Optimal Law Enforcement with Sophisticated and Naïve Offenders*

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

March 2, 2020

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 the actual enforcement effort into account, 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, Underinference, Salience

*We gratefully acknowledge comments and suggestions by the associate editor und by two referees. 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. The economic analysis of crime and deterrence in the tradition of Becker (1968) assumes that potential offenders perfectly observe or anticipate the authority's enforcement effort (see the survey by Polinsky and Shavell, 2007), or at least that they form unbiased expectations (Bebchuk and Kaplow, 1992; Garoupa, 1999). In this paper, we extend the economic model of crime by taking into account three behavioral observations from criminology: First, perceptions of the risk of apprehension are often biased. For instance, respondents in a survey by Kleck et al. (2005) assumed apprehension frequencies of 51.9% for murder and 37.9% for burglary, compared to actual arrest rates of 90.2% and 14.2%, respectively. Overall, the risk of apprehension seems to be rather over- than underestimated (Apel and Nagin, 2011). Second, perceptions of potential offenders are only sluggishly adjusted to changes in the enforcement policy (Chamlin and Langworthy, 1996; Kleck and Barnes, 2014). Third, while sluggish, these perceptions of potential offenders can still be influenced by e.g. media reports and official announcements about (changes in) the risk of apprehension and sanctions (see Apel, 2013 and the survey by Chalfin and McCrary, 2017).

Motivated by these three observations, this paper addresses two research questions: First, if risk perceptions of (some) potential offenders respond only imperfectly to the actual intensity of law enforcement, what are the consequences for the socially optimal enforcement policy? Second, under which circumstances should authorities release additional information on the probability of apprehension that is likely to influence the risk perceptions of potential offenders?

To this end, we develop a model with two offender types, *sophisticates* and *naïves*: Sophisticated offenders always have correct perceptions about the risk of apprehension and hence behave as in the canonical Becker (1968) framework. Conversely, the perceptions of *naïve* offenders are biased as they are not (fully) driven by the actual enforcement activities. Rather, they put too much weight on other (exogenous) factors such as, for example, own experience (Matsueda et al., 2006; Pogarsky et al., 2004; Anwar and Loughran, 2011) and observations in the social network (Stafford and Warr, 1993; Paternoster and Piquero, 1995; Apel and Nagin, 2011).[1] Such behavior is not only well-documented in criminology,

[1] The distinction between naïve and sophisticated types is borrowed from a literature in behavioral industrial organization (Gabaix and Laibson, 2006; Heidhues, Kőszegi, and Murooka, 2012; Armstrong and Vickers, 2012; Heidhues, Kőszegi, and Murooka, 2017), where not all customers familiarize themselves with all terms of a contract.

but is more generally also in line with the *underinference* bias (also called conservatism bias) known in behavioral economics. Individuals subject to this bias form beliefs that are not sufficiently responsive to relevant information.[2] The empirical relevance of this bias is, for example, documented by Benjamin (2019), who summarizes the empirical literature on probability formation by concluding that "underinference is by far the dominant direction of bias".

For our first research question, we consider the case where the authority does not actively attempt to influence the risk perceptions of naïve offenders. The optimal enforcement effort then decreases in the share of naïve offenders as this effort matters more for sophisticates. Whether naïves over- or underestimate the probability of apprehension is endogenous and depends both on their risk perception itself and on their share in the population: If the share of naïves is large, then the authority's optimal effort is low and naïve offenders tend to overestimate the actual effort. In this case, sophisticates largely benefit from the existence of naïves and are less deterred than naïves. Conversely, if the share of naïves is low, the optimal enforcement effort tends to exceed the naïves' perception who would then commit more offenses than sophisticates. These properties give rise to a (maximized) social welfare function which is U-shaped in the share of naïves. In equilibrium, there is either a high enforcement effort underestimated by naïves (when their share is low) or a low enforcement effort overestimated by them (when their share is high).

As for our second research question, we analyze the agency's incentive to publicly reveal information about its enforcement policy.[3] In the empirical literature, such communication has been identified as an important tool to influence deterrence (see e.g. Apel and Nagin, 2011; Apel, 2013; Chalfin and McCrary, 2017). For example, in the "Operation Ceasefire" as discussed in Braga et al. (2001), the Boston police implemented a number of specific enforcement measures against criminal youth gangs, and the police also let gang members know that they would do everything legally possible to punish them collectively. As a second example, the police department of the city of Berlin informed

[2] In the main body of the text, we derive our results for the simplest case in which the perceptions of naïve offenders are entirely independent of the authority's actual enforcement effort. In Appendix B, however, we consider a richer model where the actual effort does affect the perceptions of naïves. We show that all results remain unchanged as long as the perceptions of naïves' do not respond too strongly to the actual effort.

[3] While some information about enforcement activities and detection rates is publicly available from official sources (see e.g. `https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017`), it tends to be highly aggregated, so that many offenders might find it difficult to correctly infer their actual risk of apprehension.

the public that the police force in criminal hot spots would be tripled.[4] Moreover, in the context of tax fraud, the German federal state of Brandenburg publicly announced a considerable increase of its budget to carry out tax investigations.[5] Finally, the public transport authority of the German city of Frankfurt informed the public about a 52% increase in the number of ticket inspectors.[6]

Note that, in all of these examples, the announced policy involves a credible pledge to high(er) enforcement effort. This phenomenon arises as an equilibrium feature of our model. In particular, the enforcement authority simultaneously chooses both its effort and whether to keep it hidden or reveal it. Under hiding, a given effort is observed by sophisticates only, while under revelation, it is observed by all offenders (thereby making naïves and sophisticates identical). We show that two types of optimal enforcement policies emerge:[7] The first type is to set a high effort and to reveal it (as in the above examples), thereby generating an upwards shift in the risk perceptions of naïves. The second type is to set a low effort, which is kept hidden, and which is below the perceptions of naïves, who therefore in equilibrium overestimate the probability of apprehension. The first policy, choosing high effort which is revealed, is more costly but yields higher deterrence of both naïves and sophisticates. We show that it is optimally chosen by the agency as long as naïves would not strongly overestimate their (equilbrium) probability of apprehension under hiding. Revelation tends to be optimal when the share of naïves is not too large, such that even the optimal effort under hiding is relatively high.

Our results are in line with some of the empirical findings quoted above, and they also suggest scope for potentially beneficial policy interventions. Consider first the empirical findings that offenders' perceptions of the probability of apprehension do often not correspond to the actual level and that overestimation of the probability of apprehension seems to be more prevalent than underestimation. In our model, it turns out that the optimal enforcement policy never matches the expectations of naïves: Either the optimal effort is higher than these perceptions and revealed, or it is lower and remains hidden. Since the

[4]See e.g. https://www.rbb24.de/politik/beitrag/2019/02/polizei-reform-berlin.html.

[5]See e.g. https://www.pnn.de/brandenburg/land-erhoeht-zahl-der-steuerfahnder-67-millionen-euro/21315690.html.

[6]See e.g. http://www.fr.de/rhein-main/verkehr/rmv-und-vgf-schaerfere-kontrollen-gegen-schwarzfahrer-a-567360.

[7]The question whether or not to reveal the enforcement effort may sound trivial once we assume that more detection is desirable and that revelation is costless: A fixed effort should be revealed if and only if it is larger than expected. In reality, however, the enforcement effort is typically not fixed, but an important dimension of the enforcement policy that is to be chosen. This is taken into account in our framework.

incorrect expectations in the former case are corrected, underestimation of enforcement effort does not occur in equilibrium.[8]

From a policy perspective, in areas where the probability of apprehension was found to be overestimated, our results suggest that it can indeed be optimal for enforcement agencies to hide the low effort actually chosen, and to instead "free-ride" on the deterrence arising from the perceptions of many offenders. By contrast, when offenders tend to underestimate the probability of apprehension, an agency might prefer to make it more salient that it commits considerable resources to enforcement and even increase them, and that many offenders are indeed detected.

We extend our basic model in three empirically relevant directions. First, providing salient information may facilitate evasion activities. For example, the New York City Council recently introduced a policy aimed at increasing the transparency and accountability over the NYPD's use of powerful new surveillance tools. This policy aims at reducing the risk of information abuse but officials fear that it may also reduce the efficacy of their effort.[9] We find that the reduction in the efficacy due to revelation leads to an interesting non-monotonicity for the regime comparison as revealing the enforcement effort is only optimal for intermediate shares of naïves.[10] This also leads to naïves underestimating the enforcement effort as an equilbrium feature of the model.

Second, we consider the case where the size of the sanction is a further policy variable of the authority. We find that the classical result on the optimality of maximum fines (Becker, 1968) does not necessarily hold in our setting with sophisticated and naïve offenders: If the share of naïves is large, then the authority has an incentive to choose low effort and impose a high fine. But if the fine becomes too high, then even naïve offenders with high benefits from committing the act would be (inefficiently) deterred as they are unaware of the actually low probability of apprehension. Perception biases hence also influence the optimal fine.

Third, we allow for differences in the naïves' risk perceptions. This seems important as

[8] As shown in two model extensions (explained in more detail below), this feature of our baseline model changes when either revelation has additional downsides or when there is heterogeneity in the perceptions of naïves.

[9] See `https://www.huffingtonpost.com/entry/new-york-city-is-making-its-citizens-safer-by-overseeing-police-technology_us_58e23f04e4b0ba359596583b`, lastly retrieved on May, 28, 2018.

[10] Nussim and Tabbach (2009) demonstrate that higher enforcement effort may actually trigger higher crime levels if criminals can invest in detection avoidance activities. Tabbach (2010) finds that costly avoidance strategies can even increase social welfare if they are not completely deterred. The intuition is that anything that increases an offenders's costs can serve as substitute for (costly) punishment, so that avoidance activities should be made expensive but should not be completely eliminated.

heterogeneity is known to be an important determinant of offenses which is, for example, reflected in the fact that young survey respondents who assume lower probability of apprehension are more likely to commit auto theft (Lochner, 2007). After showing that our main results prevail, we demonstrate that heterogeneity reinforces an inefficiency that arises in the regime with hidden enforcement effort: some offenders with high private benefits may be deterred, while others with low benefits commit the act.

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

2 Relation to the Literature

A large body of empirical work has provided systematic evidence that supports crucial ingredients of our framework: First, perceptions on the probability of apprehension depend on the own personal environment of potential offenders (Apel and Nagin, 2011; Nagin, 2013) and are often only vaguely related to the actual risk (Chalfin and McCrary, 2017; Kleck and Barnes, 2014). Second, studies taking endogeneity issues into account still find a deterrence effect of police strength (Levitt and Miles, 2007; Rosenfeld et al., 2007). Third, perceptions can be influenced by the authority's communication strategy (Chalfin and McCrary, 2017).

There are also a number of theoretical approaches to which our paper is related. In a model with noisy but unbiased signals about the true probability of apprehension, Garoupa (1999) shows that revealing the effort is always welfare-enhancing. The reason is that heterogeneous beliefs misallocate offenses among subjects with high and low private benefits. This property also exists in our model. But since we allow for systematic misperceptions, this gives the authority the possibility to free-ride on the perceptions of naïve offenders when keeping the enforcement effort hidden. This can indeed be the optimal policy when this effect is sufficiently strong. As we do in our second extension, Bebchuk and Kaplow (1992) find a rationale for fines below the maximum one. In their model, maximum fines reinforce the negative impact of heterogeneous perceptions. In our approach maximum fines may inefficiently deter subjects who overestimate the actual risk of apprehension. Lang (2017) shows that legal uncertainty about whether an activity is forbidden or not may enhance social welfare as only offenders with high private benefit will commit an offense.

There are also a number of papers that consider heterogenous perceptions in a dynamic context:[11] Sah (1991) develops a rich model where perceptions may change over time, but treats the authority's policy as exogenous. Our model is complementary in the sense that it takes offenders' (potentially biased) perceptions as given, and focusses on the optimal enforcement policy. In Ben-Shahar (1997), heterogeneous perceptions also 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 increase first-period arrests by setting low fines, which in turn increases the percentage of individuals who commit the offense in the second period only when their private gains are large. By contrast, our static model considers the revelation of the enforcement effort as part of the optimal policy in a setting with heterogeneous perceptions. 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 who underestimate the true probability have higher violation incentives, so that their bias is reduced while the overestimation of non-violators persists. Our model highlights a different reason for the phenomenon of over-estimation: when an authority decides to choose an effort below the perceptions of naïves, it might not have an incentive to reveal that information.

An important assumption of our model is that the authority can credibly reveal its detection effort. While it seems sensible to assume that, in many countries, official statistics are trustworthy, authorities may well emphasize some facts (such as the number of newly hired inspectors) while keeping silent about others (such as the fact that fare dodging still pays in expectation). Our model covers this only in a stylized way since revelation leads to a perfect update of the probability of apprehension. Baumann and Friehe (2013) consider the opposite case where the regulator's announcements are considered as cheap talk and hence need to be incentive compatible. Baumann and Friehe (2013) also find that not revealing the actual detection probability might be welfare maximizing, but for a quite different reason. In their model, communication of the exogeneous detection probability can cause under-deterrence or over-deterrence. While they focus on credibility issues, we are mainly interested in the interdependency of the decisions on the probability of appre-

[11] As shown in a variety of empirical studies, history matters largely for risk perceptions and behavior (see e.g. Lochner, 2007, Anwar and Loughran, 2011, Apel, 2013).

hension and its revelation. In our view, the two approaches are complementary as there seems to be a spectrum of more or less credible communication: while general statements like “crime does not pay” are on the cheap talk side, revealing concrete information concerning the enforcement policy (e.g. the number of ticket inspectors) seems much more credible (at least in countries where one can reasonably assume that authorities do not deliberately release false information). Interestingly, the two approaches exhibit different comparative statics properties: In Baumann and Friehe (2013) revelation is only incentive compatible for low levels of harm and high costs of sanctioning, while in our model, revelation emerges as an equilibrium feature when the harm is high and enforcement costs are low.

The possibility to either hide or reveal the enforcement effort 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 (2019) 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.[12]

The decision whether or not to reveal the enforcement effort is also discussed in the context of tax evasion. Cronshaw and Alm (1995) assume that a tax authority, whose effort costs are private information, decides on its effort simultaneously with tax payers choosing the honesty of their report. They then show that in a Bayesian Nash-mixed strategy equilibrium, reducing uncertainty on the authority’s effort costs increases compliance under reasonable assumptions. Snow and Warren (2005) show that uncertainty on the detection probability increases tax compliance with ambiguity averse offenders. As uncertainty also enhances betrayal incentives for ambiguity lovers, however, they conclude that increasing uncertainty is no sound policy option. Lang and Wambach (2013) extend

[12]The effects of the salience of policies has also been studied in other contexts such as taxation and incentive schemes, see e.g. Brown, Hossain, and Morgan (2010), Chetty, Looney, and Kroft (2009), Englmaier, Roider, and Sunde (2016).

the analysis to a market setting where insurance companies decide on whether to commit to an auditing strategy. They find that no commitment can be a market equilibrium due to the deterrence effect on ambiguity-averse offenders. Thus, while considering different settings, also these papers provide rationales for why it might be optimal to not reveal the chosen enforcement effort.

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.[13]

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 $(1 - a)$ 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.[14] When it remains hidden, they perceive it to be $\hat{e} \geq 0$ instead. 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 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 actual enforcement effort does affect the perception of naïves, as long as this effect is not too strong. For simplicity, all naïve agents have the same perception $\hat{e}$ in the baseline model, while the extension to heterogeneous perceptions is analyzed in Section 5.3. Moreover, the gain distribution G applies to both sophisticated and naïve offenders.

Irrespective of the type, each offender is detected with probability $p(e)$, which satisfies $p(0) = 0$ and $lim_{e\to\infty} p(e) = 1$. Moreover, $p(e)$ is assumed to be twice continuously

[13] 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).

[14] 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 ($a = 0$).

differentiable and strictly increasing, with $0 < p'(e) < \infty$, for all effort levels $e > 0$. 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 $C(0) = 0$ and the Inada condition $C'(0) = 0$. Moreover, $C(e)$ is assumed to be strictly increasing and weakly convex, i.e. $C'(e) > 0$, $C''(e) \geq 0$ for all $e > 0$.

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.[15] 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-peaked under both regimes $T = H, R$ such that we get a unique interior optimum with

[15] 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. Even when gains from an offense receive a weight $\pi < 1$, we always treat the payment of fines as purely redistributive.

respect to the enforcement effort.[16]

4 Optimal Enforcement Policy

4.1 Basic Properties of Optimal Enforcement Effort

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 $W_H^*(a, \hat{e}) := W_H(e_H^*(a); a, \hat{e})$.[17] 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:

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

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

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

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.

[17] 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 $\hat{e}$.

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 results are nevertheless available. If there were no costs of enforcement, the enforcement authority would deter all potential offenders whose (weighed) gain is below social harm and not deter the others, i.e. the optimal enforcement effort would just balance the (weighed) gains of the indifferent offender $\pi g_R^j = p(e)f$ with the social harm h such that $\pi p(e)f = h$. Reaching this high level of deterrence is only possible when the exogenous fine is not too small: $f > \frac{h}{\pi}$, which we assume from now on.[18] For notational convenience, we define $e^{max} := p^{-1}\left(\frac{h}{\pi f}\right)$ 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 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. To simplify the analysis, 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 $\hat{e} < e^{max}$.*

Assumption 1 ensures that the (weighted) gain of the indifferent naïve offender is lower than the social harm. This implies that naïves are not over-deterred. Thus, in our analysis of the baseline model we can focus on the settings in which more deterrence is always desirable for the social planner.[19]

4.2 Optimal Enforcement Effort under Regime H

We now use the insights from the previous section to characterizes the optimal enforcement effort under regime H, i.e. when this effort is not revealed to the potential offenders:

[18] This assumption also implies $\pi \cdot 1 \cdot f > h$, which does not affect our results, but puts the focus on the relevant scenarios. It excludes the somewhat unrealistic scenario in which even with maximal effort, it is impossible to deter all offenders that should be deterred when effort is costless. By setting $\pi = 1$ as is standard in the literature, the assumption reads $f > h$, which is a typical feature of enforcement policies.

[19] This assumption will be relaxed in the model extension considered in Subsection 5.2 below.

Proposition 1. ***(Optimal Enforcement Effort 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. Given Assumption 1, which ensures that naïves are not over-deterred, this directly implies that the maximum attainable surplus under hiding increases in $\hat{e}$.

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. $g^n_H = p(\hat{e})f > p(e^*_H(a))f = g^s_H$ 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 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 ($g^n_H = p(\hat{e})f < p(e^*_H(a))f = g^s_H$), 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).

4.3 Optimal Regime Choice

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 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 for all offenders, while $\hat{e}$ is no longer relevant. Therefore, enforcement effort under regime H is decreasing in the share of naïves a and satisfies $e^*_H(a) < e^*_R$ for all $a \in (0, 1]$.

Intuitively, there are thus two candidate optimal enforcement policies (i.e. an effort choice and the decision whether to reveal it or to keep it hidden) for the agency: Either it sets a high effort and reveals it, or it chooses a low effort and keeps it hidden. The former policy is more costly, but it leads to higher deterrence for both types of offenders. The following result provides a detailed characterization under which conditions each of these two candidate policies is optimal.

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 $\tilde{a}(\hat{e}) \in (0, 1)$ implicitly defined by $W_H^*(\tilde{a}(\hat{e}), \hat{e}) = W_R^*$ such that it is optimal to hide (reveal) the effort when the share of naïves is sufficiently large (small), i.e. $W_H^*(a, \hat{e}) > (<) W_R^* \; \forall a > (<) \tilde{a}(\hat{e})$.

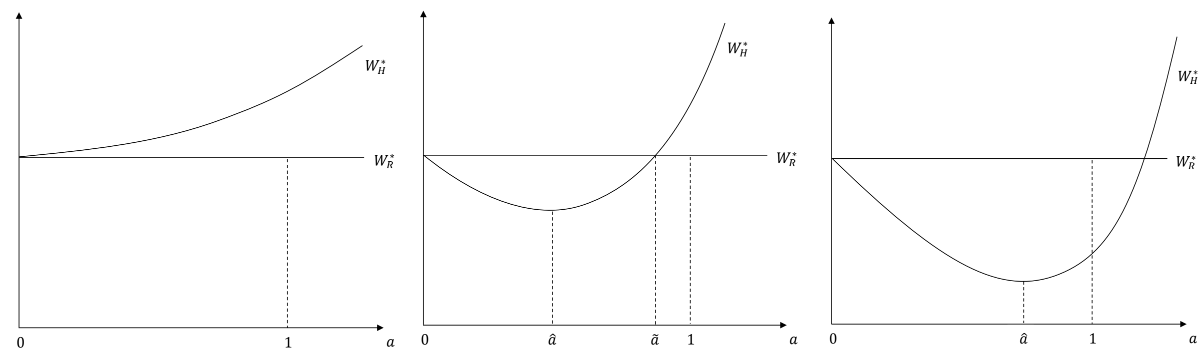
(iii) If $\hat{e} < e_H^*(0)$ and $W_H^*(1, \hat{e}) < W_R^*$ hold, then revealing the effort is always optimal, i.e. $W_H^*(a, \hat{e}) < W_R^* \; \forall a \in (0, 1]$.

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). This result relies on the model feature that more deterrence is always desirable for the social planner, which holds due to Lemma 1 and Assumption 1.[20]

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 where $W_H^*(a, \hat{e}) = W_R^*$ holds lies inside or outside the range $a \in (0, 1)$ (see panels (ii) 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.

To summarize the result in an intuitive way, ignore the trivial case (i) and recall that the two regimes coincide in the special case that there are no naïves ($a = 0$). In this case, a high effort is optimal as it is effective for all potential offenders. When the share of naïves is not zero but small, the optimal effort under hiding is still relatively high (since the many sophisticates do respond to it). As the few naives would then underestimate the effort, it is optimal to reveal it. In fact, the effort under revelation is even slightly higher than under hiding since now all offenders respond to it. As the share of naïves increases, the optimal effort and the resulting cost under hiding decrease (Proposition 1). This generates a low level of deterrence for (fewer and fewer) sophisticates. However, the (more and more) naïves eventually overestimate the probability of apprehension, which

[20] Suppose Assumption 1 is violated, i.e. $\hat{e} > e^{max}$. Then there is an additional downside of regime H as some naïves are over-deterred. To change the regime comparison, however, this effect would have to outweigh the welfare loss from under-deterrence that always emerges in regime R (see Lemma 1).



(i) hiding is always optimal (ii) hiding is optimal if $a > \widetilde{a}$ (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.

generates constant deterrence for this group at no additional cost for the agency. When the share of naïves is sufficiently large, this effect might become sufficiently strong so that hiding in combination with a low effort indeed constitutes the optimal policy.[21] While this policy saves on enforcement costs, it induces a misallocation of crime in the sense that the minimum private benefit for committing the crime is lower for sophisticates than for naïves.

These findings can lead to empirically testable predictions.As an example, suppose 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).

[21] Note that the cutoff level $\widetilde{a}$ for the optimal regime switch does not coincide with the point $\hat{a} < \widetilde{a}$, which determines whether naïves over- or underestimate the optimal effort under regime H. Hence, the underestimation of this effort is a necessary, but not sufficient, condition for the optimality this regime.

4.4 Implications

The properties of the optimal enforcement policy can be useful to rationalize some of the empirical findings discussed above (e.g. Kleck et al., 2005; Apel and Nagin, 2011), and also to suggest scope for potentially beneficial policy interventions.

First, one implication of Proposition 2 is that the optimal enforcement effort never matches the expectations of naïves. That is, for any ex ante perception $\hat{e}$ of naïves, the authority will always optimally choose an enforcement effort that differs from it.[22] In the light of this finding, it might appear less surprising that empirical evidence shows that offenders' perceptions of the probability of apprehension do often not correspond to the actual level.

Second, a further implication of the proposition is that the enforcement effort is never underestimated in equilbrium. This is broadly in line with the empirical evidence showing that overestimation of the probability of apprehension seems to be more prevalent than underestimation (Apel and Nagin, 2011). But of course, this somewhat strong result is also due to the fact that, in the baseline model, revealing the effort does not reduce its effectiveness. This assumption is relaxed Section 5.1 below.

Third, our analysis supports the empirical evidence in situations where the probability of apprehension is often overestimated. As the proposition shows, it can indeed be optimal for enforcement agencies to hide the low effort actually chosen, and to instead "free-ride" on the deterrence arising from the seemingly high perceptions of many offenders.

Fourth, a policy recommendation arises for situations where the empirical evidence suggests that many offenders underestimate the probability of apprehension (and hence the underlying enforcement effort). Our results suggest that it might be better to even increase the enforcement effort, and accompany this with a campaign that makes the high probability of apprehension more salient to potential offenders. The examples for the revelation of enforcement effort as discussed in the Introduction indeed show that revelation is often combined with a high or increased level of enforcement effort.

Finally, using comparative statics analysis, one can also investigate how the regime comparison is affected by changes in the model parameters. To fix ideas, consider the cost function $C(e) = kc(e)$. It is then easily shown that, ceteris paribus, revealing the

[22]To see this, suppose first that regime H and effort $e_H^*(a) = \hat{e}$ are chosen. By Proposition 2, this is the case for $a = \hat{a}$ as illustrated in Figure 1. As it can be seen from the figure, in that case revelation together with a higher effort e_R^* would instead be optimal. Suppose now that regime R and effort $e_R^* = \hat{e}$ are chosen. The enforcement authority could replicate the same welfare with the regime H and the effort $e_H = \hat{e}$, but further improve it by lowering the effort to its optimal level e_H^*.

enforcement effort becomes relatively more attractive when the costs of enforcement (k) or the weight of offenders' gains in the social welfare function (π) are low, and when the harm from the offense (h) is high.[23] These effects are all intuitive once we account for the key result that revelation is optimally combined with a high enforcement effort.

Generally, our results suggest that a change in the communication policy should be accompanied by a change in the level of enforcement activity.

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

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

which, again, equalizes the expected social loss from the marginal offense with the marginal

[23]Formally, let $e_R^* > \hat{e}$ (otherwise, regime H is always optimal) and define $\Delta^* := W_R^* - W_H^*$ which measures the relative attractiveness of regime R. It is then straightforward to show that Δ^* is decreasing in k and π, and increasing in h.

cost of effort. In comparison to the baseline model, a lower effectiveness of the enforcement effort reduces the attainable maximum welfare, as the following lemma shows.

Lemma 2. *When revealing the effort reduces its effectiveness, the maximum surplus is smaller compared to the basic model, i.e.* $\widetilde{W}_R^* < W_R^*$.

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.* $W_H^*(a, \hat{e}) > \widetilde{W}_R^* \ \forall a \in (0, 1]$.

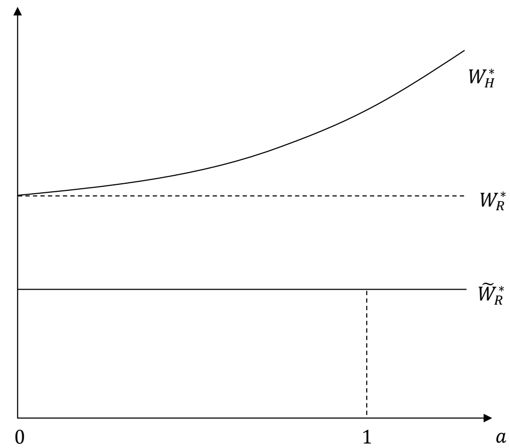
(ii) For $\hat{e} < e_H^*(0)$*, it is optimal to hide the effort if the attainable surplus under revealed effort is sufficiently low, i.e. if* $\widetilde{W}_R^* < W_H^*(\hat{a}, \hat{e})$*, where* $\hat{a}$ *is the share of naïves at which* $e_H^*(\hat{a}) = \hat{e}$.

(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 hide the effort if the share of naïves is either smaller than the first threshold or larger than the second (i.e.* $a < a_1$ *or* $a > a_2$*), but not in-between.*

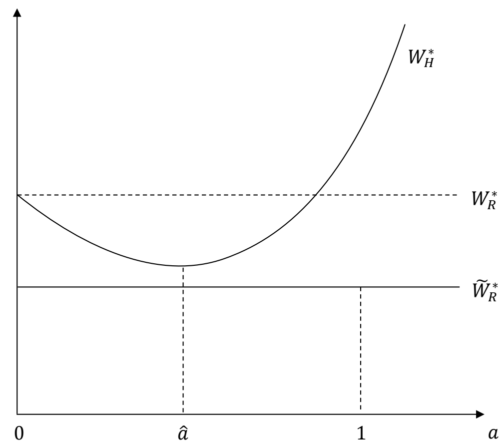
(iv) By contrast, if the two conditions from part (iii) hold, but if $W_H^*(1, \hat{e}) < \widetilde{W}_R^*$ *(implying* $a_2 > 1$*), it is optimal to hide (reveal) the effort if the share of naïves is sufficiently small (large).*

Proposition 3 is illustrated in Figure 2, and has a similar structure as Proposition 2. Again, 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 compared to the baseline model: 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.

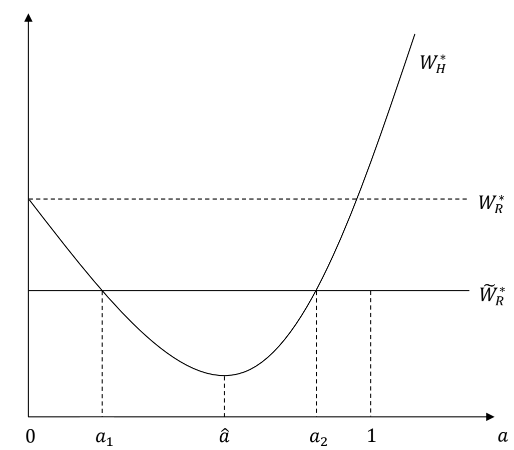
Second, as for parts (iii) and (iv), when hiding is not globally optimal, the regime comparison becomes non-monotonic in the share of naïves (a). In particular, there now also exists an interval of small values of $[0, a_1]$ where hiding is optimal. This qualitative



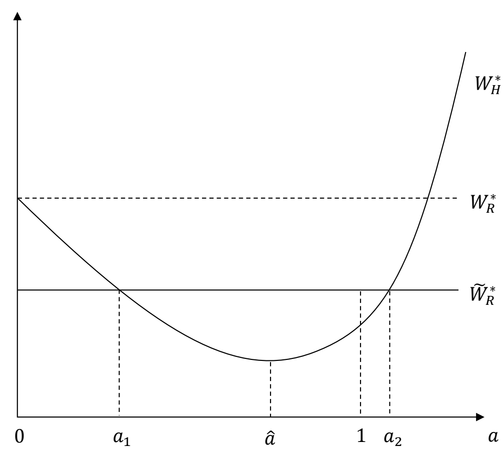
(i) hiding is always optimal



(ii) hiding is always optimal



(iii) hiding is optimal if $a < a_1$ and if $a > a_2$



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

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.

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. Note that, in this interval, the effort optimally remains hidden and is underestimated by naïve offenders in equilbrium (since a_1 is smaller than $\hat{a}$, for which W_H^* is minimum). This feature did not arise in the baseline model. Finally, 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 only chooses its enforcement effort e, while fines have been chosen by other parties such as legislators. In other settings, it is the enforcement authority which 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 it is typically less costly to increase the fine compared to increasing the effort. As a consequence, it is generally 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. Intuitively, 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.[24] As for the regime comparison, we find that allowing for endogenous fines is either neutral or 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 be given by law or by the wealth of offenders.[25]

[24]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.

[25]Since f is now a choice variable, unlike in the baseline model, we do not force it to satisfy Assumption 1 or the assumption $f > \frac{h}{\pi}$.

The optimization problem of the enforcement authority from the baseline model (see the surplus function (1)) then needs to be adapted as follows:

$$\max_{\substack{e\geq 0, f\geq 0\\ T\in\{H,R\}}} \quad W_T(e) := (1-a)\cdot\left[\int_{p(e)\cdot f}^{\infty}(\pi g - h)G'(g)dg\right] + a\cdot\left[\int_{g_T^n}^{\infty}(\pi g - h)G'(g)dg\right] - C(e),$$

$$\text{subject to } f \leq F.$$

The authority can now also affect the respective threshold for the marginal offenders (lower bounds of the two integrals) through its choice of f (recall that $g_T^n \in \{p(e)f, p(\hat{e})f\}$).[26] 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)*** *Suppose $a \in (0,1)$ and $\hat{e} > 0$. When the enforcement authority also chooses the fine $f \in [0, F]$ in addition to its enforcement effort e and the regime $T \in \{H, R\}$, 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.[27]

(ii) In regime H, when the maximal fine F is not too large, then setting $f_R^* = F$ is optimal. In this case 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 maximal fine F is sufficiently large, then the optimal fine f_H^* is interior, i.e. $f_H^* < F$. In this case the optimal enforcement effort is below the naïves' 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. $\pi p(e_H^*) f_H^* < h < \pi p(\hat{e}) f_H^*$.

(iv) For the extended model with a maximal possible fine F, consider the corresponding baseline model where the fixed fine $\bar{f}$ equals F.[28] *If regime H leads to higher welfare than regime R in the baseline model, then this also holds in the corresponding extended model.*

[26] Formally, this is a constrained optimization problem. Of course, besides the crucial constraint $f \leq F$, there are two more constraints, $e \geq 0$ and $f \geq 0$. However, as explained in the proof of Proposition 4 they are never binding in our setting, and hence are left out here for the sake of readability.

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

[28] The constructed baseline model need not satisfy Assumption 1.

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.

As it can be seen from the proof, the relevant first order conditions in each of these two cases are not different from the corresponding first order conditions, (3) and (5), in the baseline model. Also the second order conditions boil down to concavity of the surplus function $W_T(e, f)$ in e, as it was the case in the baseline model.

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 $\pi p(\hat{e})f \geq h$.[29] Further increasing the fine then involves a trade-off between deterring more sophisticates, which is still desirable (as $\pi p(e_H^*)f_H^* < h$), and deterring more naïves. The optimal interior fine $f_H^* < F$ satisfies two relevant first order conditions, one of which shows this novel trade-off:

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

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).[30]

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 other situations where not committing an offense leads to high opportunity costs, e.g. when violating a production standard of minor importance leads to a large cost reduction.

Finally, part (iv) of Proposition 4 shows that endogenizing the fine works in favor of regime H, as it gives the enforcement authority 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,

[29] In the baseline model, Assumption 1 precluded this case. It has become possible because the fine can be chosen to be high.

[30] 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. The second order conditions are considered in the proof of Proposition 4.

this may also be the case, but we have just seen that it may also be optimal to implement a lower fine. Overall, extending the model to endogenous fines leads largely to the choice of a maximum fine, which is a replication of the baseline model (with $f^* = F = \bar{f}$), but it may also lead to a lower fine in regime H in order to counteract over-deterrence.

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.[31] 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 given by:

$$W_H(e) := (1-\sum_{l=1}^{L} a_l)\cdot\left[\int_{p(e)f}^{\infty}(\pi g - h)G'(g)dg\right]+\sum_{l=1}^{L} a_l\cdot\left[\int_{p(\hat{e}_l)f}^{\infty}(\pi g - h)G'(g)dg\right]-C(e), \quad (9)$$

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.*

[31] 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.

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, since 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. Intuitively, the optimal policy reflects the trade-off that enforcement costs can be saved under hiding (due to the low optimal effort), but leads to low deterrence of sophisticates.

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 be 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 (behavioral) 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. This could make the analysis of the optimal enforcement policy (and the decision concerning revelation) more intricate due to a potentially non-monotone relationship between the enforcement effort and the number of apprehensions (see e.g. Buechel and Muehlheusser, 2016). Second, a similar effect on the optimization problem arises when the fine is no longer purely re-distributive as in our framework, but for example receives the same weight (π) as the offenders' gains. In this case, each apprehended offender generates a welfare gain, so that the enforcement authority also cares about the number of apprehensions. Third, 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. Fourth, 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.[32] 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. Fifth, the probability of apprehension might not only depend on the enforcement effort and the sanction, but moreover be weakly decreasing in the number of actual and/or potential offenders due to capacity constraints in the enforcement activity. To our knowledge, this has not yet been studied in an economic model of law enforcement. 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.

[32] 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. □

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

In the model with reduced effectiveness, consider first optimal welfare:

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

Now, we construct effort e' in the baseline model that leads to the same probability of apprehension as the optimal effort $\widetilde{e}_R^*$ under reduced effectiveness, i.e. $p(e') = \tilde{p}(\widetilde{e}_R^*)$. This is possible because $p(0) = 0 < \tilde{p}(\widetilde{e}_R^*) < 1 = \lim_{e \to \infty} p(e)$. Evaluating $W_R(e)$ from the surplus function (4) at $e = e'$ yields

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

Taking the difference $W_R(e') - \widetilde{W}_R^*$ yields $-C(e') + C(\widetilde{e}_R^*)$ (because $\widetilde{p}_R(\widetilde{e}_R^*)f = p(e')f$ by construction). Since $\widetilde{p}_R(e) < p(e)$ $\forall e > 0$ and both functions are strictly increasing, we have $e' < \widetilde{e}_R^*$. This implies that $-C(e') + C(\widetilde{e}_R^*) > 0$.

As a final step, since we have shown that $W_R(e') > \widetilde{W}_R^*$ 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

$$W_R^* \geq W_R(e') > \widetilde{W}_R^*.$$

□

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).[33]

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

Before we prove the parts of the proposition, we formally state the optimization problem and the conditions for its solution.

$$\max_{\substack{e \geq 0, f \geq 0 \\ T \in \{H,R\}}} \quad W_T(e) := (1-a) \cdot \left[\int_{p(e) \cdot f}^{\infty} (\pi g - h) G'(g) dg \right] + a \cdot \left[\int_{g_T^n}^{\infty} (\pi g - h) G'(g) dg \right] - C(e),$$
$$\text{subject to } f \leq F.$$

Recall that the lower bound of the second interval g_T^n stands for the gain of the indifferent naïve offender, which is $g_R^n = p(e)f$ in regime R and $g_H^n = p(\hat{e})f$ in regime H, respectively. Formally, we have a constrained optimization problem with three constraints, $e \geq 0$, $f \geq 0$, and $f \leq F$. In our setting, however, the first and the second constraint are never binding. Due to the Inada condition on the cost function, zero effort is not optimal for any positive fine. Similarly, a fine of zero is not optimal for any positive effort since a slightly higher fine increases deterrence at zero cost. Finally, the case of $e = f = 0$ can be excluded since a simultaneous slight increase of both variables e and f increases deterrence. Therefore, we can focus the analysis on the crucial constraint $f \leq F$, which might or might not be binding in the optimum.[34]

The Lagrangian function then is $\partial L(e, f, \lambda) = W_T(e, f) + \lambda[F - f]$. The Kuhn-Tucker

[33] 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.

[34] This leads to a simpler exposition but to the same results as the constrained optimization problem that includes all three constraints explicitly.

conditions are

$$\frac{\partial L(e,f,\lambda)}{\partial e} = \frac{\partial W_T(e,f)}{\partial e} = 0 \tag{A.3}$$

$$\frac{\partial L(e,f,\lambda)}{\partial f} = \frac{\partial W_T(e,f)}{\partial f} - \lambda = 0 \tag{A.4}$$

$$\lambda[F-f] = 0 \tag{A.5}$$

$$\lambda \geq 0 \tag{A.6}$$

$$F-f \geq 0 \tag{A.7}$$

We distinguish two cases: Either $f < F$ (case 1: interior fine) or $f = F$ (case 2: maximum fine).

Case 1 (interior fine): We first search for a solution (e^*, f^*, λ^*) that satisfies $f^* < F$. By (A.5) this implies $\lambda^* = 0$. Using this in (A.4) yields the condition $\frac{\partial W_T(e,f)}{\partial f} = 0$, which together with (A.3) yields the first order conditions of the maximization problem if it were unconstrained. Since no constraint is binding, the optimal point (e^*, f^*) is the same as it would be in an unconstrained optimization problem. The Hessian matrix is

$$H = \begin{pmatrix} \frac{\partial^2 W_T(e,f)}{\partial e \partial e} & \frac{\partial^2 W_T(e,f)}{\partial e \partial f} \\ \frac{\partial^2 W_T(e,f)}{\partial f \partial e} & \frac{\partial^2 W_T(e,f)}{\partial f \partial f} \end{pmatrix}.$$

A well-known sufficient condition for a local maximum at the point (e^*, f^*) is that principal minors of the Hessian matrix at this point alternate in sign, beginning with a negative sign (see e.g. Jehle and Reny, 2001, pp. 480). The first principal minor of the Hessian is $D_1 = |\frac{\partial^2 W_T(e,f)}{\partial^2 e}| < 0$, which is local concavity at the optimal effort.

The second (and final) principal minor is $D_2 = |H| = \frac{\partial^2 W_T(e,f)}{\partial^2 e} \cdot \frac{\partial^2 W_T(e,f)}{\partial^2 f} - \left(\frac{\partial^2 W_T(e,f)}{\partial ef}\right)^2 > 0$. For this to hold, we need first of all $\frac{\partial^2 W_T(e,f)}{\partial^2 e} < 0$ and $\frac{\partial^2 W_T(e,f)}{\partial^2 f} < 0$, which means local concavity in e and f, and furthermore the standard requirement that the direct effect (i.e. the first two terms) dominates the indirect effect (i.e. the third term).

Case 2 (maximal fine): We now search for a solution (e^*, f^*, λ^*) with $f^* = F$. The first order conditions are given by Eq. (A.3)–(A.5). For the second order conditions, we consider the bordered Hessian matrix $\bar{H}$.

$$\bar{H} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & \frac{\partial^2 W_T(e,f)}{\partial e \partial e} & \frac{\partial^2 W_T(e,f)}{\partial e \partial f} \\ -1 & \frac{\partial^2 W_T(e,f)}{\partial f \partial e} & \frac{\partial^2 W_T(e,f)}{\partial f \partial f} \end{pmatrix}$$

Since we have only two variables and one binding constraint, a sufficient condition for a local maximum is that the determinant of the bordered Hessian is positive, i.e. $|\bar{H}| > 0$

must hold at (e^*, f^*, λ^*) (see e.g. Jehle and Reny, 2001, pp. 496). In our optimization problem, this yields $|\bar{H}| = -\frac{\partial^2 W_T(e,f)}{\partial^2 e} > 0$. Hence, the second order conditions boil down to local concavity of the surplus function at the optimal effort.

Now, we turn to the parts of the proposition, which are to prove.

Part (i): In regime R, the optimization problem is

$$\max_{e \geq 0, f \in [0,F]} W_R(e,f) = \int_{p(e)f}^{\infty} (\pi g - h) G'(g) dg - C(e). \tag{A.8}$$

We first observe that there is no over-enforcement in the optimum, i.e. $\pi p(e^*) f^* < h$. Suppose to the contrary that $\pi p(e^*) f^* \geq h$. Then a slight reduction of e^*, while keeping f^* constant, would weakly increase social benefits and strictly decrease the costs.

This observation, together with Equation (A.4) yields for any candidate equilibrium (e^*, f^*, λ^*),

$$\lambda^* = \frac{\partial W_R(e,f)}{\partial f} = [(h - \pi p(e) f) \cdot G'(p(e) f) \cdot p(e)] > 0.$$

Furthermore, $\lambda^* > 0$ implies $f^* = F$ by equation (A.5). Hence, in regime R the maximal fine is optimal. The optimal effort is determined by (A.3), which yields

$$[(h - \pi p(e) F) \cdot G'(p(e) F) \cdot p'(e) F] = C'(e). \tag{A.9}$$

Finally, observe that this condition coincides with Eq. (5) that determines the optimal effort in regime R of the baseline model, when we set $\bar{f} \equiv F$.

Part (ii): In regime H, the optimization problem is

$$\max_{e \geq 0, f \in [0,F]} W_H(e,f) = (1-a) \cdot \left[\int_{p(e)f}^{\infty} (\pi g - h) G'(g) dg \right] + a \cdot \left[\int_{p(\hat{e})f}^{\infty} (\pi g - h) G'(g) dg \right] - C(e). \tag{A.10}$$

Eq. (A.4) yields for any candidate equilibrium (e^*, f^*, λ^*),

$$\lambda^* = (1-a) \left[(h - \pi p(e) f) \cdot G'(p(e) f) \cdot p(e) \right] + a \left[(h - \pi p(\hat{e}) f) \cdot G'(p(\hat{e}) f) \cdot p(\hat{e}) \right]. \tag{A.11}$$

Observe that the RHS of Eq. (A.11) is strictly positive for $f = 0$ and, by continuity, it remains strictly positive for f small enough. Hence supposing that the maximal fine F is not too large, yields $\lambda^* > 0$. Further, $\lambda^* > 0$ implies $f^* = F$ by Eq. (A.5). Hence, the maximal fine is optimal. The optimal effort e_H^* is then implicitly given by condition (A.3) which yields

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

Finally, 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). Suppose now that the maximal fine F is large enough. Observe that the RHS of Eq. (A.11) becomes negative for f sufficiently large. (Indeed the first summand of (A.11) is bounded from above by $(1-0)(h-0)\cdot G'(\cdot)\cdot 1$ by $G'(\cdot) < \infty$; while the second summand becomes negative and arbitrarily large in absolute terms for large f as the term $-\pi p(\hat{e})f < 0$ dominates then.) Hence, $f = F$ would imply $\lambda < 0$ by Eq. (A.11) and thus contradict condition (A.6). Thus, there is no solution where the maximal fine is chosen in this case.

Searching for a solution with $f^* < F$ means searching for a solution with $\lambda^* = 0$ by (A.5). Hence, the RHS of of Eq. (A.11) has to be zero. Indeed, for F sufficiently large, there exists a combination of f and e that renders it zero, as we show next. For small f both summands on the RHS are positive. For any level of deterrence of sophisticates, $p(e)f$, we can now simultaneously increase f and decrease e such that their level of deterrence stays constant. Thereby, the first summand of (A.11) continuously decreases by this move, but stays positive since $(h - \pi p(e)f)$ and $G'(p(e)f)$ have not changed, while $p(e)$ has decreased. The second summand, also continuously decreases by simultaneously increasing f and decreasing e. It becomes negative at some point (precisely at $f = \frac{h}{\pi p(\hat{e})}$) and at some further point equals the first summand in absolute terms such that $\lambda = 0$.

Therefore, the candidate solution (e^*, f^*, λ^*) to the maximization problem (A.10) satisfies $\lambda^* = 0$ and the two first order conditions (A.3) and (A.4), which by Leibniz' rule become:

$$(1-a)\left[(h - \pi p(e)f)\cdot G'(p(e)f)\cdot p'(e)f\right] = C'(e), \tag{A.12}$$

$$(1-a)\left[(h - \pi p(e)f)\cdot G'(p(e)f)\cdot p(e)\right] = a\left[(\pi p(\hat{e})f - h)\cdot G'(p(\hat{e})f)\cdot p(\hat{e})\right]. \tag{A.13}$$

The RHS of Eq. (A.12), $C'(e)$, is positive for $e = e^*_H > 0$. Hence its LHS must also be positive. The LHS of Eq. (A.12) is only positive if $\pi p(e)f < h$. Now, observe that $\pi p(e)f < h$ implies that the LHS of Eq. (A.13) is also positive. In turn, the RHS of Eq. (A.13) 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^b_R and e^b_H the respective optimal efforts in regime R and in regime H. By assumption, $W_R(e^b_R, \bar{f}) < W_H(e^b_H, \bar{f})$ 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 $\max_{e>0, f\in[0,F=\bar{f}]} W_R(e,f) = W_R(e_R^b, \bar{f})$. Hence,

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

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. (9) in the heterogeneity extension. Observe that the optimal effort in both Eq. (2) and Eq. (9) 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

$$(9) - (2) = \sum_{l=1}^{L} a_l \cdot \left[\int_{p(\hat{e}_l)f}^{\infty} (\pi g - h) G'(g) dg\right] - a \cdot \left[\int_{p(\hat{e})f}^{\infty} (\pi g - h) G'(g) dg\right].$$

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 (9), 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. □

B Robustness: Endogenous Perceptions of Naïves

The decision whether or not to commit an offense is one under uncertainty, as it involves the risk of being apprehended. In our model, sophisticated offenders are assumed to correctly assess this risk, as determined by the actual enforcement effort, and hence they behave as in the standard Becker (1968) approach. By contrast, when the actual enforcement effort remains hidden, the risk perception of naïves is not fully driven by this effort, but also depends on other factors. In this section we motivate our modeling assumptions of naïveté by a very common bias in probabilistic judgment: *Under-inference*. An individual with under-inference bias (also known as *conservatism bias*) does not update sufficiently from a new piece of information. In their classical survey, Peterson and Beach (1967) describe this bias as follows: "when statistical man and subjects start with the same prior probabilities for two population proportions, subjects revise their probabilities in the same direction but not as much as statistical man does[.]"

From an empirical point of view, it is well known that not all people draw the right inferences when it comes to handling probabilities (see e.g. Tversky and Kahneman, 1974). In this respect, under-inference is a particularly common and strong deviation from Bayesian updating. For example, when summarizing the empirical literature on how real people update probabilities, Benjamin (2019) concludes as "Stylized Fact 1. Under-inference is by far the dominant direction of bias" and documents it by conducting a meta-analysis of several experiments and also by surveying additional experiments. In our context, an offender that exhibits under-inference bias will not fully respond to the actually chosen effort and its implied probability of apprehension, but rather will put too much weight on his prior belief about it. In the main text, we have made the simplifying assumption that naïves rely fully on their prior beliefs, so that their perceived risk of apprehension is independent of the actual enforcement effort. In this section, we show that our results also extend to the case where the actual effort does affect the risk perceptions of naïves, as long as this effect is not too strong.

Assuming that some agents suffer from this error in probabilistic judgment, motivates our assumptions on the difference between naïves and sophisticates and in particular on the behavior of naïves. In our model, a "statistical man" has an accurate (posterior) belief about the probability of apprehension $p(e)$ and is called *sophisticated*. By contrast, a *naive* agent's posterior belief, denoted by $\hat{p}(e)$, is a convex combination between the (exogenous) prior belief $\bar{p} \in [0, 1]$ and the actual probability of apprehension $p(e)$, i.e.

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

where $\delta \in [0, 1)$ is the weight placed on the actual probability $p(e)$. Hence, under-inference is here expressed by $\delta < 1$ and it means that a naïve agent will not fully learn the actual detection probability.[35] The stronger the bias, the smaller the parameter δ. For $\delta = 0$, our baseline model is nested in this more general specification simply by setting the prior belief equal to the naïves' perceived probability of apprehension, i.e. $\bar{p} = p(\hat{e})$. It now remains to show that our results extend to this more general model, as long as under-inference is sufficiently strong.[36]

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. In full analogy to Assumption 1 above, we now assume:

Assumption B.1. *The prior belief $\bar{p}$ satisfies $\pi\bar{p}f < h$.*

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

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

[35] Anwar and Loughran (2011) also use a Bayesian approach to predict the perceived probability of apprehension of criminals. In their data, they find belief updating that is consistent with Bayesian learning, in particular, the direction of revision of the belief is as predicted. Note that these findings are consistent with both our model of sophisticates who are Bayesian and our model of naïves who update in the same direction as Bayesian, but not as far as sophisticates.

[36] Similarly, we could assume that the perceived effort $\hat{e}$ is a convex combination between some exogenous component $\bar{e}$ and the actual effort e. As we have shown in an earlier working paper version, our results do extend to such a generalization, given that the weight on the exogenous component is sufficiently large.

with $\hat{p}(e) = \delta p(e) + (1-\delta)\bar{p}$. We denote the maximizer of surplus function (B.2) by $e^*_H(a)$ and the resulting maximum surplus by $W^*_H(a, \bar{p})$. Importantly, 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

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

Intuitively, $e^*_H(a)$ equates the marginal benefit of deterring more sophisticates *and* naïves (LHS) with the marginal cost of deterrence (RHS). Observe that deterrence of naïves is less effective than deterrence of sophisticates in the sense that $\hat{p}'(e) = \delta p'(e) < p'(e)$. For further reference, this first order condition can be rewritten as

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

where $b(\delta) := (h - \pi(\delta p(e) + (1-\delta)\bar{p})f) \cdot G'(\delta p(e) + (1-\delta)\bar{p}f) \cdot \delta \cdot p'(e) \cdot f$.

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 under-inferrence is sufficiently strong, i.e. δ sufficiently small, then the (weighed) gain of the indifferent sophisticated and naïve offender resulting under the optimal enforcement effort is below the social harm, i.e. $\pi p(e^*_H(a))f < h$ and $\pi \hat{p}(e^*_H(a))f < h$, as well as $\pi p(e^*_R)f < h$.*

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) $h - \pi \hat{p}(e)f > 0$ (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. Naïves were weakly over-deterred if $\pi \hat{p}(e)f \geq h$. By definition of $\hat{p}(e)$, this is equivalent to $\pi\delta p(e)f + \pi(1-\delta)\bar{p}f \geq h$. Recalling that $\pi\bar{p}f < h$ by Assumption B.1 and that $\delta \in [0, 1)$ for naïves, a necessary condition for $\pi\hat{p}(e)f \geq h$ (i.e. over-deterrence of naïves) is $\pi p(e)f > h$ (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 $\pi\hat{p}(e)f < h$. 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 $b(\delta)$ as introduced in condition (B.4) above. Note first that no over-deterrence of naïves (i.e. $\pi\hat{p}(e^*_H(a))f < h$) implies that $b(\delta) \geq 0$ for all $\delta \in [0,1]$. Moreover, for shrinking δ it converges to zero: $\lim_{\delta\to 0} b(\delta) = (h - \pi\bar{p}f)\cdot G'(\bar{p}f)\cdot\delta\cdot p'(e)\cdot f = 0$ (since $G'(\cdot) < \infty$ and $p'(\cdot) < \infty$). Hence, in condition (B.4), since $C'(e) > 0$, there always exist some $\bar{\delta} > 0$ such that the RHS is strictly positive for all $\delta \in [0,\bar{\delta}]$. This implies that for all $\delta \in [0,\bar{\delta}]$, 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 naïve offenders under-infer the actual probability of apprehension. Lemma B.1 uses Assumption B.1, which is analogous to Assumption 1 in the basic model, and additionally assumes that the weight δ on the actual probability is not too large. To see why an additional restriction is needed, suppose that $\bar{p}$ is low, so that substantial deterrence of naïves requires large effort e, and that such an effort is affordable at relatively low costs $C(e)$. When the share of naïves a is high, the authority may then choose a very high effort and even tolerate that sophisticates are over-deterred. This is excluded if the weight on the prior belief 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 also hold for the augmented model, provided that the prior belief ($\bar{p}$) gets sufficient weight in the determination of the posterior belief of naïves:

Proposition B.1. ***(Optimal Policy under Regime*** H*) Let Assumption B.1 hold and let the naïves' under-inference be sufficiently strong, i.e.* δ *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* $\delta = 0$, *then* $e^*_H(1) = 0$.[37]

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

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

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

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

(v) When the prior belief $\bar{p}$ is sufficiently small, i.e. $\bar{p} < p(e_H^*(1))$, the social welfare under hiding $W_H^*(a, \bar{p})$ 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 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

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

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 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 negative (such that they become positive when applying the negative sign that is before them). This is the case if and only if only if

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

where we have used $\hat{p}'(e_H^*(a)) = \delta \cdot p'(e_H^*(a))$. The RHS of inequality (B.5) can be substituted by $b(\delta)$ (see condition B.4 above). As shown in the proof of Lemma B.1,

$\lim_{\delta\to 0} b(\delta) = 0$. 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 δ 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 $\delta = 0$ (such that there are only naïves and their posterior belief is unaffected by the chosen effort). In this special case, any positive effort $e > 0$ is suboptimal since there is no benefit but only costs of enforcement; thus, we have $e_H^*(1) = 0$.

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

$$\begin{aligned}\frac{\partial W_H^*}{\partial \bar{p}} &= -a(\pi(\delta p(e) + (1-\delta)\bar{p})f - h) \cdot G'((\delta p(e) + (1-\delta)\bar{p})f) \cdot (1-\delta) \cdot f \\ &= a(h - \pi\hat{p}(e)f) \cdot G'(\hat{p}(e)f) \cdot \hat{p}'(e) \cdot f\end{aligned}$$

which is strictly positive by Lemma B.1.

(iii-v) Using the envelope theorem and taking the derivative of $W_H^*(a, \bar{p})$ w.r.t. a yields

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

the sign of which is solely determined by comparing the two respective lower bounds of the integrals. By Lemma B.1, we have $\pi p(e_H^*(a))f - h < 0$ and $\pi\hat{p}(e_H^*(a))f - h < 0$ such that the first integral is bigger (in absolute terms) than the second one if and only if $p(e_H^*(a))f < \hat{p}(e_H^*(a))f$. Hence, $W_H^*(a, \bar{p})$ is increasing in a if and only if $p(e_H^*(a)) < \bar{p}$, and they are identical for $p(e_H^*(a)) = \bar{p}$.

Part (iii) supposes that $\bar{p} > p(e_H^*(0))$. Proposition B.1 above shows that $e_H^*(a)$ is decreasing. Hence, we have $\bar{p} > p(e_H^*(a))$ for all $a \in [0,1]$ in this case. Thus, $W_H^*(a, \bar{p})$ is strictly monotone increasing in a. Part (iv) supposes that $p(e_H^*(1)) < \bar{p} < p(e_H^*(0))$. Since $p(e_H^*(a))$ is continuously decreasing in a, there must be a threshold $\hat{a}(\bar{p})$ such that $p(e_H^*(\hat{a})) = \bar{p}$. $W_H^*(a, \bar{p})$ is strictly decreasing in a for $a < \hat{a}(\bar{p})$ and strictly increasing when the inequality is reversed. Part (v) finally supposes that $\bar{p} < p(e_H^*(1))$. Since $p(e_H^*(a))$ is decreasing in a, we have $\bar{p} < p(e_H^*(a))$ for all $a \in [0,1]$. Thus, $W_H^*(a, \bar{p})$ is strictly monotone decreasing in a.

□

Hence, as long as δ 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 $\delta > 0$. 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 probability of apprehension $p(e_H^*(a))$ might always lie below the naïves' posterior belief $\hat{p}(e_H^*(a))$, or, equivalently, below their prior belief $\bar{p}$. 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**) Let Assumption B.1 hold and let the naïves' under-inference be sufficiently strong, i.e. δ sufficiently small.*

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

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

(iii) If $\bar{p} < p(e_H^*(0))$ and $W_H^*(1, \bar{p}) < W_R^*$ 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 prior belief $\bar{p}$ is in the intermediate range, i.e. $p(e_H^*(1)) < \bar{p} < p(e_H^*(0))$. 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 prior belief $\bar{p}$ is sufficiently small, i.e. $\bar{p} < p(e_H^*(1))$.

To summarize, when the naïves' under-inference is strong enough, 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.

Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics-Foundations and Applications 2*, 69.

Braga, A., D. Kennedy, E. Waring, and A. M. Piehl (2001): "Problem-oriented policing, deterrence, and youth violence: An evaluation of Boston's Operation Ceasefire," *Journal of Research in Crime and Delinquency*, 38, 195–225.

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.

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

Chamlin, M. and R. Langworthy (1996): “The police, crime, and economic theory: A replication and extension,” *American Journal of Criminal Justice*, 20, 165–182.

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

Cronshaw, M. and J. Alm (1995): “Tax compliance with two-sided uncertainty,” *Public Finance Quarterly*, 23, 139–166.

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

Dur, R. and B. Vollaard (2019): “Salience of law enforcement: A field experiment,” *Journal of Environmental Economics and Management*, 93, 208–220.

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.

Jehle, G. and P. Reny (2001): *Advanced Microeconomic Theory (2nd Edition)*, Addison Wesley.

Kleck, G. and J. Barnes (2014): “Do more police lead to more crime deterrence?” *Crime & Delinquency*, 60, 716–738.

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.

Lang, M. and A. Wambach (2013): “The fog of fraud – mitigating fraud by strategic ambiguity,” *Games and Economic Behavior*, 81, 255–275.

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

Levitt, S. D. and T. J. Miles (2007): “Empirical study of criminal punishment,” in *Handbook of Law and Economics*, ed. by A. M. Polinsky and S. Shavell, Elsevier, vol. 1, 455–495.

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.

Peterson, C. R. and L. R. Beach (1967): “Man as an intuitive statistician.” *Psychological Bulletin*, 68, 29.

Pogarsky, G., A. Piquero, and R. Paternoster (2004): "Modeling change in perceptions about sanction threats: The neglected linkage in deterrence theory," *Journal of Quantitative criminology*, 20, 343–369.

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.

Rosenfeld, R., R. Fornango, and A. Rengifo (2007): "The impact of order-maintenance policing on New York City homicide and robbery rates: 1988-2001," *Criminology*, 45, 355–384.

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

Snow, A. and R. Warren (2005): "Ambiguity about audit probability, tax compliance, and taxpayer welfare," *Economic Inquiry*, 43, 865–871.

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. (2010): "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.