perceptions about this effort. While a useful starting point. this assumption is highly stylized. In particular, while the empirical evidence discussed above suggests that offenders' beliefs concerning their probability of apprehension are often biased, this does not generally imply that these beliefs are independent of actual enforcement effort.Consequently, we also consider a model extension where the effort as perceived by naïves is a convex combination of the actual effort and some exogenous component, which can be interpreted as an "anchor" in the sense of Tversky and Kahneman (1974). We show that the main results of the baseline model are robust also in the extended model, as long as the weight put on the anchor is sufficiently large.

The remainder of the paper is organized as follows: Section 2 discusses the related literature. Section 3 introduces the basic model. Section 4 characterizes the optimal enforcement policy. Section 5 considers three model extensions: a lower effectiveness of revealed enforcement effort (Section 5.1), endogenous fines (Section 5.2), and heterogeneity with respect to naïves' perceptions (Section 5.3). We conclude and point to further research in Section 6. All proofs are in Appendix A. Appendix B contains the analysis of

[7]The previous literature has already identified a number of reasons why maximum fines may not be optimal, for example, costs of fine collection, the requirement that the punishment should reflect the severity of the offense, offenders' risk aversion, offenders' heterogeneity with respect to wealth, or offenders who engage in socially undesirable avoidance activities. See the survey by Polinsky and Shavell (2007) for a detailed discussion of these factors.

the more general modeling of belief-formation of naïve offenders.

2 Relation to the Literature

Apel and Nagin (2011), Nagin (2013), and Chalfin and McCrary (2017) provide comprehensive surveys of the criminological literature that has motivated our paper. Our model framework itself is related to the following strands of literature:

First, there are studies where potential offenders get noisy but unbiased signals about the true probability of apprehension. For example, Garoupa (1999) shows that revealing the effort is then always welfare enhancing as heterogeneous beliefs mis-allocate offenses among subjects with high and low private benefits. This is different in our model with biased beliefs as the ranking between revealing and hiding depends on the prior and the share of naïves in the population. The model with uncertainty by Bebchuk and Kaplow (1992) is related to our third extension as they also find that heterogeneous beliefs provide a rationale for fines below the maximum one. While the reason in their approach is that higher fines generally aggravate the problem of different perceptions, maximum fines in our model may deter too many naïves. Therefore, it may be superior to deter sophisticates with higher effort than with higher fines. Lang (2017) shows that legal uncertainty, defined as uncertainty about whether an activity is forbidden or not, may enhance social welfare as only offenders with high private benefit will commit an offense.

Second, with respect to heterogeneous perceptions of offenders, Ben-Shahar (1997) analyzes a setting where heterogeneous perceptions yield a misallocation of offenses between individuals with high and low private gains. In his two-period model, individuals learn the apprehension risk after having been caught once. This sets incentives to set low fines in the first period as this increases first-period arrests, thereby increasing the percentage of individuals who commit the offense in the second period only when their private gains are large. D'Antoni (2018) shows that initially unbiased uncertainty about the probability of apprehension translates into a systematic overestimation in a dynamic process where only violators learn the true probability. The reason is that offenders underestimating the true probability have higher violation incentives, so that their bias is reduced while the overestimation of non-violators persists.

Third, with respect to commitment, in our framework the authority can credibly reveal its detection effort. Conversely, Baumann and Friehe (2013) consider a cheap-talk game and show that credible information transmission requires certain conditions on the levels of harm, sanctions, and the social costs of fines. As in our model, hiding the information may be optimal for the authority, but offenders always benefit from revelation as they can adjust their decisions accordingly. In our model, naïve offenders may either benefit or suffer when the authority reveals its effort: On the one hand, they can adjust their behavior accurately to the actual probability of apprehension. On the other hand, the probability of apprehension is always higher with revelation.

Fourth, the possibility to either hide or reveal the enforcement policy, thereby affecting its salience, is also a crucial issue in studies of so-called crackdowns (Eeckhout, Persico, and Todd, 2010; Lazear, 2006). These are phases and/or regions of very high enforcement effort, e.g., controls for speeding for one day in one part of a city, which are announced in advance. The announcement leads to higher deterrence for the group targeted by the enforcement authority but may reduce the deterrence of other offenders. Revealing a focused effort is optimal when many potential offenders otherwise perceive a low overall probability of apprehension. In our framework, this corresponds to the case of a large share of naïves with a low perception of enforcement effort. Moreover, in a field experiment in the context of littering, Dur and Vollaard (2018) find that increasing the salience of the enforcement policy through simple warning labels close to garbage collection areas reduces the number of illegally disposed garbage bags. In line with our framework, this points to underestimation of the probability of apprehension, before the policy was revealed.[8]

Fifth, with respect to the origin of perceptions of naïve offenders, we treat these as exogenously given, without explicitly considering where they come from. This is complementary to Sah (1991), who develops a rich model where perceptions may change over time, but treats the authority's policy as exogenous. By contrast, our model is static, but we derive the optimal enforcement policy.

Finally, our distinction between two offender types and whether crucial information should be revealed to overcome naiveté is drawn from a recent literature in behavioral industrial organization (Gabaix and Laibson, 2006; Heidhues, Kőszegi, and Murooka, 2012; Armstrong and Vickers, 2012; Heidhues, Kőszegi, and Murooka, 2017), where consumers differ in their ability to fully grasp all attributes of a sales contract. This literature focuses on firm behavior and profits in a competitive environment, and also explores potential welfare implications. In our setting, there is no competition and the policy is chosen by a monopolistic authority that acts as a social welfare maximizer.[9] In line with this litera-

[8]The effects of the salience of policies has also been studied in other contexts such as taxation and incentive schemes, see e.g. Brown et al. (2010), Chetty et al. (2009), Englmaier et al. (2016).

[9]Buehler and Eschenbaum (2017) study a model which encompasses both welfare maximization and profit maximization as special cases.

ture, we also find that through the induced policy decisions, in equilibrium, sophisticated agents may benefit from the presence of naïve ones.

To summarize, while there are some similarities with models from several strands of literature, our study is the first to investigate optimal law enforcement in a setting with sophisticated and naïve offenders.

3 Model

Law enforcement is conducted by an authority which takes two decisions: a level of enforcement effort $e \geq 0$, and whether to *hide* (H) or to publicly *reveal* (R) it to the potential offenders.[10]

There is a unit mass of (risk-neutral) individuals who differ in their gains $g \in \mathbb{R}$ from committing an offense. Gains are distributed according to a cumulative distribution function G, which is twice continuously differentiable and strictly increasing, satisfying $G'(g) < \infty$ for all g. Each offense leads to a social harm $h > 0$.

We distinguish two types of offenders. A fraction $(1 - a)$ (with $0 \leq a \leq 1$) is *sophisticated* in the sense that they always take into account the authority's true enforcement effort e, irrespective of whether or not it has been revealed. The remaining fraction $a \in [0, 1]$ of offenders is *naïve* in the sense that they take into account the true enforcement effort e only when it is publicly revealed by the authority.[11] When it remains hidden, they perceive it to be $\hat{e} \geq 0$ instead. In the basic model we assume that $\hat{e}$ is exogenously given from the viewpoint of the enforcement authority. However, we show in Appendix B that the results are qualitatively the same, when the perceived effort is a convex combination between the true effort and some exogenous component, as long as there is sufficient weight on the latter.[12] As mentioned above, the level of $\hat{e}$ might well depend on the type of offense under consideration and other situation-specific factors. In the baseline model all naïve agents have the same perception $\hat{e}$ (see Section 5.3 for the case of heterogeneous perceptions. Moreover, the gain distribution G applies to both sophisticated and naïve offenders.

[10] In the behavioral industrial organization literature discussed above, the revealing (hiding) of the chosen policy is often referred to as unshrouding (shrouding), see e.g., Gabaix and Laibson (2006); Heidhues, Kőszegi, and Murooka (2017).

[11] In this respect, the standard economic model of law enforcement (see e.g., Becker, 1968; Polinsky and Shavell, 2000, 2007) is nested in our model when either the effort is revealed or when all offenders are sophisticates ($a = 0$).

[12] As will be discussed in Appendix B below, this exogenous component can be naturally interpreted as an anchor in the sense of Tversky and Kahneman (1974).

Irrespective of the type, each offender is detected with probability $p(e)$, which satisfies $p(0) = 0$ and which is twice continuously differentiable and strictly increasing for all effort levels e satisfying $p(e) < 1$ and $p'(e) < \infty$. In deciding whether or not to commit the offense, sophisticates always take into account the actual probability $p(e)$. By contrast, naïves perceive it to be $p(\hat{e})$ when the enforcement effort is hidden, and $p(e)$ when it is revealed. Each detected offender is subject to a fine $f > 0$. In the basic model, we treat f as exogenous, while the case where it also becomes a choice variable of the enforcement authority is considered in Section 5.2. The cost of enforcement effort e is given by a function $C(e)$, which is twice continuously differentiable, satisfying $C(0) = 0$, $C'(e) > 0$, $C''(e) \geq 0$, as well as the Inada conditions $C'(0) = 0$ and $\lim_{e \to p^{-1}(1)} C'(e) = \infty$.

The sequence of events is as follows: At stage 1, the enforcement authority decides on its effort and on whether to reveal or to hide it. At stage 2, each individual decides on whether or not to commit the offense.

At stage 2 each offender will commit the offense when her gain exceeds the expected (respectively perceived) punishment, i.e., for $g \geq g_T^j$, where the threshold gain g_T^j in general depends on both the regime $T = H, R$ and the offender types $j = s, n$ where s (n) indicates sophisticates and naïves, respectively. Thereby, the deterrence of sophisticates is independent of whether or not the effort is revealed. By contrast, the deterrence of naïves is determined by the true enforcement effort e under regime R and by the perceived effort ($\hat{e}$) under regime H. In summary, this leads to $g_H^s = g_R^s = g_R^n = p(e) \cdot f$ and $g_H^n = p(\hat{e}) \cdot f$, which is independent of e.

The enforcement authority chooses its policy (T, e) to maximize the overall expected surplus

$$W_T(e) := (1-a) \cdot \left[\int_{p(e) \cdot f}^{\infty} (\pi g - h) G'(g) dg \right] + a \cdot \left[\int_{g_T^n}^{\infty} (\pi g - h) G'(g) dg \right] - C(e), \qquad (1)$$

where the first (second) term gives the surplus generated by sophisticates (naïves) and where $\pi \in (0, 1]$ denotes the weight placed on the offenders' gains.[13] With respect to candidate maximizers of the surplus function (1), the above Inada conditions rule out corner solutions of e. Moreover, we also assume that the surplus function (1) is single-

[13]Most scholars would agree that benefits from severe crimes should not be considered as part of social welfare (see e.g., Stigler, 1970). However, things might be different for smaller offenses such as, for example, violations of environmental standards leading to a monetary gain in the form of a lower production cost. As a result, the gains are usually included (with weight 1) in the social surplus function (see e.g., Polinsky and Shavell, 2000, 2007). Our slightly more general formulation hence allows to capture different forms of offenses or different preferences of the social planner.

peaked under both regimes $T = H, R$ such that we get a unique interior optimum with respect to the enforcement effort.[14]

4 Optimal Enforcement Policy

We consider first the regime where the enforcement authority hides its effort ($T = H$), so that the two offender types face (different) threshold values, $g_T^s = p(e)f$ and $g_T^n = p(\hat{e})f$. The authority therefore chooses its enforcement effort e to maximize

$$W_H(e) := (1-a) \cdot \left[\int_{p(e)f}^{\infty} (\pi g - h) G'(g) dg \right] + a \cdot \left[\int_{p(\hat{e})f}^{\infty} (\pi g - h) G'(g) dg \right] - C(e). \quad (2)$$

We denote the (unique) maximizer of surplus function (2) by $e_H^*(a)$ and the resulting maximum surplus by $W_H^*(a, \hat{e}) := W_H(e_H^*(a); a, \hat{e})$.[15] The interior solution for $e_H^*(a)$ is implicitly given by the first-order condition

$$(1-a)\left[(h - \pi p(e)f) \cdot G'(p(e)f) \cdot p'(e)f\right] = C'(e), \quad (3)$$

i.e. when the marginal benefit of deterring sophisticates equals the marginal effort cost.

Consider next the regime where the enforcement authority reveals its effort ($T = R$), so that it is observed by both offender types. As a result, they both face the same threshold $g_R^s = g_R^n = p(e)f$. The optimal enforcement effort therefore maximizes

$$W_R(e) := \int_{p(e)f}^{\infty} (\pi g - h) G'(g) dg - C(e), \quad (4)$$

and we denote the (unique) maximizer by e_R^* and the resulting maximum surplus by $W_R^* := W_R(e_R^*)$. Since also naïve offenders learn the actual effort under this regime, both e_R^* and W_R^* are independent of $\hat{e}$ and a. The interior solution e_R^* solves the first-order condition:

$$\left[(h - \pi p(e)f) \cdot G'(p(e)f) \cdot p'(e)f\right] = C'(e). \quad (5)$$

While the exact characterization of $e_H^*(a)$ and e_R^* depend on the details of the cost function $C(e)$, the distribution of gains G and the other model parameters, some general

[14] Single-peakedness is for example ensured when the surplus function (1) is globally concave, i.e., when

$$(1-a)f\left[-\pi p'(e) \cdot G'(p(e)f) + (h - \pi p(e)f)\left[G''(p(e)f)\left(p'^2 f \cdot p'(e)f + G'(p(e)f)\, p''(e)\right]\right]\right] < C''(e)$$

holds for all $e > 0$. For example, this condition is satisfied as long as the distribution of gains G is not too convex or as long as the effort cost function $C(e)$ is sufficiently convex.

[15] Notice that the first integral in the surplus function (2) is independent of $\hat{e}$ (as $\hat{e}$ does not affect the behavior of sophisticates), while the second is independent of e (as the deterrence of naïves is solely determined by $\hat{e}$). As a result, $e_H^*(a)$ does not depend on $\hat{e}$.

results are nevertheless available. Define $e^{max} := p^{-1}(\frac{h}{\pi f})$ as the enforcement effort under which the indifferent (sophisticated) offender's (weighed) gain just equals the social harm.

Lemma 1. *(**No Over-Deterrence**) For both regimes H and R, the (weighed) gain of the indifferent sophisticated offender resulting under the optimal enforcement effort is below the social harm, i.e. $\pi p(e^*_H(a))f < h$ and $\pi p(e^*_R)f < h$ or, equivalently, $e^*_H(a), e^*_R < e^{max}$.*

Lemma 1 corresponds to a well-known result of the literature (see e.g. Polinsky and Shavell, 2007), which focuses on the case where all offenders are sophisticates. Intuitively, because enforcement is costly, some offenses with gains below social harm are not deterred in the social optimum. Thereby, the optimal enforcement level decreases in the weight π which the authority puts on the offenders' gains. To avoid uninteresting case distinctions, we assume that the naïves' perceived enforcement effort $\hat{e}$ satisfies the same property.

Assumption 1. *The perceived enforcement effort of naïve offenders $\hat{e}$ satisfies $\pi p(\hat{e})f < h$, which is equivalent to $\hat{e} < e^{max}$.*

The next result characterizes the properties of the optimal policy under regime H.

Proposition 1. *(**Optimal Policy under Regime** H)*

(i) The optimal (interior) effort level $e^*_H(a)$ is strictly decreasing in the share of naïves a and satisfies $e^*_H(0) = e^*_R$, $e^*_H(a) < e^*_R$ for all $a \in (0, 1]$, and $e^*_H(1) = 0$.

(ii) For any given share of naïves $a > 0$, the resulting maximum surplus $W^*_H(a, \hat{e})$ is strictly increasing in the naïves' perceived enforcement effort $\hat{e}$.

(iii) When the perceived effort of naïve offenders is sufficiently large (i.e., $\hat{e} > e^*_H(0)$), the social welfare under hiding $W^*_H(a, \hat{e})$ is strictly increasing in the share of naïves a for all $a \in [0, 1]$.

(iv) Otherwise (i.e., for $0 < \hat{e} < e^*_H(0)$), there exists a threshold for the share of naïves $\hat{a} \in (0, 1)$, implicitly defined by $e^*_H(\hat{a}) = \hat{e}$, such that $W^*_H(a, \hat{e})$ is strictly decreasing (increasing) in a for all $a < (>)\hat{a}$.

As for part (i), when there are no naïve agents ($a = 0$), the two surplus functions (2) and (4) coincide, so that $e^*_H(0) = e^*_R$ (and $W^*_H(0, \hat{e}) = W^*_R$) must hold. Moreover, $e^*_H(a)$ is decreasing in a as the authority's effort matters only for sophisticates. Thus, the optimal effort is always smaller when it is hidden rather than revealed (i.e., $e^*_H(a) < e^*_R$

for all $a \in (0,1]$). In the polar case where all offenders are naïve ($a = 1$), the optimal effort is zero as the deterrence for the whole population of offenders no longer depends on it, so that a positive effort level would not lead to more deterrence. Note also that under regime H, the fraction of sophisticates that is deterred from committing the offense decreases in the fraction of naïves: Sophisticates with gains $g > p(e_H^*)f$ benefit from the presence of naïves as they face a lower detection probability than they would if there were no naïves. Part (ii) expresses the fact that a higher enforcement effort as perceived by naïves ($\hat{e}$) increases their deterrence, while at the same time not affecting the behavior of sophisticates. As a result, the maximum attainable surplus under hiding increases (given Assumption 1).

As for part (iii), not even the maximum effort the authority would choose exceeds the perceived effort of naïves ($e_H^*(0) < \hat{e}$). Then, naïves will always overestimate the probability of apprehension, so that their deterrence is higher compared to sophisticates (i.e., $g_H^n = p(\hat{e})f > p(e_H^*(a))f = g_H^s$ for all $a \in [0,1]$). As a larger share of naïves hence leads to higher deterrence, the (maximum) social surplus is monotone increasing in a (again under Assumption 1).

As for the case of non-excessive levels of $\hat{e}$ considered in part (iv), it now depends on the share of naïves (a) whether the optimal enforcement effort is higher or lower than naïves' perception about it. In particular, $e_H^*(a) < \hat{e}$ and the argument from part (iii) only applies when a is sufficiently large such that it exceeds the threshold level $\hat{a}(\hat{e})$. By contrast, for a small (i.e., $a < \hat{a}(\hat{e})$), the optimal enforcement effort exceeds the perception of naïves about it ($e_H^*(a) > \hat{e}$), which leads them to underestimate the probability of apprehension. Moreover, also the deterrence is weaker for naïves compared to sophisticates ($g_H^n = p(\hat{e})f < p(e_H^*(a))f = g_H^s$), so that in this parameter region an increase in a leads to a net reduction of both deterrence and social welfare. Combining these two cases together give rise to a (maximized) social welfare function which is U-shaped in the share of naïves (a).

In summary, the crucial feature of regime H is that, in equilibrium, there is either a high enforcement effort leading naïves to underestimate the probability of apprehension (when the share of naïves a is low) or a low enforcement effort leading naïves to overestimate it (when their share is high). Assuming that offender naïveté of potential offenders is less prevalent for severe crimes, this is in line with empirical evidence surveys according to which individuals underestimate the probability of apprehension for homicide and overestimate it for burglary (see e.g., Kleck et al., 2005).

In a next step, we characterize the authority's optimal regime choice by comparing the resulting maximum surplus under the optimal enforcement levels $e^*_H(a)$ and e^*_R, respectively. Recall that under regime H, the optimal (costly) enforcement effort e^*_H is only effective for sophisticated offenders, while naïves are deterred by their perceived effort $\hat{e}$. By contrast, under regime R, only the actual effort e^*_R matters, while $\hat{e}$ is no longer relevant. This leads to the following result:

Proposition 2. ***(Optimal Regime Choice)***

(i) When the perceived effort of naïve offenders is sufficiently large (i.e., $\hat{e} > e^*_H(0)$), then it is always optimal to hide the effort, i.e., $W^*_H(a, \hat{e}) > W^*_R$ $\forall a \in [0, 1]$.

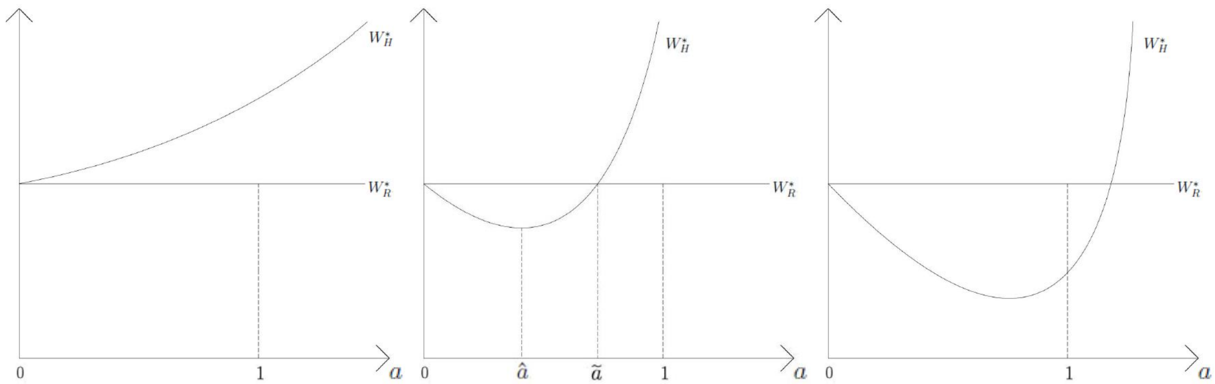
(ii) Otherwise (i.e, for $\hat{e} < e^*_H(0)$), either regime can be optimal. For $W^*_H(1, \hat{e}) > W^*_R$, there exists a threshold $\widetilde{a}(\hat{e}) \in (0, 1)$ implicitly defined by $W^*_H(\widetilde{a}(\hat{e}), \hat{e}) = W^*_R$ such that it is optimal to hide (reveal) the effort when the share of naïves is sufficiently large (small), i.e., $W^*_H(a, \hat{e}) > (<) W^*_R$ $\forall a > (<) \widetilde{a}(\hat{e})$.

(iii) If $\hat{e} < e^*_H(0)$ and $W^*_H(1, \hat{e}) < W^*_R$ hold, then revealing the effort is always optimal, i.e., $W^*_H(a, \hat{e}) < W^*_R$ $\forall a \in [0, 1]$.

The proposition is illustrated in Figure 1, where each of the three cases is represented by one panel. For $\hat{e} > e^*_H(0)$, even the maximum enforcement effort which the enforcement authority would choose under regime H is lower than the naïves' perception about it. As these high perceptions of naïves are a powerful deterrence device, regime H is then optimal (see panel (i) of Figure 1).

Conversely, for lower levels of $\hat{e}$ satisfying $\hat{e} < e^*_H(0)$, we know from Proposition 1 that the maximum social surplus under hiding ($W^*_H(a, \hat{e})$) is U-shaped in a. Since the social surplus when revealing the effort is independent of the share of naïves (a), revealing the effort is optimal as long as this share is sufficiently small. Whether or not the regime with hiding eventually becomes optimal for large levels of a depends on whether the value of a (apart from $a = 0$) where $W^*_H(a, \hat{e}) = W^*_R$ holds lies inside or outside the feasible range $a \in [0, 1]$ (see panels (i) and (iii) of Figure 1). A necessary and sufficient condition for the former case is $W^*_H(1, \hat{e}) > W^*_R$ as stated in the proposition.

Overall, the intuition for the optimal enforcement policy choice can be summarized as follows: When the share of naïves in the population of potential offenders is small, the optimal policy stipulates a relatively high effort, which would be underestimated by the naïves, and is therefore revealed to all offenders. By contrast, when the share of naïves in



(i) hiding is always optimal (ii) hiding is optimal iff $a > \widetilde{a}$ (iii) revealing is always optimal

Figure 1: Illustration of Proposition 2.

Notes: Welfare comparison when hiding (W_H^*) respectively revealing (W_R^*) the enforcement effort. The horizontal axis represents the share of naïves a. Each panel corresponds to one part of Proposition 2.

the population of potential offenders is sufficiently large, the optimal policy stipulates a relatively low effort, which is overestimated by the naïves, and is therefore kept hidden.

Our model leads to empirically testable predictions. Suppose, for instance, that there is a new app that provides its users with information about the current intensity of speeding controls in some metropolitan area. This technological innovation could be interpreted as a decrease of the share of naïve motorists in that area. The local enforcement authority is hence predicted to increase its enforcement effort, even if it remains hidden to the non-users of the app (cf. Proposition 1). Moreover, if the decrease of the share of naïves is sufficiently strong, the enforcement authority is predicted to increase the enforcement effort discontinuously and switch the regime by actively revealing it to all customers (cf. Proposition 2).

5 Extensions

5.1 Extension A: Revealed Enforcement Effort Reduces its Effectiveness

So far, revealing the enforcement effort only changed the perception (and thereby the deterrence level) of naïve offenders, but not the effectiveness of the actual effort in terms of the detection of offenses. However, it is also argued that revealing the effort might compromise police investigations as this allows offenders to adapt their behavior in order to avoid detection.

We now account for the possibility that the revelation of the effort reduces its effectiveness. In particular, when revealing its effort, the authority detects each offender only with probability $\widetilde{p}_R(e)$, where $\widetilde{p}_R(e) < p(e)\ \forall e > 0$. The detection function $\widetilde{p}_R(e)$ is assumed to satisfy the same properties as the function $p(e)$. The reduction in effectiveness affects both offender types so that, as in the basic model, the distinction between sophisticates and naïves vanishes when the enforcement effort is revealed.

The enforcement authority's maximization problem is then given by

$$\max_e \ W_R(e) := \int_{\widetilde{p}_R(e)f}^{\infty} (\pi g - h) G'(g) dg - C(e). \tag{6}$$

We denote the (unique) maximizer by $\widetilde{e}_R^*$ and the resulting maximum surplus by $\widetilde{W}_R^* := W_R(\widetilde{e}_R^*)$. As in the basic model, $\widetilde{e}_R^*$ and $\widetilde{W}_R^*$ are independent of a and $\hat{e}$. If interior, $\widetilde{e}_R^*$ satisfies the first order condition

$$-\left[(\pi\widetilde{p}_R(e) - h) \cdot G'(\widetilde{p}_R(e)f) \cdot \widetilde{p}_R'(e)f\right] = C'(\widetilde{p}_R(e)), \tag{7}$$

which leads to the following Lemma:

Lemma 2. *When revealing the effort reduces its effectiveness, both the optimal effort and the resulting maximum surplus are smaller compared to the basic model, i.e.* $\widetilde{e}_R^* < e_R^*$ *and* $\widetilde{W}_R^* < W_R^*$ *hold.*

From Lemma 2, it follows immediately that hiding the effort is strictly superior to revelation when the population of offenders consists of sophisticates only ($a = 0$). Proposition 3 characterizes the optimal regime choice in more detail:

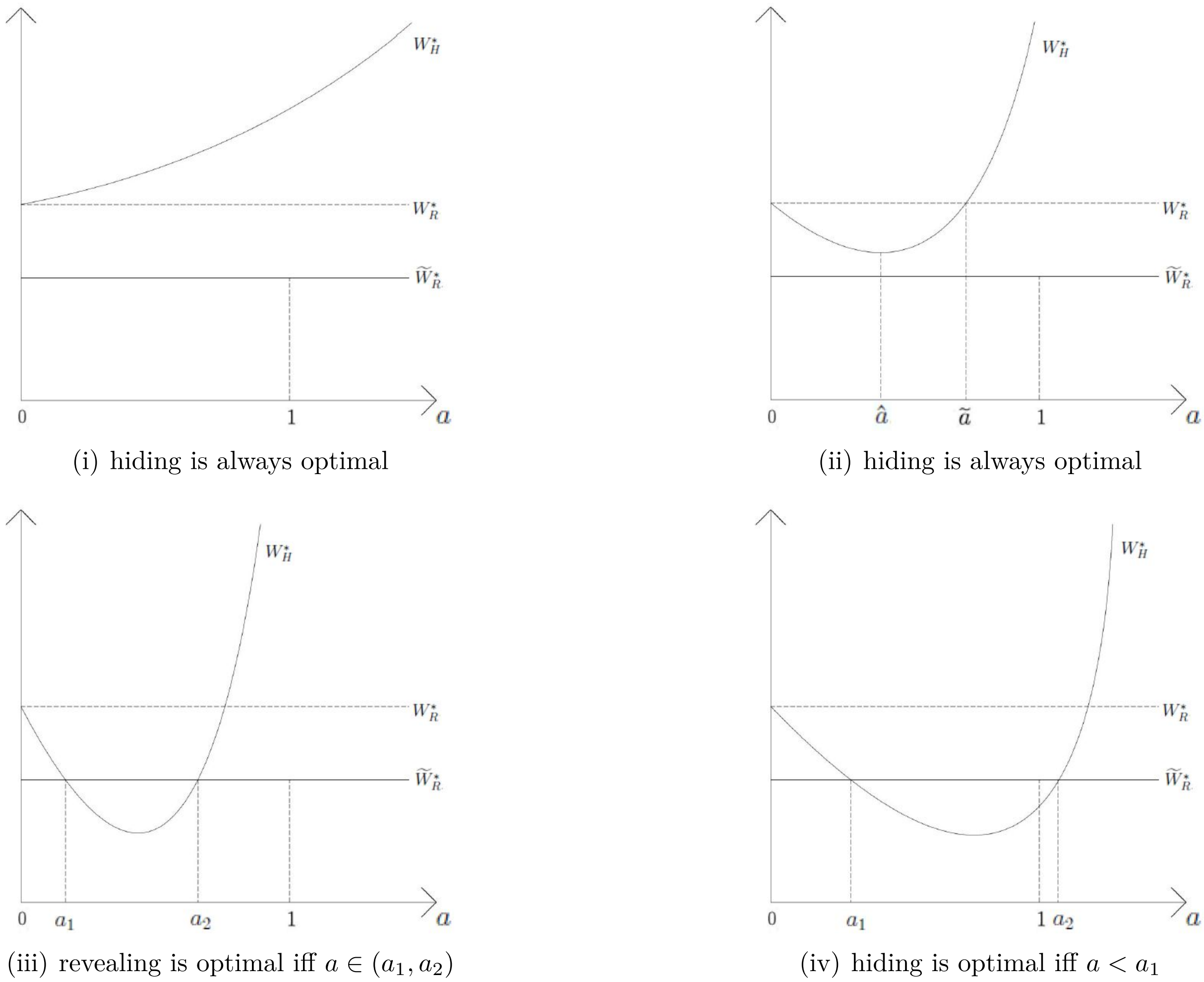
Proposition 3. ***(Regime Comparison with Reduced Effectiveness)***

(i) *For* $\hat{e} > e_H^*(0)$, *it is optimal to hide the effort, i.e.* $W_H^*(a, \hat{e}) > \widetilde{W}_R^*\ \forall a \in [0, 1]$.

(ii) *For* $\hat{e} < e_H^*(0)$, *it is optimal to hide the effort if the attainable surplus under revealed effort is sufficiently low, i.e. if* $\widetilde{W}_R^* < W_H^*(\hat{a}, \hat{e})$, *where* $\hat{a}$ *is the share of naïves at which* $e_H^*(\hat{a}) = \hat{e}$.

(iii) *For* $\hat{e} < e_H^*(0)$ *and* $\widetilde{W}_R^* > W_H^*(\hat{a}, \hat{e})$, *either regime can be optimal. In particular, for* $W_H^*(1, \hat{e}) > \widetilde{W}_R^*$, *there exist two thresholds* a_1 *and* a_2 *implicitly defined by* $W_H^*(a_1, \hat{e}) = W_H^*(a_2, \hat{e}) = \widetilde{W}_R^*$ *with* $0 < a_1 < a_2 < 1$ *such that it is optimal to reveal the effort if the share of naïves is neither too large nor to small, i.e.* $\widetilde{W}_R^* > W_H^*(a, \hat{e})$ $\forall a \in (a_1, a_2)$. *Otherwise, it is optimal to hide the effort.*

(iv) By contrast, if the two conditions from part (iii) hold, but if $W_H^*(1, \hat{e}) < \widetilde{W}_R^*$ (implying $a_2 > 1$), it is optimal to hide (reveal) the effort if the share of naïves is sufficiently small (large), i.e. $W_H^*(a, \hat{e}) > (<) \widetilde{W}_R^*$ $\forall a < (>) a_1$.



(i) hiding is always optimal

(ii) hiding is always optimal

(iii) revealing is optimal iff $a \in (a_1, a_2)$

(iv) hiding is optimal iff $a < a_1$

Figure 2: Illustration of Proposition 3.

Notes: Welfare comparison when either hiding the enforcement effort (W_H^*) or revealing it with reduced effectiveness ($\widetilde{W}_R^*$) (the case of revealing enforcement effort with unchanged effectiveness of the baseline model (W_R^*) is kept as a benchmark). The horizontal axis represents the share of naïves a. Each panel corresponds to one part of Proposition 3.

Proposition 3 is illustrated in Figure 2: As in the basic model (see Figure 1 above), hiding the effort is optimal for sufficiently large share of naïves, and/or when their perceived enforcement effort ($\hat{e}$) is sufficiently large. However, there are also qualitative changes: First, due to $\widetilde{W}_R^* < W_R^*$, the parameter range where hiding the effort is optimal increases. In particular, as shown in part (ii), even when W_H^* is U-shaped, hiding can still be optimal for all shares of naïves a when the negative impact of revelation of the effort on its effectiveness is sufficiently large.

As for parts (iii) and (iv), when hiding is not globally optimal, the regime comparison becomes non-monotonic in the share of naïves (a). In particular, there now also exists an interval of small values of $[0, a_1]$ where hiding is optimal. This qualitative difference to the baseline model (compare with Figure 1) is due to the fact that revealing the effort would indeed improve deterrence for the few naïves, but the deterrence for all sophisticates would decrease due to the lower probability of apprehension ($\widetilde{p}(e) < p(e)$). In the interval $[0, a_1]$ this second effect is larger so that hiding is optimal. Furthermore, as in the baseline model, for intermediate values of $a \in [a_1, a_2]$ revealing is optimal, while hiding becomes again optimal for a sufficiently large when $a_2 < 1$ holds.

5.2 Extension B: Endogenous Choice of Fine

We have so far treated the fine f as exogenously given. This is appropriate in settings where the enforcement authority chooses its enforcement effort e, while the fines have been chosen by other parties such as legislators. In other cases, it is the enforcement authority which simultaneously decides on both fine and effort. For such settings, the classic insight of Becker (1968) is that any level of deterrence $p(e)f > 0$ can also be reached with a slightly lower effort and a slightly higher fine. Moreover, such a change leads to higher welfare since increasing the fine is costless, while decreasing the effort saves enforcement costs. As a consequence, it is always optimal to set the largest possible fine.

In this section, we analyze a model extension in which the authority simultaneously decides on both the fine and its enforcement effort. We show that in our setting with sophisticated and naïve offenders, Becker's argument does not always apply, i.e., it might be optimal for the authority to set the fine strictly below its maximum level. (As mentioned in footnote 7 above, the literature has already identified several other scenarios in which the classic reasoning that "fines should be maximal" might not apply.) As for the regime comparison, we find that endogenous fines works in favor of regime H.

Consider an authority that chooses effort $e \geq 0$ and fine $f \in [0, F]$, where the maximal possible fine F might for example by given by law or by the wealth of offenders. The optimization problem of the enforcement authority from the baseline model (see the surplus function (1)) then needs to be adapted as follows:

$$\max_{T,e,f} \quad W_T(e) := (1-a) \cdot \left[\int_{p(e)\cdot f}^{\infty} (\pi g - h) G'(g) dg \right] + a \cdot \left[\int_{g_T^n}^{\infty} (\pi g - h) G'(g) dg \right] - C(e), \quad (8)$$

where the authority can now also affect the respective threshold for the marginal offenders (lower bounds of the integrals) through its choice of f (recall that $g_T^n \in \{p(e)f, p(\hat{e})f\}$). Denoting by e_T^* and f_T^* the respective optimal choices under regime $T = R, H$, we have the following result:

Proposition 4. ***(Endogenous Fine)*** *When the enforcement authority also chooses the fine $f \in [0, F]$ in addition to its enforcement effort e, then:*

(i) In regime R, the maximal fine is optimal, $f_R^* = F$. All results of the baseline model hold by substituting the exogenous fine $\bar{f}$ with the maximal fine F.[16]

(ii) In regime H, when the maximal fine is optimal, $f_H^* = F$, then all results of the baseline model hold by substituting the exogenous fine $\bar{f}$ with the maximal fine F.

(iii) In regime H, when the optimal fine f_R^* is interior (i.e., $f_R^* < F$), then the optimal enforcement effort is below the perceived enforcement effort, i.e., $e_H^* < \hat{e}$. The gain of the indifferent sophisticated offender is below social harm, while the gain of the indifferent naïve offender is above social harm, i.e., $\pi p(e_H^*) f^* < h < \pi p(\hat{e}) f^*$.

(iv) When regime H leads to higher welfare than regime R in the baseline model for a fixed fine $\bar{f}$, then this also holds when $\bar{f}$ is the maximal possible fine, i.e. when the fine f is chosen from the interval $[0, F = \bar{f}]$.

Parts (i) and (ii) of the proposition provide an additional justification for considering fixed fines in the baseline model: As the fine optimally chosen is just equal to the maximum amount (which is exogenously given), assuming an exogenous fine in the baseline model can be interpreted as a reduced form.

Part (iii) of the proposition, however, reveals a novel case where the optimal fine is below the maximal one. While increasing small fines is beneficial by deterring more offenders, naïves are over-deterred when the fine becomes too large: Their private benefit from the offense might exceed social harm, but they are nevertheless deterred as the fine has reached the point where $\pi p(\hat{e}) f \geq h$. Further increasing the fine then involves a trade-off between deterring more sophisticates, which is still desirable (as $\pi p(e^*) f^* < h$), and deterring more naïves. The optimal interior fine $f^* < F$ satisfies two first order conditions, one of which shows this novel trade-off:

$$(1-a)\left[(h - \pi p(e) f) \cdot G'(p(e) f) \cdot p(e)\right] = a\left[(\pi p(\hat{e}) f - h) \cdot G'(p(\hat{e}) f) \cdot p(\hat{e})\right], \qquad (9)$$

[16] To emphasize the difference to the baseline model where the fine was exogenous, we now use notation $\bar{f}$ when referring to an exogenous fine.

i.e. the marginal benefit of deterring sophisticated agents by increasing fine f (LHS) equals the marginal loss of deterring naïve agents with a high benefit from crime (RHS).[17]

The model with sophisticated and naïve agents thus reveals a new reason for why Becker's classic argument for maximal fines does not always apply. With too high a fine, one would deter inefficiently many offenders who are not aware that effort e^* is actually low. While over-deterrence seems unlikely for severe crimes, it may well be relevant in situations where violating a production standard leads to a large cost reduction, or when not committing an offense leads to high opportunity costs.

Finally, part (iv) of Proposition 4 provides a regime comparison, which favors hiding the enforcement effort. This result follows from the fact that endogenizing the fine gives the enforcer more flexibility. Under regime R, however, the fine is always chosen maximally, so that the authority just replicates the welfare from the baseline model by choosing $f^* = F = \bar{f}$ and effort optimal as before. Under regime H, this may also be the case, but we have just seen that it may also be optimal to implement a lower fine. Thus, an endogenous fine works in favor of hiding the enforcement effort.

5.3 Extension C: Heterogeneity of Perceptions

In the baseline model, all naïve individuals share the same perception $\hat{e}$ when enforcement effort is hidden. In the following, we first show that the results of the baseline model carry over to heterogeneous perceptions. Then, we demonstrate that heterogeneity sets additional incentives to reveal the effort.

Let there be L groups of naïves with perceptions $\hat{e}_1, ..., \hat{e}_L$ and group sizes $a_1, ..., a_L$. All other model features are as in the baseline model (see Section 3). In particular, gains from crime are distributed according to a cdf G for the sophisticated agents as well as for all groups of naïve individuals. Furthermore, all naïve individuals learn the actual effort e_R in case of revelation.[18] Finally, we extend Assumption 1 to all perceptions $\hat{e}_l$ i.e., $0 < \hat{e}_l < e^{max}$ holds for any group l. As the marginal offender's gain from crime is thus below social harm, deterrence is socially desirable.

Welfare under revealing $W_R(e)$ is unaffected from heterogeneity of perceptions, so that the maximum welfare in this case is still $W_R(e_R^*) = W_R^*$. Welfare under regime H is now

[17] The other first order condition reflects the common trade-off between costs and benefits of deterrence: $(1-a)\left[(h - \pi p(e)f) \cdot G'(p(e)f) \cdot p'(e)f\right] = C'(e)$, i.e. the marginal benefit of deterring sophisticated agents by increasing effort e equals the marginal effort cost.

[18] We assume that revelation of effort is public in the sense that it is impossible to reveal it to just some groups of naïves. While partial revelation could never be optimal in the baseline model, there could now be an incentive to reveal the effort only to naïves who underestimate the effort.

given by:

$$W_H(e) := (1 - \sum_{l=1}^{L} a_l) \cdot \left[\int_{p(e)f}^{\infty} (\pi g - h) G'(g) dg\right] + \sum_{l=1}^{L} a_l \cdot \left[\int_{p(\hat{e}_l)f}^{\infty} (\pi g - h) G'(g) dg\right] - C(e), \tag{10}$$

which is a straightforward generalization of the surplus function (2).

To analyze the impact of different degrees of heterogeneity, the following definition is useful. Two models are referred to as *welfare equivalent* when they lead to the same welfare when the same policies (T, e) are chosen in the two models.

Proposition 5. ***(Reduction of General Model to Baseline Model)*** *For every model with heterogeneous perceptions $(\hat{e}_1, ..., \hat{e}_L)$ of the fractions $(a_1, ..., a_L)$ of naïve agents, there is a unique model with a homogeneous perception $\tilde{e}$ of the fraction of naïve agents $a := \sum_{l=1}^{L} a_l$ that is welfare equivalent to it.*

Intuitively, any model with heterogeneous perceptions can be mirrored by the unique (baseline) model with homogeneous perceptions where $\hat{e} = \tilde{e}$ and $a = \sum_{l=1}^{L} a_l$. This implies that all results for this baseline model (Section 4) carry over. In particular, Proposition 1 characterizes the optimal policy under hiding, and Proposition 2 characterizes the optimal regime choice. Moreover, we can generate some additional comparative statics insights by studying how the parameters of the extended model, with $(\hat{e}_1, ..., \hat{e}_L)$ and $(a_1, ..., a_L)$, affect the parameters of the welfare equivalent model, with a and $\tilde{e}$. This leads to the following observations. First, as the optimal effort e_H^* depends only on the percentage of sophisticated agents, it is strictly decreasing in a_l for each group l. Second, welfare W_H^* is strictly decreasing (increasing) in a_l when the group's perception satisfies $\hat{e}_l < \hat{e}_H^*$ $(\hat{e}_l > \hat{e}_H^*)$. Finally, welfare W_H^* is strictly increasing in the perception $\hat{e}_l (< e^{max})$ of any group l (with $a_l > 0$).

The reduction result expressed in Proposition 5 shows that the insights from of our baseline model are robust. However, we have not yet specified how the degree of heterogeneity affects welfare, i.e. which (welfare-equivalent) homogeneous perception $\tilde{e}$ corresponds to the given heterogeneous perception levels $\hat{e}_1, ..., \hat{e}_L$. It turns out that the perception level $\tilde{e}$ is not simply the mean of the heterogeneous perceptions $\sum_{l=1}^{L} \hat{e}_l$, but lower than that, because increasing the dispersion of perceptions has a deteriorating effect on welfare. To illustrate the intuition behind this insight, we analyze a simple set-up, in which the dispersion of the perceptions is varied while the mean level of perceptions is kept constant.

We study $L = 2$ groups of equal sizes ($a_1 = a_2 > 0$) and with perceptions $\hat{e}_1 = \hat{e} - \sigma$ and $\hat{e}_2 = \hat{e} + \sigma$. Let σ be small enough such that $0 < \hat{e}_1 < \hat{e}_2 < e^{max}$. The construction is such that the mean perception level is $\hat{e}$ and the distance of each group to the mean is σ. While we already know that a higher mean of perceived effort $\hat{e}$ is welfare enhancing (since this holds true for the perception $\hat{e}_l$ for any group l), we now turn to the impact of the distance σ. We show that increasing the distance σ reduces welfare with hidden effort W_H^* under very mild assumptions.

Proposition 6. ***(Heterogeneity of Perceptions Reduces Welfare)*** *Let there be $L = 2$ groups of naïve agents of equal sizes ($a_1 = a_2$) and with perceptions $\hat{e}_1 = \hat{e} - \sigma$ and $\hat{e}_2 = \hat{e}+\sigma$. Suppose that the gain distribution G and the detection function $p(e)$ are not too convex (they may well be linear, concave or slightly convex), i.e. $\frac{G'(p(\hat{e}+\sigma)f)}{G'(p(\hat{e}-\sigma)f)} \cdot \frac{\partial p(\hat{e}+\sigma)/\partial\sigma}{-\partial p(\hat{e}-\sigma)/\partial\sigma} < \frac{h-\pi p(\hat{e}-\sigma)f}{h-\pi p(\hat{e}+\sigma)f}(> 1)$. Then, welfare W_H^* under regime H is strictly decreasing in the distance σ of the perceptions to the mean.*

Heterogeneity among naïve offenders reduces welfare under hiding because it induces different thresholds for the indifferent offenders in each group of naïves ($g_H^{n_2} = p(\hat{e}_2)f > p(\hat{e}_1)f = g_H^{n_1}$). As a result, some naïves with $\hat{e}_2$ and large gains are deterred, while others with $\hat{e}_1$ and small gains are not. As in the baseline model, for any given number of offenses the total surplus is highest when the offenses are committed by the offenders with the largest gains. Since this condition is violated for any $\sigma > 0$, there is some inefficiency. Moreover, as σ increases, so does the wedge between the two threshold values and the resulting inefficiency, so that overall welfare decreases. As for the regime comparison, as welfare with revealed effort is independent of the perceptions and their distribution, revealing is more likely to be optimal when the heterogeneity of perceptions is large. Our finding also extends the result by Garoupa (1999) on the benefits of revelation to our setting with naïves and sophisticates. While revelation is always optimal in Garoupa (1999) if it incurs no cost, hiding may still be optimal in our setting if perceptions of naïves tend to be large and are not too dispersed.

Note that different perceptions on enforcement effort would not lead to different welfare if gains from offenses were ignored (i.e. if $\pi = 0$) and the density of benefits was constant, as it would then not make a difference from a social welfare perspective who commits an offense. For several applications such as environmental or product liability, however, it can be argued that private gains do matter as they often come in the form of lower avoidance costs. If firms have different perceptions of the authority's enforcement effort, this is likely to lead to inefficiencies in the form of inefficient care levels. Our results suggest that the larger heterogeneity in perceptions, the more likely is it that revealing the true effort to firms is optimal.

6 Conclusion

The economic literature on law enforcement assumes that potential offenders are either fully informed about the agency's enforcement effort (and, hence, the probability of apprehension) or form unbiased beliefs in case of uncertainty. At the same time, criminologists emphasize that the perceived probability of apprehension differs considerably among individuals and is often not systematically related to the true probability. We propose a model that combines both perspectives by distinguishing between sophisticated and naïve offenders, and characterize the optimal enforcement policy. Thereby, in addition to determining its enforcement effort, the enforcement authority can also decide whether to hide or reveal it to the offenders.

We show that the welfare-maximizing authority chooses either a policy (R, e_R^*) in which the enforcement effort (e_R^*) is relatively high and is revealed to the offenders; or a policy (H, e_H^*) in which the enforcement (e_H^*) is relatively low and remains hidden. The reason for the low effort under regime H is that it only affects the deterrence of a fraction of the agents, the sophisticates, whereas under regime R it is effective for all agents. The advantage of hiding is that enforcement costs can be saved due to the low effort. However, it also has two disadvantages compared to revealed effort: First, low effort leads to low deterrence of sophisticates. Second, hiding induces different gain thresholds for the indifferent sophisticated and naïve offender, respectively, and hence leads to a distortion in the sense that a given number of acts is not committed by the offenders with the highest gains. The regime comparison is then driven by the relative importance of these different effects. In particular, hiding becomes more attractive when the share of naïves is high.

In extensions, we consider several additional factors that affect the regime comparison just described. First, revealing the effort may reduce its effectiveness as offenders learn how to avoid detection. This makes hiding favorable, not only when the share of naïves is high, but also when it is low. Second, when the fine also becomes part of the enforcement policy, the authority may prefer to set a fine below the maximal level to mitigate the effect that inefficiently many naïve offenders are deterred. Third, the naïve offenders may differ with respect to their perceptions about the enforcement effort, which reinforces the issues that it might no longer the subjects with the highest benefits who commit the act. This works in favor of revelation. Overall, our results show that, when deciding on their effort and communication strategy, authorities should take into account the number of offenders with mis-perceptions and their degree of mis-perception. Those parameters, might well differ across different types of offenses (Kleck et al., 2005). We view our paper as contributing to the overall agenda of integrating the perspectives from law & economics and criminology (see e.g., Chalfin and McCrary, 2017) in the academic debate on law enforcement and deterrence.

Our framework could be extended in several directions. First, we assume that the authority maximizes social welfare, which neglects potential principal-agent issues between society and the law enforcement authority. In particular in the context of private law enforcement, the authority may have an incentive to signal its competency by focusing on the number of detected offenders instead of overall welfare.[19] Second, it would be interesting to relax the assumption that the perceptions of naïve offenders are exogenous and static. Instead, they could adapt their beliefs upon receiving noisy signals on the actual enforcement level, for example based on experiences of their own or within their social network as in Sah (1991). Enforcement authorities would then face a dynamic optimization problem that has not yet been solved. Third, the naïveté of offenders may only refer to some but not to all enforcement technologies. As an example, consider Ben Gurion airport where all arriving vehicles must first pass through a preliminary security checkpoint where armed guards search the vehicle and exchange a few words with the driver and occupants to gauge their mood and intentions.[20] As this effort is observable to everyone, our distinction between naïve and sophisticated offenders is likely to be of minor importance. In addition, however, plain clothes officers patrol the area outside the terminal building, assisted by hidden surveillance cameras which operate around the clock, and not all offenders might be aware of this effort. The general question is then how the authority should divide its effort between the directly observable and the not directly observable technology, and how the incentive to reveal information on the latter depends on the relative efficacy and costs of the two technologies. Finally, our model could be extended to include precaution on the side of the victims. Potential victims might invest into safety technologies like alarm equipment, but they may also have inaccurate perceptions on the benefits of those technologies.

[19] As argued by Buechel and Muehlheusser (2016), the number of detections might not be too informative about the underlying enforcement effort.

[20] See e.g., `https://edition.cnn.com/travel/article/ben-gurion-worlds-safest-airport-tel-aviv`.

Appendix

A Proofs

A.1 Proof of Lemma 1

As for regime H, suppose first that $a = 1$. Then the maximizer of surplus function (2) is $e_H^*(1) = 0$. Hence, $\pi p(e_H^*(a))f = \pi p(0)f = 0 < h$. Now, let $a < 1$. Then surplus function (2) is increasing at $e = 0$ because of the Inada condition $C'(0) = 0$. Hence, the maximizer satisfies $e_H^*(a) > 0$, i.e. the optimal effort is interior, and satisfies the first order condition Eq. (3). Our assumptions on the cost function $C(e)$ ensure that the RHS of Eq. (3) is always strictly positive for all $e > 0$. Hence, the condition can only be satisfied when the LHS is also strictly positive. Since $G'(\cdot) > 0$, and $p'(e) > 0$ and $f > 0$, it follows that also $h - \pi p(e)f > 0$ must hold at $e = e_H^*(a)$ for the LHS to be strictly positive. This, however, is just equivalent to the statement in the Lemma. The proof for regime R is completely analogous to the case $a < 1$ in regime H and hence omitted. □

A.2 Proof of Proposition 1

(i) First suppose $a < 1$. Then surplus function (2) is increasing at $e = 0$ because of the Inada condition $C'(0) = 0$. Hence, the maximizer satisfies $e_H^*(a) > 0$, i.e. the optimal effort is interior, and satisfies the first order condition Eq. (3). That the optimal effort $e_H^*(a)$ is strictly decreasing in a can be established as follows: From Eq. (3), applying the implicit function theorem, one gets

$$\frac{\partial e_H^*(a)}{\partial a} = -\frac{(-1)\left[(h - \pi p(e)f) \cdot G'(p(e)f) \cdot p'(e)f\right]}{W_H''(e)} < 0.$$

To verify the sign of this expression, note first that the denominator is just the second derivative of the surplus function (2). At the optimal effort $e^*(a)$, this is strictly negative since this is the condition for a maximum. Furthermore, the numerator is strictly negative since $G'(\cdot) > 0$, and $p'(e) > 0$ and $f > 0$ and by Lemma 1. Moreover, for $a = 0$, the surplus functions (2) and (4) coincide and so must the optimal enforcement levels. The property $e_H^*(a) < e_R^*$ for $a > 0$ then follows directly from the above arguments. Finally, for $a = 1$, the claim $e_H^*(1) = 0$ can be established by contradiction: Suppose, there are only naive offenders in the population ($a = 1$) and some $e > 0$ were optimal. Then social welfare would be strictly higher when e is reduced, since it would lead to lower cost (since $C(e)$ is strictly increasing), but to no loss in deterrence. The reason is that as under regime H, the deterrence of naïves only works through $\hat{e}$, while the actual enforcement effort e has no impact.

(ii) Using the envelope theorem and taking the derivative of $W_H^*(a, \hat{e})$ w.r.t. $\hat{e}$ yields

$$\frac{\partial W_H^*}{\partial \hat{e}} = -a(\pi p(\hat{e})f - h) \cdot G'(p(\hat{e})f) \cdot p'(\hat{e})f$$

which is strictly positive under Assumption 1.

(iii) and (iv): Using the envelope theorem and taking the derivative of $W_H^*(a, \hat{e})$ w.r.t. a yields

$$\frac{\partial W_H^*}{\partial a} = -\int_{p(e_H^*(a))f}^{\infty} (\pi g - h)G'(g)dg + \int_{p(\hat{e})f}^{\infty} (\pi g - h)G'(g)dg,$$

the sign of which is solely determined by comparing the two respective lower bounds of the integrals. By Lemma 1, we have $\pi p(e_H^*(a))f - h < 0$ and by Assumption 1 we have $\pi p(\hat{e})f - h < 0$ such that the first integral is bigger (in absolute terms) than the second one if and only if $p(e_H^*(a))f < p(\hat{e})f$. Hence, $W_H^*(a, \hat{e})$ is increasing in a if and only if $e_H^*(a) < \hat{e}$, and they are identical for $e_H^*(a) = \hat{e}$. Part (iii) supposes that $\hat{e} > e_H^*(0)$. Proposition 1 above shows that $e_H^*(a)$ is decreasing. Hence, we have $\hat{e} > e_H^*(a)$ for all $a \in [0, 1]$. Thus, $W_H^*(a, \hat{e})$ is strictly monotone increasing in a. Part (iv) supposes that $0 < \hat{e} < e_H^*(0)$. Proposition 1 above shows that since $e_H^*(a)$ is decreasing with $e_H^*(1) = 0$. Together, we have $e_H^*(0) > \hat{e} > e_H^*(1)$, and there must be a threshold $\hat{a}(\hat{e})$ such that $e_H^*(\hat{a}) = \hat{e}$. Hence, $W_H^*(a, \hat{e})$ is strictly decreasing in a for $a < \hat{a}(\hat{e})$ and strictly increasing when the inequality is reversed. $\square$

A.3 Proof of Proposition 2

Part (i): Recall first that the two regimes coincide for $a = 0$, i.e., when there are no naïve offenders ($W_H^*(0, \hat{e}) = W_R^*$). As shown in Proposition 1, when $\hat{e} > e_H^*(0)$, $W_H^*(a, \hat{e})$ is strictly increasing in a for all $a \in [0, 1]$ and hence it is optimal for the enforcement authority to hide its enforcement effort.

Parts (ii) and (iii): When $0 < \hat{e} < e_H^*(0)$, then as shown in Proposition 1 above, $W_H^*(a, \hat{e})$ is U-shaped and strictly decreasing in the interval $[0, \hat{a})$ and increasing for $a > \hat{a}$. Hence, there must exist a (second) point of intersection between $W_H^*(a, \hat{e})$ and W_R^* at some point $\widetilde{a}(\hat{e}) > 0$. The condition $W_H^*(1, \hat{e}) > W_R^*$ is a necessary and sufficient condition for $\widetilde{a}(\hat{e})$ to lie in the relevant range $(0, 1)$, which is also illustrated in Figure 1. When it is satisfied, as assumed in part (ii), then $W_H^*(a, \hat{e}) < (>) W_R^*$ for all $a < (>) \widetilde{a}(\hat{e})$. When it is not satisfied, as assumed in part (iii), then $W_H^*(a, \hat{e}) < W_R^*$ for all $a \in (0, 1]$. For the special case $\hat{e} = 0$, $W_H^*(a, \hat{e})$ is strictly decreasing in $[0, 1)$, starting at $W_H^*(0, \hat{e}) = W_R^*$. Hence, this case is treated in part (iii). □

A.4 Proof of Lemma 2

First, due to the Inada condition $C'(0) = 0$, optimal effort is strictly positive for both surplus functions (4) and (6). Comparing the first order conditions Eq. (5) and Eq. (7) reveals that, since $\widetilde{p}_R(e) < p(e)$ for all $e > 0$, the marginal benefit is pointwise smaller in Eq. (7) compared to Eq. (5). Hence, the point of intersection with the marginal cost curve $C'(e)$ must occur at a smaller value of e.

Second, we now show that $W_R(e = \widetilde{e}_R^*) > \widetilde{W}_R^*$ holds, and hence a fortiori, $W_R^* > \widetilde{W}_R^*$ must hold. Note that we have

$$\widetilde{W}_R^* = \int_{\widetilde{p}_R(\widetilde{e}_R^*)f}^{\infty} (\pi g - h) G'(g) dg - C(\widetilde{e}_R^*). \tag{A.1}$$

and evaluating $W(e)$ from the surplus function (4) at $e = \widetilde{e}_R^*$ yields

$$W_R(\widetilde{e}_R^*) = \int_{p(\widetilde{e}_R^*)f}^{\infty} (\pi g - h) G'(g) dg - C(\widetilde{e}_R^*). \tag{A.2}$$

Taking the difference $W_R(\widetilde{e}_R^*) - \widetilde{W}_R^*$ yields

$$\int_{p(\widetilde{e}_R^*)f}^{\infty} (\pi g - h) G'(g) dg - \int_{\widetilde{p}_R(\widetilde{e}_R^*)f}^{\infty} (\pi g - h) G'(g) dg. \tag{A.3}$$

Note that since $\widetilde{p}_R(e) < p(e)$ for all $e > 0$, the lower bound is strictly larger in the first integral. Hence, we can rewrite the difference as

$$\int_{p(\widetilde{e}_R^*)f}^{\infty} (\pi g - h) G'(g) dg - \left(\int_{p(\widetilde{e}_R^*)f}^{\infty} (\pi g - h) G'(g) dg + \int_{\widetilde{p}_R(\widetilde{e}_R^*)f}^{p(\widetilde{e}_R^*)f} (\pi g - h) G'(g) dg \right). \tag{A.4}$$

Since the first two terms cancel, this is equal to

$$-\int_{\widetilde{p}_R(\widetilde{e}_R^*)f}^{p(\widetilde{e}_R^*)f} (\pi g - h) G'(g) dg. \tag{A.5}$$

A sufficient condition for this expression to be strictly positive is that the value of the integrand at the upper bound $p(\widetilde{e}_R^*)f$ is negative, i.e. if $\pi p(\widetilde{e}_R^*)f - h < 0$. Note that we have established in Lemma 1 above that $\pi p(e_R^*)f - h < 0$ holds. Since we have shown at the beginning of this proof that $\widetilde{e}_R^* < e_R^*$, a fortiori $\pi p(\widetilde{e}_R^*)f - h < 0$ must hold as $p(\cdot)$ is an increasing function. This in turn implies that expression (A.5) is indeed strictly positive.

As a final step, since we have shown that $W_R(\widetilde{e}_R^*) > \widetilde{W}_R^*$ holds, this must a fortiori be true for the maximum surplus under revealed enforcement effort in the basic model (W_R^*), i.e. we have $W_R^* \geq W_R(\widetilde{e}_R^*) > \widetilde{W}_R^*$. □

A.5 Proof of Proposition 3

Part (i): From Proposition 1, when $\hat{e} > e_H^*(0)$, $W_H^*(a, \hat{e})$ is strictly increasing in a for all $a \in [0, 1]$. Moreover, as shown in Lemma 2, $W_R^* > \widetilde{W}_R^*$ holds, so that we have $W_H^*(a, \hat{e}) \geq W_R^* > \widetilde{W}_R^*$ for all $a \in [0, 1]$ and hence it is always optimal for the enforcement authority to hide its enforcement effort.

Part (ii): When $0 < \hat{e} < e_H^*(0)$, then from Proposition 1, $W_H^*(a, \hat{e})$ is U-shaped in a, and it takes its minimum value at $a = \hat{a}$. When this minimum value still exceeds $\widetilde{W}_R^*$ (i.e. when $W_H^*(\hat{a}, \hat{e}) > \widetilde{W}_R^*$), then hiding the enforcement effort is again globally optimal. For the special case $\hat{e} = 0$, $W_H^*(a, \hat{e})$ is decreasing in a and $e_H^*(1) = \hat{e} = 0$, i.e. $\hat{a} = 1$. Thus, the statement also holds.

Part (iii): When $0 < \hat{e} < e_H^*(0)$ (so that $W_H^*(a, \hat{e})$ is U-shaped in a, but $W_H^*(\hat{a}, \hat{e}) < \widetilde{W}_R^*$), then there must exist a threshold $a_1 > 0$ such that $W_H^*(a, \hat{e}) > \widetilde{W}_R^*$ for all $a \in [0, a_1)$ (recall that $W_H^*(0, \hat{e}) = W_R^* > \widetilde{W}_R^*$). Moreover, if also the condition $W_H^*(1, \hat{e}) > \widetilde{W}_R^*$ is satisfied, then there must exist a second threshold a_2 with $a_1 < a_2 < 1$ such that hiding the effort is also optimal for all $a \in (a_2, 1]$, and revealing it is optimal in the intermediate range (a_1, a_2).[21]

Part (iv): This part refers to the setting of part (iii), but where the condition $W_H^*(1, \hat{e}) > \widetilde{W}_R^*$ does not hold, i.e. we have $W_H^*(1, \hat{e}) < \widetilde{W}_R^*$. Then a range where hiding the enforcement effort is again optimal for sufficiently large a does not exist (i.e. $a_2 > 1$), so that hiding is optimal (only) for $a \in [0, a_1)$ and revealing is optimal for $a > a_1$. This also holds for the special case $\hat{e} = 0$ (given that $W_H^*(1, \hat{e}) < \widetilde{W}_R^*$). $\square$

A.6 Proof of Proposition 4

Part (i): In regime R, the authority chooses its enforcement effort e and the fine f to maximize welfare.

$$\max_{e \geq 0, f \in [0, F]} W_R(e, f) = \int_{p(e)f}^{\infty} (\pi g - h) G'(g) dg - C(e). \tag{A.6}$$

We first observe that due to the Inada condition $C'(0) = 0$, the optimal effort is interior, i.e. $e_R^* > 0$. We next show that there is no over-enforcement, i.e. $\pi p(e_R^*) f^* < h$. Suppose to the contrary that $\pi p(e_R^*) f^* \geq h$. Then a slight reduction of e_R^*, while keeping f^* constant, would weakly increase social benefits and strictly decrease the costs. Suppose

[21] This cannot occur in the special case $\hat{e} = 0$ since for $\hat{a} = 1$, we cannot have $W_H^*(\hat{a}, \hat{e}) < \widetilde{W}_R^*$ and $W_H^*(1, \hat{e}) > \widetilde{W}_R^*$ at the same time.

now that the optimal fine f^* was not maximal, i.e. $f^* < F$. By continuity, the induced level of deterrence, $p(e_R^*)f^*$, can also be reached by a lower effort $e' < e_R^*$ and a higher fine $f' > f^*$. This increases welfare since the increase in fine is costless, while the decrease in effort saves costs. Hence, the optimal fine must be maximal: $f^* = F$. Consequently, the optimal effort equals the optimal effort of the baseline model when we set fine $\bar{f}$ (of the baseline model) to F.

Part (ii): In regime H, the authority chooses its enforcement effort e and the fine f to maximize welfare.

$$\max_{e\geq 0, f\in[0,F]} W_H(e,f) = (1-a)\cdot\left[\int_{p(e)f}^{\infty}(\pi g - h)G'(g)dg\right] + a\cdot\left[\int_{p(\hat{e})f}^{\infty}(\pi g - h)G'(g)dg\right] - C(e). \tag{A.7}$$

Observe first that $f = 0$ is never optimal. (Indeed, for every effort e and perceived effort $\hat{e}$, there is a (small) fine $f^\epsilon > 0$ with $\pi p(e)f^\epsilon < h$ and $\pi p(\hat{e})f^\epsilon < h$ such that $W_H(e,0) < W_H(e,f^\epsilon)$). Hence the optimal fine f^* is either interior or maximal.

By assumption of part (ii) of Proposition 4, the optimal fine is maximal: $f^* = F$. In this case, the solution for e_H^* is implicitly given by the first-order condition

$$(1-a)\left[(h - \pi p(e)F)\cdot G'(p(e)F)\cdot p'(e)F\right] = C'(e).$$

Observe that this condition coincides with Eq. (3) that determines the optimal effort in regime H of the baseline model, when we set $\bar{f} \equiv F$.

Part (iii): By assumption of part (iii) of Proposition 4, the optimal fine is interior, i.e. $f^* < F$. In this case, the maximization problem (A.7) has a solution (e^*, f^*) that satisfies the following two first order conditions:

$$(1-a)\left[(h - \pi p(e)f)\cdot G'(p(e)f)\cdot p'(e)f\right] = C'(e), \tag{A.8}$$

$$(1-a)\left[(h - \pi p(e)f)\cdot G'(p(e)f)\cdot p(e)\right] = a\left[(\pi p(\hat{e})f - h)\cdot G'(p(\hat{e})f)\cdot p(\hat{e})\right]. \tag{A.9}$$

Both equations (A.8) and (A.9) follow from Leibniz's rule. The RHS of Eq. (A.8), $C'(e)$, is positive for $e = e_H^* > 0$. Hence its LHS is also positive. The LHS of Eq. (A.8) is only positive if $\pi p(e)f < h$. Now, observe that $\pi p(e)f < h$ implies that the LHS of Eq. (A.9) is also positive. In turn, the RHS of Eq. (A.9) must also be positive, which implies that $(\pi p(\hat{e})f - h) > 0$. Hence, for the optimal effort e^* and the optimal fine f^*, the two first order conditions imply $\pi p(e^*)f^* < h < \pi p(\hat{e})f^*$. And finally, $e^* < \hat{e}$.

Part (iv): For the baseline model, let us denote by e_R^b and e_H^b the respective optimal efforts in regime R and in regime H. By assumption, $W_R(e_R^b, \bar{f}) < W_H(e_H^b, \bar{f})$ for the fixed fine $\bar{f}$. By part (i) of this proposition, the optimal fine in regime R is maximal, i.e. equal to F. Together with $F = \bar{f}$, this yields $\max_{e>0, f\in(0,F=\bar{f}]} W_R(e, f) = W_R(e^b_R, \bar{f})$. Hence,

$$\max_{e>0, f\in(0,\bar{f}]} W_R(e, f) = W_R(e^b_R, \bar{f}) < W_H(e^b_H, \bar{f}) \leq \max_{e>0, f\in(0,\bar{f}]} W_H(e, f). \quad \square$$

A.7 Proof of Proposition 5

Welfare $W_R(e)$ under regime R is unaffected from heterogeneity of perceptions and hence neither is optimal welfare $W_R(e^*_R) = W^*_R$ in this case. Welfare under regime H is given by Eq. (2) in the baseline model and by Eq. (10) in the heterogeneity extension. Observe that the optimal effort in both Eq. (2) and Eq. (10) then only depends on the first and on the last term, which coincide in both equations for $1 - a = 1 - \sum_{l=1}^{L} a_l$. Thus, optimal effort e^*_H is the same in both scenarios. For $a = \sum_{l=1}^{L} a_l$, the difference between the two becomes

$$(10) - (2) = \sum_{l=1}^{L} a_l \cdot \left[\int_{p(\hat{e}_l)f}^{\infty} (\pi g - h)G'(g)dg\right] - a \cdot \left[\int_{p(\hat{e})f}^{\infty} (\pi g - h)G'(g)dg\right].$$

Observe that this difference is independent of the actual effort e. By assumption on the exogenous perceptions $\hat{e}$ and $\hat{e}_l$, both expressions in brackets are negative (cf. Assumption 1). For $\hat{e} \equiv 0 (< \min\{\hat{e}_1, ..., \hat{e}_L\})$, the left term is larger in absolute terms than the right one such that the difference is positive. For $\hat{e} \equiv e^{max} (> \max\{\hat{e}_1, ..., \hat{e}_L\})$, the difference is negative. Since the difference is a continuously decreasing function in $\hat{e}$, there must exist a unique level $\hat{e} \equiv \tilde{e}$ that satisfies that the difference is just zero.

We have thus constructed a model with homogeneous perceptions $\tilde{e}$ that is welfare equivalent to the given model with heterogeneous perceptions. $\square$

A.8 Proof of Proposition 6

Welfare under regime H is given by the surplus function (10), which for two groups becomes

$$W_H(e) = (1 - a_1 - a_2) \cdot \left[\int_{p(e)f}^{\infty} (\pi g - h)G'(g)dg\right] + \sum_{l=1}^{2} a_l \cdot \left[\int_{p(\hat{e}_l)f}^{\infty} (\pi g - h)G'(g)dg\right] - C(e).$$

Setting $\hat{e}_1 = \hat{e} - \sigma$ and $\hat{e}_2 = \hat{e} + \sigma$ and applying Leibniz's rule yields $\frac{\partial W_H(e)}{\partial \sigma} =$

$$a_1 \cdot \left[-[\pi p(\hat{e} - \sigma)f - h]G'(p(\hat{e} - \sigma)f)\frac{\partial p(\hat{e} - \sigma)f}{\partial \sigma}\right] + a_2 \cdot \left[-[\pi p(\hat{e} + \sigma)f - h]G'(p(\hat{e} + \sigma)f)\frac{\partial p(\hat{e} + \sigma)f}{\partial \sigma}\right].$$

Using $a_1 = a_2$, we get $\frac{\partial W_H(e)}{\partial \sigma} < 0$ if and only if

$$[h - \pi p(\hat{e} - \sigma) f] G'(p(\hat{e} - \sigma) f) f \frac{\partial p(\hat{e} - \sigma)}{\partial \sigma} + [h - \pi p(\hat{e} + \sigma) f] G'(p(\hat{e} + \sigma) f) f \frac{\partial p(\hat{e} + \sigma)}{\partial \sigma} < 0,$$

which is (by rearranging such that every factor is positive)

$$[h - \pi p(\hat{e} + \sigma) f] G'(p(\hat{e} + \sigma) f) \cdot \frac{\partial p(\hat{e} + \sigma)}{\partial \sigma} < [(h - \pi p(\hat{e} - \sigma) f] G'(p(\hat{e} - \sigma) f) \cdot (-\frac{\partial p(\hat{e} - \sigma)}{\partial \sigma}).$$

The last inequality is equivalent to the condition stated in the proposition that G and $p(e)$ are "not too convex." Hence, this condition implies that $\frac{\partial W_H(e)}{\partial \sigma} < 0$ for any e. Since the optimal effort e_H^* is independent of σ, also the maximum welfare $W_H^* = W_H(e_H^*)$ is decreasing in σ. $\square$

B Robustness: Endogenous Perceptions of Naïves

In our basic model, the perceived effort of naïves $\hat{e}$ under regime H is independent of the actual effort e. We now relax this assumption by assuming that $\hat{e}$ is a convex combination of the actual effort e and an exogenous component $\bar{e} \geq 0$, i.e.,

$$\hat{e}(e) := \gamma e + (1 - \gamma) \bar{e}, \tag{B.1}$$

where $\gamma \in [0, 1]$ is the weight placed on the actual effort e. For example, $\bar{e}$ can be interpreted as an anchor in the sense of Tversky and Kahneman (1974), so that $\gamma < 1$ reflects the well-known anchoring bias according to which individuals are sluggish in adjusting their beliefs, thereby putting too much weight on the anchor. In the basic model, the implicit assumption was that there is no adjustment away from the anchor at all, and it is hence nested as the special case $\gamma = 0$ of this more general specification. As the perception $\hat{e}$ is no longer fully exogenous, we now develop all results with respect to the exogenous parameter $\bar{e}$.[22] The following assumption is hence in full analogy to Assumption 1 above.

Assumption B.1. *The anchor $\bar{e}$ satisfies $\pi p(\bar{e}) f < h$, which is equivalent to $\bar{e} < e^{max}$.*

[22] In the baseline model, we had $\hat{e} = \bar{e}$ by construction and all results were described in dependence of this exogenous parameter, which could be labeled either $\bar{e}$ or $\hat{e}$. In the more general model considered here, the exogenous anchor $\bar{e}$ and the perception $\hat{e}$ need not coincide. Again, we describe all results in dependence of the exogenous parameter, which here is $\bar{e}$. Hence, if the results of the baseline model turn out to be robust, we can restate them in the more general framework using $\bar{e}$, where we used to have $\hat{e}$.

The aim of the extension is to show that the results from the basic model are robust as long as there is sufficient weight on the anchor $\bar{e}$. As the perceptions of naïves matter only for regime H the analysis for regime R remains unaffected. If appropriate for consistency, the subsequent formal statements also refer to regime R but the proofs confine attention to regime H.

The problem of the enforcement authority in regime H is then to choose effort e such that the following surplus function is maximized:

$$W_H(e) := (1-a) \cdot \left[\int_{p(e)f}^{\infty} (\pi g - h) G'(g) dg \right] + a \cdot \left[\int_{p(\hat{e}(e))f}^{\infty} (\pi g - h) G'(g) dg \right] - C(e). \quad \text{(B.2)}$$

Formally, the difference to the basic model (see Eq. 2) is that the actual enforcement effort now also affects the behavior of naïves, such that the second integral term also depends on e. Focusing again on interior solutions for the optimal effort $e^*_H(a)$, the respective first-order condition is

$$\begin{aligned} (1-a) \cdot [(h - \pi p(e) f) \cdot G'(p(e)f) \cdot p'(e) \cdot f] & \\ +a \cdot [(h - \pi p(\hat{e}(e)) f) \cdot G'(p(\hat{e}(e)f) \cdot p'(\hat{e}(e)) \cdot \hat{e}'(e) \cdot f] &= C'(e). \end{aligned} \quad \text{(B.3)}$$

Intuitively, $e^*_H(a)$ equates the marginal benefit of deterring more sophisticates *and* naïves (LHS) with the marginal cost of deterrence (RHS). For further reference, this condition can be rewritten as

$$(1-a) \cdot [(h - \pi p(e) f) \cdot G'(p(e)f) \cdot p'(e) \cdot f] = C'(e) - a \cdot b(\gamma), \quad \text{(B.4)}$$

where $b(\gamma) := (h - \pi p(\hat{e}(e))f) \cdot G'(p(\hat{e}(e))f) \cdot p'(\hat{e}(e)) \cdot \gamma \cdot f$ and $\hat{e}'(e) = \gamma$. Note that for $\gamma = 0$, conditions (B.3) and (B.4) coincide with the respective condition (3) for the basic model.

We can now show under which conditions the optimal enforcement effort $e^*_H(a)$, will over-deter neither naïves nor sophisticates:

Lemma B.1. ***(No Over-Deterrence)*** *If the weight on the actual effort γ is sufficiently small, then the (weighed) gain of the indifferent sophisticated or naïve offender resulting under the optimal enforcement effort is below the social harm, i.e. $\pi p(e^*_H(a))f < h$ and $\pi p(\gamma e^*_H(a) + (1-\gamma)\bar{e})f < h$, as well as $\pi p(e^*_R)f < h$ or, equivalently, $e^*_H(a), \hat{e}, e^*_R < e^{max}$.*

Proof. Our assumptions on the cost function $C(e)$ still ensure that the maximizer of (B.2), $e^*_H(a)$, is indeed interior, hence satisfying the first-order condition (B.3). Moreover, the

RHS of Eq. (B.3) is always strictly positive for all $e > 0$. Hence, condition (B.3) can only be satisfied when the LHS is also strictly positive. Since $G'(\cdot) > 0$, and $p'(e) > 0$ and $f > 0$, it follows that either (i) $h - \pi p(e) f > 0$ (i.e., sophisticates are not over-deterred) or (ii) $h - \pi p(\hat{e}(e)) f > 0$ (i.e., naïves are not over-deterred) or (iii) both. Hence, it cannot occur that both groups are over-deterred under the optimal policy $e_H^*(a)$. It remains to show that indeed case (iii) is the relevant one.

No over-deterrence of naïves. Recall that the perceived effort of naïves $\hat{e}$ is a convex combination of the actual effort e and the anchor $\bar{e}$ (i.e., $\hat{e}(e) = \gamma e + (1 - \gamma)\bar{e}$), and that $\bar{e} < e^{max}$ by Assumption B.1. Since $y \in [0, 1]$, a necessary condition for $\hat{e}(e) \geq e^{max}$ (i.e., over-deterrence of naïves) is $e > e^{max}$ (i.e., over-deterrence of sophisticates as well). But this case has been shown to be inconsistent with condition (B.3) above. Hence, it follows that Assumption B.1 is sufficient to ensure that $\hat{e}(e) < e^{max}$ must hold, i.e. naïves are not over-deterred.

No over-deterrence of sophisticates. To rule out that sophisticates are over-deterred, consider the function $b(\gamma)$ as introduced in condition (B.4) above. Note first that no over-deterrence of naïves (i.e., $\pi p(\hat{e}(e_H^*(a))) f < h$) implies that $b(\gamma) \geq 0$ for all $\gamma \in [0, 1]$. Moreover, the first three factors of $b(\gamma)$ are bounded (since $p(\cdot) \in [0, 1]$, $G'(\cdot) < \infty$, and $p'(\cdot) < \infty$), while the fourth ($\gamma = \hat{e}'(e)$), converges to zero for γ small, so that $\lim_{\gamma \to 0} b(\gamma) = 0$. Hence, in condition (B.4), since $C'(e) > 0$, there always exist some $\bar{\gamma} > 0$ such that the RHS is strictly positive for all $\gamma \in [0, \bar{\gamma}]$. This implies that for all $\gamma \in [0, \bar{\gamma}]$, the LHS must also be positive and hence $h - \pi p(e_H^*(a)) f > 0$ must hold, i.e., sophisticates are not over-deterred as claimed. □

Lemma B.1 establishes that the property of no over-deterrence of neither offender type is preserved also for the augmented model where the actual enforcement effort does affect the perception of naïve offenders ($\hat{e}$). Lemma B.1 uses Assumption B.1, which is analogous to Assumption 1 in the basic model, and additionally assumes that the weight γ on the actual effort is not too large. To see why this additional restriction is needed, suppose that $\bar{e}$ is low, so that it is important to increase the deterrence of naïves by choosing a large effort. And as such an increase in effort matters more for the deterrence of naïves if γ is high, the authority may then even tolerate that sophisticates are over-deterred. This is excluded if the weight on the anchor is sufficiently large as the optimal effort is then mainly driven by the deterrence of sophisticates.

In a next step we can restate Proposition 1, the characterization of optimal enforcement effort in regime H. Again, with one minor exception, the results from the basic model (see Proposition 1) also hold for the augmented model, provided that the anchor ($\bar{e}$) gets sufficient weight in the determination of the perception of naïves:

Proposition B.1. ***(Optimal Policy under Regime*** H*) Let Assumption B.1 hold and let the naïves' weight on the actual effort γ be sufficiently small.*

(i) The optimal (interior) effort level $e_H^*(a)$ is strictly decreasing in the share of naïves a and satisfies $e_H^*(0) = e_R^*$, $e_H^*(a) < e_R^*$ for all $a \in (0,1]$, and when $\gamma = 0$, then $e_H^*(1) = 0$.[23]

(ii) For any given share of naïves $a > 0$, the resulting maximum surplus $W_H^*(a, \bar{e})$ is strictly increasing in the naïves' anchor $\bar{e}$.

(iii) When the anchor $\bar{e}$ is sufficiently large, i.e., $\bar{e} > e_H^*(0)$, the social welfare under hiding $W_H^*(a, \bar{e})$ is strictly increasing in the share of naïves a for all $a \in [0,1]$.

(iv) When the anchor $\bar{e}$ is in the intermediate range, i.e. $e_H^*(1) < \bar{e} < e_H^*(0)$, there exists a threshold for the share of naïves $\hat{a} \in (0,1)$, implicitly defined by $e_H^*(\hat{a}) = \bar{e}$, such that $W_H^*(a, \bar{e})$ is strictly decreasing (increasing) in a for all $a < (>)\hat{a}$.

(v) When the anchor $\bar{e}$ is sufficiently small, i.e., $\bar{e} < e_H^*(1)$, the social welfare under hiding $W_H^*(a, \bar{e})$ is strictly decreasing in the share of naïves a for all $a \in [0,1]$.

Proof. The crucial part of the proposition is part (i). Once this is established, the proofs for parts (ii), (iii) and (iv) are completely analogous to the proof of Proposition 1 and part (v) will follow immediately.

(i) As shown in the proof of Lemma B.1 above, the optimal enforcement policy under regime H, $e_H^*(a)$, is interior and satisfies the first order condition (B.3). As for the comparative static properties with respect to the share of naïves a, applying the implicit function theorem yields

$$\begin{aligned}\frac{\partial e_H^*(a)}{\partial a} &= -\frac{(-1)\left[(h - \pi p(e_H^*(a)) \cdot f) \cdot G'(p(e_H^*(a))f) \cdot p'(e_H^*(a)) \cdot f\right]}{W_H''(e_H^*(a))} \\ &- \frac{\left[(h - \pi p(\hat{e}(e_H^*(a)))f) \cdot G'(p(\hat{e}(e_H^*(a)))f) \cdot p'(\hat{e}(e_H^*(a))) \cdot \gamma \cdot f\right]}{W_H''(e_H^*(a))}\end{aligned}$$

which is claimed to be negative. To verify this claim, note first that the denominator in each term is just the second derivative of the surplus function (B.2). At the

[23] Only the last piece differs from the baseline model, where we always had $e_H^*(1) = 0$, while we now need the additional condition $\gamma = 0$. For $\gamma > 0$, we have $e_H^*(a) > 0$ even if $a = 1$.

optimal effort $e_H^*(a)$, this must be negative for $e_H^*(a)$ to be a local optimum. It remains to show that the sum of the two numerator terms is also negative. This is the case if and only if only if

$$\begin{aligned}(h-\pi p(e_H^*(a))\cdot f)\cdot G'(p(e_H^*(a))f)\cdot p'(e_H^*(a))\cdot f &> \\ (h-\pi p(\hat{e}(e_H^*(a)))f)\cdot G'(p(\hat{e}(e_H^*(a)))f)\cdot p'(\hat{e}(e_H^*(a)))f\cdot\gamma &\end{aligned} \tag{B.5}$$

The RHS of inequality (B.5) can be substituted by $b(\gamma)$ (see condition B.4 above). As shown in the proof of Lemma B.1, $\lim_{\gamma\to 0} b(\gamma) = 0$. Moreover, Lemma B.1 establishes that $\pi p(e)f < h$, which makes the LHS of (B.5) strictly positive. Hence, the inequality must hold for γ sufficiently small.

Moreover, for $a = 0$, the surplus functions $W_H(a)$ and W_R coincide and so must the optimal enforcement levels. The property $e_H^*(a) < e_R^*$ for $a > 0$ then follows directly from the above arguments.

Finally, we consider the case $a = 1$ and $\gamma = 0$ (such that there are only naïves and their perception is unaffected by the chosen effort). In this special case, we are back to the baseline model for which we have $e_H^*(1) = 0$ by Proposition 1 part (i).

(ii) Fully analogous to the proof of Proposition 1 part (ii).

(iii-v) In full analogy to to the proof of Proposition 1, we get $W_H^*(a, \bar{e})$ is increasing in a if and only if $e_H^*(a) < \bar{e}$, and they are identical for $e_H^*(a) = \bar{e}$. Part (iii) supposes that $\bar{e} > e_H^*(0)$. Proposition B.1 above shows that $e_H^*(a)$ is decreasing. Hence, we have $\bar{e} > e_H^*(a)$ for all $a \in [0, 1]$ in this case. Thus, $W_H^*(a, \hat{e})$ is strictly monotone increasing in a. Part (iv) supposes that $e_H^*(1) < \bar{e} < e_H^*(0)$. Since $e_H^*(a)$ is continuously decreasing, there must be a threshold $\hat{a}(\bar{e})$ such that $e_H^*(\hat{a}) = \bar{e}(= \hat{e})$. $W_H^*(a, \bar{e})$ is strictly decreasing in a for $a < \hat{a}(\bar{e})$ and strictly increasing when the inequality is reversed. Part (v) finally supposes that $\bar{e} < e_H^*(1)$. Since $e_H^*(a)$ is decreasing, we have $\bar{e} < e_H^*(a)$ for all $a \in [0, 1]$. Thus, $W_H^*(a, \bar{e})$ is strictly monotone decreasing in a.

□

Hence, as long as γ is not too large, Proposition B.1 is basically a restatement of Proposition 1 from the basic model, and there is only one difference: When there are only naïves ($a = 1$), the optimal effort was zero in the baseline model, but it is now strictly positive in the extended model for $\gamma > 0$. The reason is that in the extended model there is some benefit from effort even when there are only naïves, because unlike the basic model it does affect their deterrence. As a consequence, the optimal effort $e_H^*(a)$ might always lie below the perceived effort $\hat{e}$, or, equivalently, below its anchor $\bar{e}$. This leads to the new part (v) of Proposition B.1. However, as shown next, this difference has no effect on the optimal regime choice.

Proposition B.2. ***(Optimal Regime Choice)***

(i) When the anchor of naïve offenders is sufficiently large, i.e., $\bar{e} > e_H^*(0)$, then it is always optimal to hide the effort, i.e., $W_H^*(a, \bar{e}) > W_R^*$ $\forall a \in [0, 1]$.

(ii) Otherwise (i.e, for $\bar{e} < e_H^*(0)$), either regime can be optimal. For $W_H^*(1, \bar{e}) > W_R^*$, there exists a threshold $\tilde{a}(\bar{e}) \in (0, 1)$ implicitly defined by $W_H^*(\tilde{a}(\bar{e}), \bar{e}) = W_R^*$ such that it is optimal to hide (reveal) the effort when the share of naïves is sufficiently large (small), i.e., $W_H^*(a, \bar{e}) > (<) W_R^*$ $\forall a > (<) \tilde{a}(\bar{e})$.

(iii) If $\bar{e} < e_H^*(0)$ and $W_H^*(1, \bar{e}) < W_R^*$ hold, then revealing the effort is always optimal, i.e., $W_H^*(a, \hat{e}) < W_R^*$ $\forall a \in [0, 1]$.

The proof is fully analogous to Proposition 2 and hence omitted.

Proposition B.2 mirrors Proposition 2, which characterizes the regime comparison in the baseline model, in all three parts. The difference between the two settings is hidden behind part (iii) of Proposition B.2. This part includes the (standard) case, in which welfare in regime H is U-shaped in the share of naïves a. This occurs when the anchor $\bar{e}$ is in the intermediate range, i.e. $e_H^*(1) < \bar{e} < e_H^*(0)$. And it also accommodates the new case, in which welfare in regime H is decreasing for any share of naïves a (see Proposition B.1 part (v)). The latter occurs when the anchor $\bar{e}$ is sufficiently small, i.e., $\bar{e} < e_H^*(1)$.

To summarize, when the condition underlying Proposition B.1 is met (i.e., the weight γ of the actual effort on the naïves' perception is not too strong), the results for the regime comparison of the baseline model fully extend to the more general model.

References

Anwar, S. and T. Loughran (2011): "Testing a Bayesian learning theory of deterrence among serious juvenile offenders," *Criminology*, 49, 667–698.

Apel, R. (2013): "Sanctions, perceptions, and crime: Implications for criminal deterrence," *Journal of Quantitative Criminology*, 29, 67–101.

Apel, R. and D. Nagin (2011): "General deterrence: A review of recent evidence," in *Crime and Public Policy*, ed. by J. Wilson and J. Petersilia, Oxford University Press, 41–436.

Armstrong, M. and J. Vickers (2012): "Consumer protection and contingent charges," *Journal of Economic Literature*, 50, 477–493.

Baumann, F. and T. Friehe (2013): "Cheap talk about the detection probability," *International Game Theory Review*, 15, 52–67.

Bebchuk, L. A. and L. Kaplow (1992): "Optimal sanctions when individuals are imperfectly informed about the probability of apprehension," *The Journal of Legal Studies*, 21, 365–370.

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

Ben-Shahar, O. (1997): "Playing without a rulebook: Optimal enforcement when individuals learn the penalty only by committing the crime," *International Review of Law and Economics*, 17, 409–421.

Brown, J., T. Hossain, and J. Morgan (2010): "Shrouded attributes and information suppression: Evidence from the field," *Quarterly Journal of Economics*, 125, 859–876.

Buechel, B. and G. Muehlheusser (2016): "Black sheep or scapegoats? Implementable monitoring policies under unobservable levels of misbehavior," *The Journal of Legal Studies*, 45, 331–366.

Buehler, S. and N. Eschenbaum (2017): "Explaining escalating fines and prices: The curse of positive selection," *mimeo*, University of St. Gallen.

Chalfin, A. and J. McCrary (2017): “Criminal deterrence: A review of the literature,” *Journal of Economic Literature*, 55, 5–48.

Chetty, R., A. Looney, and K. Kroft (2009): “Salience and taxation: Theory and evidence,” *American Economic Review*, 99, 1145–77.

D’Antoni, M. (2018): “Asymmetric Learning and Deterrence when the Probability of Sanctions is unknown,” *mimeo*, University of Siena.

Di Tella, R. and E. Schargrodsky (2004): “Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack,” *American Economic Review*, 94, 115–133.

Donohue, J., D. Ho, and P. Leahy (2013): “Do police reduce crime? A reexamination of a natural experiment,” *Empirical Legal Analysis: Assessing the Performance of Legal Institutions*, 125–143.

Draca, M., S. Machin, and R. Witt (2011): “Panic on the streets of London: Police, crime, and the july 2005 terror attacks,” *American Economic Review*, 101, 2157–81.

Dur, R. and B. Vollaard (2018): “Salience of law enforcement: A field experiment,” *Journal of Environmental Economics and Management, forthcoming.*

Eeckhout, J., N. Persico, and P. E. Todd (2010): “A theory of optimal random crackdowns,” *American Economic Review*, 100, 1104–35.

Englmaier, F., A. Roider, and U. Sunde (2016): “The role of communication of performance schemes: Evidence from a field experiment,” *Management Science*, 63, 4061–4080.

Gabaix, X. and D. Laibson (2006): “Shrouded attributes, consumer myopia, and information suppression in competitive markets,” *Quarterly Journal of Economics*, 121, 505–540.

Garoupa, N. (1999): “Optimal law enforcement with dissemination of information,” *European Journal of Law and Economics*, 7, 183–196.

Heidhues, P., B. Kőszegi, and T. Murooka (2012): “Deception and consumer protection in competitive markets,” *The Pros and Cons of Consumer Protection*, 44.

——— (2017): “Inferior products and profitable deception,” *Review of Economic Studies*, 84, 323–356.

Kleck, G., B. Sever, S. Li, and M. Gertz (2005): “The missing link in general deterrence research,” *Criminology*, 43, 623–660.

Lang, M. (2017): “Legal uncertainty as a welfare enhancing screen,” *European Economic Review*, 91, 274–289.

Langlais, E. (2008): “Detection avoidance and deterrence: Some paradoxical arithmetic,” *Journal of Public Economic Theory*, 10, 371–382.

Lazear, E. P. (2006): “Speeding, terrorism, and teaching to the test,” *Quarterly Journal of Economics*, 121, 1029–1061.

Lochner, L. (2007): “Individual perceptions of the criminal justice system,” *American Economic Review*, 97, 444–460.

Matsueda, R., D. Kreager, and D. Huizinga (2006): “Deterring delinquents: A rational choice model of theft and violence,” *American Sociological Review*, 71, 95–122.

Nagin, D. S. (2013): “Deterrence: A review of the evidence by a criminologist for economists,” *Annual Review of Economics*, 5, 83–105.

Nussim, J. and A. D. Tabbach (2009): “Deterrence and avoidance,” *International Review of Law and Economics*, 29, 314–323.

Paternoster, R. and A. Piquero (1995): “Reconceptualizing deterrence: An empirical test of personal and vicarious experiences,” *Journal of Research in Crime and Delinquency*, 32, 251–286.

Pogarsky, G. and A. R. Piquero (2003): “Can punishment encourage offending? Investigating the “resetting” effect,” *Journal of Research in Crime and Delinquency*, 40, 95–120.

Polinsky, A. M. and S. Shavell (2000): “The economic theory of enforcement of law,” *Journal of Economic Literature*, 38, 45–76.

——— (2007): “The theory of public enforcement of law,” in *Handbook of Law and Economics*, ed. by A. M. Polinsky and S. Shavell, Elsevier, vol. 1, 403–454.

Sah, R. K. (1991): "Social osmosis and patterns of crime," *Journal of Political Economy*, 99, 1272–1295.

Sanchirico, C. W. (2006): "Detection avoidance," *New York University Law Review*, 81, 1331–1399.

Stafford, M. and M. Warr (1993): "A reconceptualization of general and specific deterrence," *Journal of Research in Crime and Delinquency*, 30, 123–135.

Stigler, G. J. (1970): "The optimum enforcement of laws," *Journal of Political Economy*, 78, 526–36.

Tabbach, A. D. (2009): "The social desirability of punishment avoidance," *The Journal of Law, Economics, & Organization*, 26, 265–289.

Tversky, A. and D. Kahneman (1974): "Judgment under uncertainty: Heuristics and biases," *Science*, 185, 1124–1131.