T

Type-I and Type-II Errors

Matteo Rizzolli
LUMSA University, Rome, Italy

Abstract

Adjudicative procedures meant at establishing truth about facts on defendants' behavior are naturally prone to errors: defendants can be found guilty/liable when they truly were not (type-I errors) or they can be acquitted when they should have been convicted (type-II errors). These errors alter the incentives of defendants to comply with norms. We review the literature with a particular focus on type-I errors.

Introduction

The word *adjudication* has its Latin roots into the words *jus* (right, justice) and *dicere* (to say). All organizations set goals and have adjudicative procedures to "establish the truth" about members' compliance. Performance appraisal committees decide whether employees earn rewards within business organizations; teachers must assess students' progress in learning; courts adjudicate whether citizens have committed crimes; disciplinary committees within sport leagues and professional organizations as well as religious tribunals assess whether members' conduct has been conforming. Adjudicative procedures must evaluate and reward something they cannot directly observe – being it effort, intention, or act – and this makes them obviously prone to errors. Although our framing will be mostly applied to criminal (See aslo "▶ Criminal Sanctions and Deterrence Voice and to Crime (Incentive to)"), administrative, and civil courts, most of the results here presented apply to any generic adjudicative procedure. We thus consider the general case of an adjudicative authority who (i) must assess whether the observed behavior of an individual *conforms* or *deviates* from the prescribed behavior and (ii) must incentivize or sanction such behavior accordingly. In judging behavior, errors inevitably arise, and they generally undermine individuals' incentives. These errors take mainly two forms: (i) the adjudicative authority may assess non-compliance when in fact the subject is duly complying and (ii) the adjudicative authority may assess compliance when in fact the subject is deviating. Individual's compliance with the prescribed behavior can be interpreted as the null hypothesis, so that the adjudicative authority can both incorrectly reject the null and sanction a complying subject (a type-I error) and incorrectly accept the null and exculpate an undeserving subject (type-II error). In the context of crime deterrence, type-I errors amount to wrongful convictions of innocents. We model the relation between type-I and type-II errors below within a standard optimal deterrence

A. Marciano, G. B. Ramello (eds.), *Encyclopedia of Law and Economics*,
https://doi.org/10.1007/978-1-4614-7883-6_66-1

framework. Finally, we discuss the empirical relevance of type-I errors.

Basic Setup

Let y_0 be the initial endowment equal for all agents and b the benefits from deviating from the prescribed behavior (e.g., committing crime). b is distributed among the agents with a generic distribution $z(b)$ and a cumulative $Z(b)$ with support $[0\bar{B}]$. Let also h be the harm/externality generated by each individual's deviation (each individual takes this decision only once). For the sake of simplicity, all individuals are audited and brought in front of an adjudicative authority. The authority observes the amount of inculpatory evidence e that is produced against a defendant, and if this overcomes a certain threshold, $\tilde{e}$ then the authority imposes a monetary sanction s. For the sake of simplicity, we also assume that there is no welfare-improving deviation as in Becker seminal paper on crime (See also "▶ Crime and Punishment" by Becker 1968 and also to "Becker, Gary S.") (this would be a crime for which $b > h$) and that monetary sanctions are transferred from the defendant to society. Furthermore, in the function of social costs, we do not consider the private benefits from crime but only its social costs.

Therefore let e have a frequency distribution of $i(e)$ for the conforming defendant (innocent) and of $g(e)$ for the deviating defendant (guilty). Let $I(e)$ and $G(e)$ be the cumulative distributions of $i(e)$ and $g(e)$, respectively, and note that $I(\tilde{e})$ and $G(\tilde{e})$ are the probabilities of being acquitted for the complying and for the deviating defendant, respectively, given the evidence threshold $\tilde{e}$. To keep notation compact, we will often use I and G for $I(\tilde{e})$ and $G(\tilde{e})$, respectively.

The evidence is stochastically distributed, albeit in general more incriminating evidence is available against deviating defendants than against complying ones. First-order stochastic dominance is assumed $I(e) > G(e)\ \forall\ e \in]0, e_{\max}[$. Without FOSD evidence would be produced randomly for the complying and the deviating alike, and therefore the whole criminal procedure would be pointless. Note also that G is the probability of type-II error and $1 - I$ is the probability of type-I error. Let us also define $\Delta(\tilde{e}) = I - G$ as the *accuracy* of the adjudicative procedure; Δ represents the ability of the procedure to distinguish complying from deviating defendants.

For our purpose, we assume the social planner optimizes deterrence only by affecting the threshold $\tilde{e}$ which in turn determines the error's trade-offs: for instance, an increase in $\tilde{e}$ generates both an increase in the number of wrongful acquittals G and a decrease in the number of wrongful convictions $1 - I$.

The risk-neutral individual does not deviate as long as the returns from deviating behavior are smaller than the expected returns of conforming. Since b varies across individuals, there exists a level of $\tilde{b}$ for which the individual is indifferent between conforming and not, and this determines the proportion of the population $Z(\tilde{b})$ who conforms.

Social welfare is thus $(1 - Z(\tilde{b}))h$: the social costs of harm caused by those defendants who deviate. On the other hand, the social planner only acts on the threshold $\tilde{e}$ which implicitly defines the trade-off between type-I and type-II errors. The link between the social planner's choice of the evidentiary standard $\tilde{e}$ which in turn determines the error's trade-off and the defendant's choice of conformity determined in $\tilde{b}$ are the ingredients to understand the role of adjudication in deterrence.

Let us begin by assuming agents to be risk-neutral utility maximizers. The returns from conforming are $E\pi_I = y_0 - (1 - I)s$, while the returns from deviating are $E\pi_G = y_0 + b - (1 - G)s$. All defendants for which $E\pi_i \geq E\pi_g$ will conform and therefore the threshold level of b which implicitly defines the conforming population is

$$\tilde{b}_{rn} = (1 - (1 - I) - G)s \tag{1}$$

By looking at Eq. 1, we can single out the typical "deterrence effect" as $\tilde{b}$ increases both with the magnitude of the sanction $(\uparrow s \Rightarrow \uparrow \tilde{b})$ and via an increase in the detection probability for the deviating defendants which in this model

corresponds to a decrease in the probability of type-II errors $\left(\downarrow G \Rightarrow \uparrow \tilde{b}\right)$. Furthermore, a "compliance effect" of type-I errors can be seen: s $\tilde{b}$ increases when the probability of being punished decreases for conforming defendants $\left(\uparrow I \Rightarrow \uparrow \tilde{b}\right)$. Also a "screening effect" can be established: the higher is the accuracy Δ, the better the procedure can discriminate between conforming and non-conforming behaviors and the greater the advantages of staying conforming ($\uparrow \Delta \Rightarrow \uparrow \tilde{b}$). Finally, by simple inspection of Eq. 1, it is evident that marginal change in either $1 - I$ or G determines an equal decrease of $\tilde{b}$ as $\frac{\partial b}{\partial(1-I)} = \frac{\partial b}{\partial G} = s$. Under risk neutrality, type-I errors $(1 - I)$ and type-II errors (G) have the same negative impact on the defendant's incentive to comply. This is because on one hand type-II errors undermine compliance inasmuch as they decrease the probability of non-conforming defendants being finally sanctioned. On the other hand, type-I errors increase the opportunity costs of conforming relative to deviating.

Now that the threshold level $\tilde{b}$ is defined, the social welfare can be computed and derived with respect to the evidence threshold:

$$\begin{aligned} \frac{\partial SW}{\partial e} &= \partial\left(1 - Z\left(\tilde{b}_{rn}\right)\right)h \\ &= -z\left(\tilde{b}_{rn}\right)(i(\tilde{e}) - g(\tilde{e}))sh \end{aligned} \tag{2}$$

Let $\tilde{e}_{\text{neutral}}$ be implicitly defined by $i(\tilde{e}) = g(\tilde{e})$. By inspection of Eq. 2, the optimal evidence threshold $\tilde{e}$ that minimizes social costs is $\tilde{e}_{\text{neutral}}$. In fact accuracy reaches its maximum level when the social planner chooses $\tilde{e}_{\text{neutral}}$ so that the distance between the two cumulative functions is maximized. If the social planner chooses a higher evidence threshold $\tilde{e}_{\text{pro-def endant}} > \tilde{e}_{\text{neutral}}$, then the error trade-off tilts in favor of the defendant as the probabilities of both correct and wrongful acquittals $-I$ and G, respectively, increase. $\tilde{e} > \tilde{e}_{\text{neutral}}$ necessarily also implies $g(\tilde{e}) > i(\tilde{e})$ by definition of the frequency distribution of $i(e)$ and $g(e)$. Notice that for levels of $\tilde{e} > \tilde{e}_{\text{neutral}}$, G grows faster than I and therefore accuracy cannot be maximal.

Evidence Thresholds, Standard of Evidence, and Error Ratios

While our analysis focuses on the evidentiary threshold $\tilde{e}$ that determines the probabilities of both type-I and type-II errors, there are other two common concepts that concern adjudication and that must be put in relation with our analysis.

The first one is the **standard of evidence**: it is generally understood as the level of certainty the adjudicative authority must reach in order to establish guilt in a criminal proceeding (or liability in civil one). Among the most common standards of proof used in different adjudicative procedures, there are the *preponderance of evidence* (*poe*) standard, the *clear and convincing evidence* (*cace*) standard, and the *beyond any reasonable doubt* (*bard*) standard. Although giving probabilistic interpretations of these standards of proof is very controversial (see Kaplow 2012, footnote 76 for a discussion), they are commonly understood to roughly coincide with the 50%, 75%, and 95% thresholds, respectively. Under *poe* (or *cace* or *bard*), the adjudicative authority must answer to the question of whether, given the evidence available, the likelihood that the defendant has deviated is larger than 50% (or 75% or 95% depending on the standard applied). These probabilities must be understood as Bayesian posterior probabilities of having deviated, and these are functions – following the Bayes' rule – of the likelihood of the signal given by the densities i and g of the evidentiary threshold $\tilde{e}$ and on the prior probability of being brought in front of the adjudicative authority. The probability that a defendant has deviated or not also depends on the base rates of the two actions, $\left(1 - Z\left(\tilde{b}\right)\right)$ and $\left(Z\left(\tilde{b}\right)\right)$, respectively. The $\tilde{b}$ are determined endogenously by defendants' decisions and ultimately depend on the evidence threshold $\tilde{e}$. Therefore in order to identify the proper threshold $\tilde{e}$ – in case the *poe* standard applies – one should ask what value of $\tilde{e}$ implicitly solves the equation $g(\tilde{e}) \cdot \left(1 - Z\left(\tilde{b}\right)\right) = i(\tilde{e}) \cdot \left(Z\left(\tilde{b}\right)\right)$. If the adjudicative authority needs to apply the *cace* or *bard* standard, one could simply multiply by either 3 or 19 the right side of the previous equation. As Lando

(2002) and Kaplow (2012) point out, the two notions – the one based on the optimal *evidence threshold* and the one based on the *standard of evidence* – are strikingly different. To begin with, the optimal evidence threshold is derived from welfare analysis and seeks to find the level of $\tilde{e}$ that maximizes social welfare. By contrast, within the *standard of evidence* framework, $\tilde{e}$ is derived by asking under what circumstances would the probability that the defendant before the adjudicative authority has actually deviated be 50% (or 75% or 95% or other conventional probabilities). In fact the optimal *evidence threshold* could be associated with any probabilistic *standard of evidence* whatsoever.

Another approach focuses on the **ratio of errors** and expresses the pro-defendant bias of adjudicative procedures in terms of error ratios. There seems to be something specific about type-I errors in the context of crime: scholars and rule makers across time and societies advocated a specific attention to the avoidance of type-I errors even at the expense of many type-II errors; arguably the most famous statement in this respect is the one of William Blackstone (1769) recommending that *it is better that ten guilty persons escape than that one innocent suffer.* Dekay (1996) systematizes the relation between the standards of evidence and the error ratios. We can interpret these as ratios of errors' frequencies where the frequency of erroneous acquittals is the conditional probability that a truly deviating defendant is acquitted (type-II error) multiplied by the base rate of the action $\left(1 - Z\left(\tilde{b}\right)\right)$, while the frequency of erroneous conviction is the conditional probability that a truly complying defendant is convicted (type-I) multiplied by the base rate $\left(Z\left(\tilde{b}\right)\right)$. So the type-I error ratio (sometimes also called the Blackstone's error ratio) is defined as $\frac{G\cdot\left(1-Z\left(\tilde{b}\right)\right)}{(1-I)\cdot Z\left(\tilde{b}\right)}$.

All else being equal, higher standards of evidence that affect the trade-off between G and I do imply higher Blackstone-like error ratios. However, it should be noticed that the optimal *evidence threshold* could be associated with many different error ratios depending on the base rates.

Risk and Loss Aversion

Subjects are known to be generally risk-averse in their utility of income. We thus introduce risk aversion in the measure of the monetary gains from crime b following Rizzolli and Stanca (2012). If b are monetary gains for which utility $U(.)$ can be derived, then the expected utility of complying is $IU(y_0) + (1 - I) \cdot U(y_0 - s)$, while the expected utility of deviating is $GU(y_0 + b) + (1 - G) \cdot U(y_0 + b - s)$. The threshold level of $\tilde{b}_{eu}$ that triggers a defendant to deviate is implicitly defined by

$$\begin{aligned} &I[U(y_0) - U(y_0 - s)] \\ &\quad - G[U(y_0 + b) - U(y_0 + b - s)] \\ &\quad \geq U(y_0 + b - s) - U(y_0 - s) \end{aligned} \tag{3}$$

Equation 3 shows that when there is an increase in either of the errors (increase in G or decrease in i) on the left-hand side of the equation, defendants find deviation convenient for lower levels of b (on the right-hand side). However, given the concavity of the utility function, the negative impact of type-I errors $(1 - I)$ on the threshold level $\tilde{b}_{eu}$ and thus on social welfare is stronger than that of type-II errors (G). To see why, note that $U(y_0) - U(y_0 - s) > U(y_0 + b) - U(y_0 + b - s)$. In order to maintain the same level of deterrence, a given percentage increase of $1 - I$ must be compensated by a smaller percentage decrease of G. Therefore, assuming risk aversion, type-I errors $(1 - I)$ create more disutility and thus induce more deviation than comparable type-II errors (G); therefore, social costs are minimized for a $\tilde{e}^* > \tilde{e}_{\text{neutral}}$. The opposite result holds if we instead assume risk-seeking behavior.

Another interesting extension concerns the introduction of loss aversion: a departure from the expected utility framework that has been incorporated in models such as the cumulative prospect theory (Dhami and al Nowaihi 2013). These models build on the empirical observation that people tend to think of possible outcomes of a choice under uncertainty relative to a certain reference point and tend to prefer the avoidance of losses (outcomes below the reference point) than the acquisition of comparable gains (outcomes

above the reference point). Incorporating reference-dependent preferences and loss aversion in the model is not trivial (see Nicita and Rizzolli 2014); however, the intuition and the results are quite simple: type-I errors always imply a potential loss relative to the status quo, while this is not necessarily true for type-II errors. To conclude, in presence of loss aversion, type-I errors $(1 - I)$ represent a net loss and impact the defendant value function more than comparable type-II errors (G); therefore, social costs are minimized for a $\tilde{e}^* > \tilde{e}_{\text{neutral}}$.

Cost of Sanctions

So far we have assumed that the sanction s is monetary and that it implies – once imposed – a costless transfer from the defendant to the society. However, the imposition of sanctions implies both private costs of punishment to defendants and to society as well. Nonmonetary sanctions are a social cost (Shavell 1987) as their imposition implies a disutility for the defendant that is not transferred to society. Furthermore, all sanctions – including monetary fines – must be administered and therefore imply a social cost (Polinsky and Shavell 1992).

Define c as the total cost (both to the defendant and to the society) of imposing a sanction. The social welfare function (assuming risk neutrality) should be rewritten as the following:

$$SW = \left[1 - Z\left(\tilde{b}\right)\right]h + \left[1 - Z\left(\tilde{b}\right)\right](1 - G)c + Z\left(\tilde{b}\right)(1 - I)c \tag{4}$$

The first term of Eq. 4 represents the harm/externality of deviating, as before. The second term represents the expected total costs of imposing sanctions on deviating defendants, and the third term represents the expected total costs of punishing complying defendant (type-I errors). The problem lies in defining the optimal $\tilde{e}$ that minimizes the expected total costs from crime, including the costs of punishment. As before, the first term is minimized for $\tilde{e} = \tilde{e}_{\text{neutral}}$. However, the second and third terms are minimized for $\tilde{e} \to \infty$. In fact for an evidence threshold $\tilde{e}$ arbitrarily high, the probability of correctly imposing a sanction on a deviating defendant $(1 - G)$ or erroneously imposing a sanction on a complying defendant $(1 - I)$ decreases to zero and – since nobody is punished – there are no costs of punishment for society. When social costs are considered, the costs of harm implied by the first term must thus be balanced against the costs of punishment of the second and third term. Therefore, in the presence of costs of punishment, the social costs of harm must be weighted against the social costs of punishment, and therefore social costs are minimized for $\tilde{e}^* > \tilde{e}_{\text{neutral}}$. This result is based on Rizzolli and Saraceno (2013).

Identity Errors

Lando (2006) introduced a distinction between *mistakes of act* and *mistakes of identity*. *Mistakes of act* happen when a defendant is judged deviating when in fact he was complying. These are adjudicative errors we have been focusing on so far, for which the main concern of the adjudicative authority is whether there actually was any deviation at all. Note that, in case of mistakes of act, the two errors are independent: an increase in wrongful convictions does not imply any change in the number of wrongful acquittals. Then there are *mistakes of identity*, by which in the presence of deviations that can be easily observed, such as a murder or a robbery in the context of crime, the wrong person can be incriminated for an act that actually did happen. These are the cases where the occurrence of the deviation cannot be denied and the authority is concerned with who committed the crime. Note that in this case the two errors for a given crime are linked, as the conviction of an innocent person implies the acquittal of the person actually responsible for it.

Suppose that at time t_1 there exists an exogenous probability $\beta_{i,g}$ that an defendant is sanctioned for a deviation that has already happened at t_0 and which the subject is not responsible for (a mistake of identity). This exogenous probability can vary depending on the decision of the defendant at t_1: it seems reasonable to assume that abstaining from a crime at t_1 reduces the

probability of a mistake of identity, so that $\beta_i \leq \beta_g$. Thus the returns from conforming at t_1 are $E\pi_I = y_0 - (1 - I)s - \beta_i s$, while the returns from deviating are $E\pi_G = y_0 + b - (1 - G)s - \beta_g s$. The threshold level of b which implicitly defines the conforming population is

$$\tilde{b}_{\text{identity}} = (1 - (1 - I) - G)s - (\beta_i - \beta_g)s \quad (5)$$

Inspection of Eq. 5 and comparison with Eq. 1 highlight the role of mistakes of identity vis-à-vis deterrence implicitly defined by $\tilde{b}_{\text{identity}}$. The first part is equal to Eq. 1, while in the second part, if $\beta_i = \beta_g$ as Lando (2006) hypothesized, then identity errors occurred at t_0 have no impact on deterrence at t_1. However, if $\beta_i < \beta_g$, then identity errors actually have a positive impact on deterrence. The reason is intuitive: the decision to deviate in t_1 triggers a net increase in the probability of being wrongfully convicted because of a mistake of identity. Of course this result is based on the assumption that the probability of identity errors in t_1 is determined exogenously, and it is not a function of $\tilde{e}$. Furthermore, identity errors impose a necessarily constraint between the input of wrongful acquittals and the output of wrongful identity convictions; Garoupa and Rizzolli (2013) show that once this constraint is considered, mistakes of identity have a net negative impact on deterrence.

Errors and the Precaution of Harm

Another main area where the role of adjudicative errors has been explored is tort law (see Png 1986; Lando and Mungan 2017, among others). The standard model of tort law substitutes the dichotomous choice between complying and deviating with a continuous choice about the level of activity/care. We will discuss the main implications below. However, some novel conclusions can be drawn also from applying the dichotomous choice model. In this framework, the defendant chooses between *conforming* to the prescribed standard of care or *deviating* and not taking any precaution. Since taking precautions is costly, we can interpret b as the opportunity cost of conforming (by deviating, the defendant saves b). Furthermore, the sanction is equal to the harm inflicted since the goal of the tort system is compensation, and the decision to conform only reduces the expected harm: when conforming, harm h_i is produced with probability α_i, while when deviating, harm h_g is produced with probability α_g, where $h_g > h_i$ and $\alpha_g > \alpha_i$. Adjudicative errors can occur in the usual way, and therefore, a risk-neutral defendant's returns from conforming are $E\pi_I = y_0 - (1 - I)\alpha_i h_i$, while the returns from deviating are $E\pi_G = y_0 + b - (1 - G)\alpha_g h_g$. All defendants for which $E\pi_i \geq E\pi_g$ will conform and therefore the threshold level of b which implicitly defines the conforming population is

$$\tilde{b}_{\text{care}} = (1 - G) \cdot \alpha_g h_g - (1 - I) \cdot \alpha_i h_i \quad (6)$$

By comparing Eq. 6 with Eq. 1, one immediately realizes that type-I errors have a smaller impact on the incentive to comply than type-II errors as $\frac{\partial b}{\partial(1-I)} = \alpha_i h_i < \frac{\partial b}{\partial G} = \alpha_g h_g$; this is because complying causes a smaller expected harm. Also social welfare changes as now also complying defendant causes harm. To find out the optimal $\tilde{e}$, we compute $\frac{\partial SW}{\partial e} = 0$ and thus

$$\begin{aligned} &\partial Z(\tilde{b})\alpha_i h_i + \partial(1 - Z(\tilde{b}))\alpha_g h_g \\ &= (\alpha_i h_i i(\tilde{e}) - \alpha_g h_g g(\tilde{e}))(\alpha_i h_i - \alpha_g h_g) = 0 \end{aligned} \quad (7)$$

Rearranging Eq. 7, we have that $i(\tilde{e}) = \frac{\alpha_g h_g}{\alpha_i h_i} g(\tilde{e})$, and since $\frac{\alpha_g h_g}{\alpha_i h_i} > 1$, the equality can be satisfied only for $\tilde{e}^* < \tilde{e}_{\text{neutral}}$. We can thus conclude that when defendants face a dichotomous choice between complying and causing a smaller expected harm and deviating and causing a larger expected harm, type-I errors impact deterrence less than type-II errors, and therefore welfare is maximized for a level of evidentiary standard smaller than the neutral one.

Precautionary Activities and Chilling of Desirable Behavior

In both the crime and the tort contexts, the choice of deviating causes social harm at least in expected terms. Compliance causes no harm in

the crime context, while it produces a smaller social harm in the tort context. In many situations, however, defendant compliance can have both harmful consequences and benign ones. One may think of the case of competition policy, where the threat of antitrust sanctions may discourage efficient, pro-competitive behavior; another case may be medical malpractice, where worries about false positives may prevent cost-effective care. Kaplow's (2011) model envisages a population that can engage in a harmful act that produces a private benefit as well as a negative externality and another population that can only engage in a benign act that produces no externality. The two types of act are initially indistinguishable to the authority, but the adjudicative procedure gives rise to an evidence signal e that is higher for harmful acts than for benign ones. As before, the authority sanctions subjects whose acts produce an evidence signal higher than a certain cutoff value $\tilde{e}$. However, now the expected sanction raises both the costs of the harmful act and that of the benign one, thus chilling desirable behavior. Kaplow (2011) shows that the optimal $\tilde{e}$ that equates the (falling) marginal benefits of deterring harmful acts with the (rising) marginal costs of chilling benign acts is such that $\tilde{e}^* > \tilde{e}_{\text{neutral}}$. Intuitively it is generally optimal to raise the sanction and simultaneously raise $\tilde{e}$, holding deterrence constant. In fact the only consequence of this policy is a reduction in chilling costs.

A similar model is proposed by Mungan (2011) where subjects can choose between *inaction* (precautionary activity) and *action*, and this second choice can produce *no externality* (desirable behavior) or a *negative externality* (harmful activity). The authority cannot distinguish with certainty whether the activity is harmful or benign but must rely on an evidence signal e and balance the usual errors' trade-off. The expectation of sanctions wrongfully imposed on desirable behavior induces subjects at the margin to switch over to precautionary activities. Again, Mungan (2011) shows that the optimal evidence threshold is such that $\tilde{e}^* > \tilde{e}_{\text{neutral}}$.

Judicial Errors When the Choice of Care Is Continuous

In the model presented so far, the defendant's choice between complying and deviating is dichotomous. However, other situations like torts are best described by a continuous choice of care level x. In the prevailing model of tort, a legal standard x is set in order to determine liability by a potential injurer: the defendant avoids liability if his level of care is equal or above the standard one which is usually equated to the optimal level x^*. Craswell and Calfee (1986) introduce legal errors in this context by proposing a model where such legal standard is uncertain in the sense that defendants who choose a level of care x only know that there is a probability $F(x)$ (decreasing in x) that they will be sanctioned so that choosing higher levels of x decreases the probability to be punished. So if they choose $\underline{x} < x^*$, there is a $1 - F(\underline{x})$ probability of type-II error (the defendant is not made liable even if he took less than the efficient level of care), while if they choose $\bar{x} > x^*$, there is a $F(\bar{x})$ probability of type-I error (the defendant is made liable even if he took enough care). Assuming that also both the sanction s and the opportunity cost of care b are increasing functions of x, Craswell and Calfee (1986) show that with respect to the socially optimal level of x, the defendant's choice of x can be either undercomplying (the defendant chooses $\underline{x} < x^*$) or overcomplying $(\bar{x} > x^*)$. This is because, on one hand, there is always a positive chance $1 - F(x)$ of acquittal, and this increases the returns of taking lower levels of care. But on the other hand, the expected sanction depends on the probability $F(x)$, and this can be driven down by increasing the level of care. The relative impact of these two countervailing effects on the final level of x can be of either sign as it depends on various features of the legal environment and in particular on the amount of uncertainty. Craswell and Calfee (1986) show that under some plausible assumptions concerning the distribution of errors, defendants will usually take an excessive level of care. In other words, while the possibility of escaping

liability when the defendant has not taken enough precaution (type-II error) has the usual adverse effect on the incentives to take precaution, the possibility of being wrongfully held liable even when one has taken enough precautions (type-I error) induces the defendant to increase the level of precautions (under some plausible conditions).

Further Effects on Evidentiary Standards and on Type-I Errors

In addition to the literature survey above, many authors have also explained the high evidence threshold usually observed in legal trials using deterrence-based arguments that point at (i) biased evidence selection (Schrag and Scotchmer 1994), (ii) parties' evidence production expenditure (Yilankaya 2002), (iii) optimal exercise of care by parties (Demougin and Fluet 2006), (iv) marginal deterrence (Ognedal 2005), (v) repeated offenders (Chu et al. 2000), and (vi) emotional costs of indignation (Nicita and Rizzolli 2014).

Furthermore, without making reference to a specific utilitarian approach based on the deterrence rationale, a consistent number of papers simply postulate that wrongful convictions of innocents are morally repugnant and thus inherently worse than type-II errors. Arguments that justify this position are reviewed in Epps (2015). These arguments are mainly deontological and transcend the utilitarian framework (See also "► Retributivism Voice") although they can still be considered in our model by overweighting the impact of type-I errors on the social welfare function along the lines of Miceli (2009).

Empirical Relevance

The role of type-II errors has been greatly analyzed empirically and experimentally; there exists a vast literature testing Becker's deterrence hypothesis with real data on incarceration and on the death penalty (Chalfin and McCrary 2017). There is also a small stream of literature testing the deterrence hypothesis in the lab (see literature cited in Khadjavi 2015); (See also "► Experimental Law and Economics Voice").

Most of this literature, however, ignores type-I errors and their impact on deterrence and behavior. This asymmetry is easy to understand once one considers that type-II errors (crimes that go unpunished) are far more easy to be observed and measured than type-I errors (wrongful convictions can in fact be mistaken for correct convictions). Empirical research on type-I errors has only recently taken off either comparing agreement rates of judges and juries (Gould et al. 2014) or by using DNA testing introduced in the 1990s. Many innocent defendants used DNA testing to clear themselves after conviction whenever biological evidence from the crime scene had been retained. By adopting this strategy, Risinger (2007) estimated the type-I error rate in capital rape-murder cases to be between 3.3% and 5% in 1982–1989. Gross and O'Brien (2008), using post-1973 US data on death sentences (See also "► Death Penalty Voice"), estimated a type-I error frequency of wrongful death sentences to be at least 2.3%. Most of this literature is concerned with measuring the magnitude of type-I errors and less with the assessment of the impact of type-I errors on general deterrence (Gould et al. 2014). A small number of controlled lab experiments try to assess the impact of type-I errors on deterrence: Grechenig et al. (2010) first showed that both errors greatly undermine deterrence in a voluntary contribution mechanism (VCM) type of game. Rizzolli and Stanca (2012) disentangled the effects of each type of error and found that type-I errors are more detrimental to deterrence than type-II errors. Marchegiani et al. (2016) found the same effect within a principal-agent setting, while Markussen et al. (2016) using a VCM design found instead the opposite effect: that type-I errors are less detrimental than type-II errors.

References

Blackstone W (1769) Commentaries on the laws of England, vol 4. Clarendon Press, Oxford

Chalfin A, McCrary J (2017) Criminal deterrence: a review of the literature. J Econ Lit 55:5–48

Chu CC, Hu S-C, Huang T-Y (2000) Punishing repeat offenders more severely. Int Rev Law Econ 20:127–140
Craswell R, Calfee JE (1986) Deterrence and uncertain legal standards. J Law Econ Org 2:279–303
Dekay ML (1996) The difference between Blackstone-like error ratios and probabilistic standards of proof. Law Soc Inq 21:95–132
Demougin D, Fluet C (2006) Preponderance of evidence. Eur Econ Rev 50:963–976
Dhami S, al Nowaihi A (2013) An extension of the Becker proposition to non-expected utility theory. Math Soc Sci 65:10–20
Epps D (2015) The consequences of error in criminal justice. Harv Law Rev 128:1065
Garoupa N, Rizzolli M (2013) Wrongful convictions do lower deterrence. J Inst Theor Econ 168:224–231
Gould JB, Carrano J, Leo RA, Hail-Jares K (2014) Predicting erroneous convictions. Iowa Law Rev 99:471–2299
Grechenig K, Nicklisch A, Thöni C (2010) Punishment despite reasonable doubt – a public goods experiment with sanctions under uncertainty. J Empir Leg Stud 7:847–867
Gross SR, O'Brien B (2008) Frequency and predictors of false conviction: why we know so little, and new data on capital cases. J Empir Leg Stud 5:927–962
Kaplow L (2011) On the optimal burden of proof. J Polit Econ 119:1104–1140
Kaplow L (2012) Burden of proof. Yale Law J 121: 738–859
Khadjavi M (2015) On the interaction of deterrence and emotions. J Law Econ Organ 31:287–319
Lando H (2002) When is the preponderance of the evidence standard optimal? Geneva Pap Risk Insur Issue Pract 27:602–608
Lando H (2006) Does wrongful conviction lower deterrence? J Leg Stud 35:327–338
Lando H, Mungan MC (2017) The effect of type-1 error on deterrence. Int Rev Law Econ 53:1
Marchegiani L, Reggiani T, Rizzolli M (2016) Loss averse agents and lenient supervisors in performance appraisal. J Econ Behav Organ 131:183–197
Markussen T, Putterman L, Tyran J-R (2016) Judicial error and cooperation. Eur Econ Rev 89:372–388
Miceli TJ (2009) Criminal procedure. Edward Elgar Publishers, vol 3 of Criminal law and economics – encyclopedia of law & economics, Edward Elgar (ed), Chaltenham
Mungan M (2011) A utilitarian justification for heightened standards of proof in criminal trials. J Inst Theor Econ 167:352
Nicita A, Rizzolli M (2014) In Dubio Pro Reo. Behavioral explanations of pro-defendant bias in procedures. CESifo Econ Stud 60:554. ift016
Ognedal T (2005) Should the standard of proof be lowered to reduce crime? Int Rev Law Econ 25:45–61
Png IPL (1986) Optimal subsidies and damages in the presence of judicial error. Int Rev Law Econ 6:101–105
Polinsky A, Shavell S (1992) Enforcement costs and the optimal magnitude and probability of fines. J Law Econ 35:133–148
Risinger DM (2007) Innocents convicted: an empirically justified factual wrongful conviction rate. J Crim Law Criminol 97:761–806
Rizzolli M, Saraceno M (2013) Better that ten guilty persons escape: punishment costs explain the standard of evidence. Public Choice 155:395–411. https://doi.org/10.1007/s11127-011-9867-y
Rizzolli M, Stanca L (2012) Judicial errors and crime deterrence: theory and experimental evidence. J Law Econ 55:311–338
Schrag J, Scotchmer S (1994) Crime and prejudice: the use of character evidence in criminal trials. J Law Econ Org 10:319–342
Shavell S (1987) The optimal use of nonmonetary sanctions as a deterrent. Am Econ Rev 77:584–592
Yilankaya O (2002) A model of evidence production and optimal standard of proof and penalty in criminal trials. Can J Econ 35:385–409