Judicial effort and the appeal system: theory and experiment *

Eberhard Feess† and Roee Sarel‡

Draft version: September 25, 2015

Abstract

This is a proposal for a paper which consists of a theoretical part and an experimental part. The paper is not yet finished, but far enough developed for presentation and discussion. The text contains introduction, theory and the preliminary experimental design, where the experiment itself still needs to be conducted.

We now turn to the content: We develop a theoretical model to analyze how the existence of an appeal system influences the judicial effort of judges in the trial court. We assume that judges care about correct decisions and face reputational losses in case the court of appeals reverses their decisions. We then show that judges may either increase or decrease their effort in response to an appeal system, depending on their effort's marginal contribution to the accuracy of a subsequent appeal. Our model entails, as special cases, most of the preceding literature on the appeal system as an error correction tool. We complement our model by a laboratory experiment which is the first to behaviorally test the effects of appeal systems on judicial effort. As mentioned, the experiment is designed but has not been conducted yet.

JEL Classification: K4, C91
Keywords: Judicial effort, appeal system, experimental economics, law and economics

1 Introduction

Imagine you are asked to order lunch (say a burger) for two colleagues using an SMS message. If you misspell the SMS, one colleague will get two burgers while the other one will stay hungry. You can prevent this if you just pay enough attention to what you are writing, but this requires effort, which you dislike. Now imagine that you receive a new Smart-phone, which has an "auto-correct function" that corrects most of your mistakes. However, it is imperfect and may either fail to correct a mistake or make a new mistake by wrongly changing your message. To make things more complicated, assume that every auto-correct (justified or not) makes a loud noise that your boss can hear, thus hurting your reputation at work. With that said, will you pay more or less attention than you did before, given the auto-correct? Moreover, will your decision change if by exerting effort you could improve the accuracy of the auto-correct? Trial judges face a similar basic dilemma given the existence of an appeal system. Traditionally, the appeal system is perceived as a tool for error correction (Shavell 1995), by harvesting private information from litigants and allowing better informed/skilled judges to review the case. Appeal

*An earlier version of this working paper has been presented in the 2015 European association of law and economics conference in Vienna. We appreciate the valuable comments of the conference participants

†Frankfurt School of Finance & Management, D- 60314 Frankfurt, Germany. email: e.feess@fs.de
‡Corresponding author. Frankfurt School of Finance & Management, D- 60314 Frankfurt, Germany. email: R.Sarel@fs.de

systems are also argued to reduce errors indirectly, by altering the incentives of trial court judges. Judges are assumed to fear reversals due to reputation concerns (e.g. Shavell 1995, 2006) and thus exert effort in order to err less and avoid reversal. In a recent paper, Chopard et al. (2014) challenge this view by arguing that reputation concerns are fully balanced by the intrinsic motivation to avoid incorrect decisions ("social orientation"). Such motivation is said to encourage free-riding on the appeal court's error correction by decreasing effort (comparable to paying less attention for spelling while relying on the auto-correct's accuracy). Judges who are more socially than reputationally oriented are then argued to *always* decrease their effort given an appeal system. This result, however, may lead to the conclusion that only "immoral" judges, who simply don't care about the result, should be appointed. Albeit the intuition behind Chopard et al. (2014)'s model is clear, the extreme nature of their result is a byproduct of their rather restrictive assumptions that (i) no opportunistic appeals (i.e. appeals of a correct result) are ever filed, such that wrongful reversals do not exist, and (ii) the appeal court's accuracy is exogenously given, and hence not affected by the trial judge's effort.

In a nutshell, the exclusion of opportunistic appeals and wrongful reversals implies that social and reputation concerns will affect the trial judge's effort incentives *equally and in opposite directions*. It is however the assumption of an exogenous appeal accuracy, which implies that it is social orientation which incentivizes effort reduction (free-riding) while reputation concerns encourage effort (rather than the other way around). It is therefore the conjunction of these assumptions which facilitates free-riding, by ensuring that (i) reversals always mean "good news" for social concerns (since all reversals reflect error correction) and (ii) effort reduces the probability of correct reversals and increases the probability of wrongful reversals (which do not exist if no opportunistic appeals are allowed). The idea that a trial judge's effort may increase any kind of reversals may seem counter-intuitive, but it is in fact simple, since reversals depend on the appeal court's accuracy such that a higher accuracy necessarily means more correct reversals and less wrongful reversals. Thus, if the appeal accuracy is (un)affected by trial effort, the probability of correct (wrongful) reversal may increase (decrease).

While the assumption of "zero opportunistic appeals" has been relaxed by many papers that allow for such appeals (and wrongful reversals) in equilibrium, the assumption of an exogenous appeal accuracy is shared by a vast majority of the literature (with one exception discussed below) - albeit it is actually highly critical. We show that a trial judge's ability to affect the accuracy of the subsequent appeal will in fact determine whether he will decrease or increase effort given an appeal system. Our main result is that when trial judges' effort marginally (weakly) increases the accuracy of the appeal court more than it increases their own accuracy, then *all* judges will increase their effort in response to an appeal system, *regardless of their reputation and social concerns*. Otherwise, judges will exert more effort only if reputation concerns sufficiently crowd-out social concerns.

We further show that anticipation of opportunistic appeals (and wrongful reversals) will incentivize socially oriented judges to exert more (and not less) effort, such that social and reputation concerns (partially) work in the same direction. Our results shed new light on the results of another recent paper by Christmann (2013), who reaches the exact opposite conclusion than Chopard et al. (2014) - i.e. that given an appeal, socially oriented judges exert more effort while reputation oriented judges exert less effort. We show that the differences between the two papers are driven by opposite simplifying assumptions, each constituting special cases of our main results.

Testing our model by examining actual judges would unfortunately be "mission impossible", due to a lack of control group (with no appeal system) and unavailability of accurate proxies for judicial effort, its marginal effect, and social vs. reputation orientation. Instead, we opt to conduct a laboratory experiment, which allows for focused manipulation and clean measurement of the relevant variables. We

thus design an experiment in which subjects must choose their effort level given different treatments - with and without an appeal - which vary in the degree of marginal contribution of effort to the accuracy in the appeal and in the negative reputation effect from reversals.

Our paper makes several important contributions. We show that relaxing the line of restrictive assumptions leads to less polarized results than predicted thus far. We can then show that the impact of an appeal system on the trial court's effort is the difference in marginal contribution to the accuracy in each court (trial and appeal). Our result seems particularly relevant for the design of an effective appeal system, especially when considering the differences between common law and civil law systems. While common law appeal courts rely on facts gathered by trial courts, civil law appeal courts fully reconsider cases. It is then plausible to assume that the appeal's accuracy will be highly endogenous in common law courts and less so in civil law courts, leading to lower effort of trial judges in civil law courts. Civil law systems should therefore be more concerned with ensuring that reputation concerns sufficiently crowd-out free-riding on error correction. On the experimental side, our paper seems to be the first to test the effects of an appeal system on judicial effort. Our paper also extends the emerging experimental literature on judicial decision making in general.

The rest of the paper is organized as follows: Section 2 reviews related literature. Section 3 presents the basic model and our main theoretical results, which are then discussed in Section 4. Section 5 describes the experimental design. Section 6 (not yet available) will discuss our experimental results. Section 7 reviews the (theoretical) implications of our paper. Section 8 (not yet available) will conclude.

2 Related Literature

Our paper is closely related to the theoretical literature on judicial effort, rational judges and the role of appeal courts. Theories are mixed and often take opposite approaches regarding judicial incentives. Posner (1993) reviews various components of a judge's utility function, including popularity, prestige, reputation and reversal aversion. He deliberately excludes the public interest from the list, since he views judges as "ordinary people" who aren't especially concerned with such issues. Epstein et al. (2013) explore the incentives of federal judges in the US and model their decision making as maximizing their satisfaction from judicial work, subject to various constraints such as reputation, working relations with colleagues and opportunity cost of time (leisure and income from private lectures). Shavell (1995) comprehensively models judges incentives given an appeal system, based on the view that such systems aim mainly at error reduction. In an extension part of his paper, Shavell (1995, pp. 391,408) incorporates the judge's disutilty from reversals, which can be reduced on expectation through judicial effort. The appeal court's accuracy is modeled as exogenous from the trial judge's perspective.[1] Drahozal (1998) compares the utility of trial and appeal judges, both preferring a correct result but constrained by costly effort. Fear of reversal then leads to more effort not because of reputational sanctions, but rather due to the additional expected effort cost if the trial judge needs to reconsider the case again after it is remanded (returned to the trial court) by the appeal court. Levy (2005) argues that albeit judges fear reversal, they may still inefficiently provoke appeals by contradicting legal precedent in order to signal their ability. Conversely, Shavell (2006) argues that reversal-averse judges will anticipate the border conditions for an appeal to be filed (e.g. litigant's appeal costs) and decide cases such that no appeals are filed, unless their estimate is imperfect. Iossa (2007) compares the effort of arbitrators and judges and assumes that judges have no reputation concerns per-se, but do care about appearing competent to the litigants, which reflects a similar concern. Fischman (2006) points out the appeal system's role in incentivizing reversal averse judges to conform with existing legal precedent. Iossa & Palumbo (2007) compare appeal systems

relying on information gathered by independent investigators v. information harvested from litigants. They show that delegating the information gathering to the litigants is superior in reducing opportunistic judicial behavior, since litigants who gather information have a better estimation of reversal probability (For additional theoretical discussions of judicial incentives and the appeal system see also: Spitzer & Talley 2000, Daughety & Reinganum 2000, Lax 2003, Randazzo 2008, Shavell 2010).

The literature on opportunistic appeals is mixed as well. Shavell (1995) explicitly assumes that appeal systems successfully discourage all opportunistic appeals by establishing appropriate conditional appeal costs.[2] A similar view is adopted in Chopard et al. (2014). Cameron & Kornhauser (2005) argue along theses lines that opportunistic appeals are rare, since they are discouraged by the expected resistance of the opposing litigants. A different approach is taken by Cameron & Kornhauser (2006), who present two opposite views regarding judicial incentives - the "principal-agent" approach, in which judicial errors stem from a conflict of interest between judges and society; and the "team" approach, in which judges strive at minimizing errors but are constrained by hidden information and costly verification (see also Kornhauser 2012). The appellant's type - true or opportunistic - constitutes such hidden information in their model, leading to a potential pooling equilibrium in which opportunistic appeals do exist. Christmann (2013) adopts a similar view. Other papers avoid the discussion on opportunistic appeals altogether by assuming perfect accuracy in the appeal court (e.g. Levy 2005, Shavell 2006, 2007), which implies that no wrongful reversals exist anyway.

Our paper is also related to the behavioral and experimental literature on judicial decision making. This line of literature has, thus far, mostly focused on aspects which are neither concerned with judicial effort nor with effects of the appeal system, for example: heuristics and biases (e.g. Guthrie et al. 2000, 2007, Zamir & Ritov 2012, Bordalo et al. 2013), costly or reciprocal punishment (e.g. Crockett et al. 2014), dilemmas in punishment imposition (e.g. Friesen 2012, Ouss & Peysakhovich 2013) and legal uncertainty (e.g. Baker et al. 2003, Feess et al. 2014, Baumann & Friehe 2015). Two experiments, however, are more closely related to our paper. The first, conducted by Sonnemans & van Dijk (2012), raises a dilemma which is comparable to judicial effort by testing whether subjects in the role of judges properly use Bayesian updating when adjudicating criminal cases. In that experiment, subjects could order "further inquiries" (i.e acquire information) for a cost prior to the decision making, which somewhat resembles the decision of whether to exert effort in trial. However, Sonnemans & van Dijk (2012) do not incorporate the possibility of an appeal nor do they induce social preferences by using other subjects in the role of litigants - both of which are at the heart of our experiment. The second and very recent, by Lewisch et al. (2015), tests the effect of a second instance on deterrence and trial judges' punishment tendencies. Our paper, however, uses a different and novel approach which is especially tailored for testing our model. First, we vary the degree of endogeneity of the appeal accuracy. Second, we let judges decide on effort rather than punishment, which bears a greater resemblance to the fact-finding mission of a trial, where legal uncertainty is endogenous (i.e. depends on the judge's effort). Third, Lewisch et al. (2015) consturct an appeal system by using a third subject in the role of an appellate judge. We, however, implement the appeal court in the form of a computerized stochastic procedure. Our approach regarding this issue holds two key advantages: (1) it allows to accurately manipulate the appeal accuracy, and (2) establishes reversal probabilities that are common knowledge for both judges and violators. This ensures that any variation in trial judges' behavior is due to the different treatments rather than potentially different estimations of reversal probabilities. Finally, our settings allows to disentangle social concerns from reputation concerns, by using also a treatment with an appeal but without a reversal loss.

The empirical literature on judicial effort (and judicial incentives in general) has also made several attempts to tackle the judge's response to a possible reversal, focusing mostly on a trade-off between

reputation costs and political preferences (e.g. Lim 2013, Zorn & Bowie 2007). Findings have been diverse. Some papers argue that judges anticipate the ruling of the appeals court and try to avoid reversals in order to protect their political views (Scott 2006, Randazzo 2008, Choi et al. 2012, e.g.) or reduce their workload (Choi et al. 2012). Some papers claim, conversely, that reputational costs may be too low to change incentives (e.g. Cross 2007), at least for the middle instance (e.g. US federal appeal courts). Other papers emphasize the judge's difficulty to assess the probability of appeal (e.g. Bowie & Songer 2008). Theses papers naturally do not contain a very accurate measurment of judicial effort, and must instead resort to proxies such as the number of citations (may reflect a higher quality of decision) (e.g. Choi et al. 2012), interviews with judges (e.g. Bowie & Songer 2008) or survey data (e.g. Bar Niv & Lachman 2009).[3] Our paper is able to mitigate these problems, by isolating different incentives and creating symmetric information with respect to reversal probabilities.

3 The model

3.1 No appeal system

We consider a risk-neutral trial court judge who adjudicates a criminal case, in which there is only one correct result. This result can be interpreted, for example, as the correct legal result (i.e. proper application of the legal standard on the facts). The judge receives a fixed salary w and must choose a level of effort $e \geq 0$ which increases his own accuracy p_e (the probability that the correct result is reached) at a cost $c_e \geq 0$. For simplicity, we assume that the result is binary (e.g. convicting or aqcuitting the defendant) and thus $0.5 \leq p_e \leq 1$, since the judge can always just "flip a coin" and have 50% probability of reaching the correct result.[4] We further assume that the accuracy function p_e is strictly increasing in effort but is concave ($p'_e > 0, p''_e \leq 0$), such that investing effort has decreasing marginal benefit. The cost function c_e increases in e as well but is convex ($c'_e > 0, c''_e > 0$).[5] If the judge makes a wrong decision, he suffers a fixed "social" loss $s \geq 0$. Note that this loss my be different across judges, since it depends on the judge's sensitivity to errors, i.e. on his intrinsic motivation to reach a correct result.

Following the literature, we assume that judges may have different degrees of aversion for type I errors (wrongful convictions) and type II errors (wrongful acquittals). Such an aversion is expressed, for example, in the famous "Blackstone formula" which states that *better that ten guilty persons escape, than that one innocent suffers* (Blackstone 1765, pp. 352)[6] as well as in the higher standard of proof which usually accompanies a criminal trial (e.g. convict only if the defendant has been proven guilty "beyond a reasonable doubt"). This difference in aversions may of course result in judges preferring to exert different levels of effort, given probability for each type of error. In other words, judges may care about the probability that the crime was committed. We thus also assume, for simplicity, that judges correctly estimate the probability that the crime was committed by the defendant and can condition their effort accordingly. Formally, we denote $R \in \{0, 1\}$ as the binary result of the trial, where $R = 1$ if the decision is correct and $R = 0$ otherwise. The trial judge's accuracy can then be defined as:

$$p_e = Pr(R = 1|e, v)$$

Where e is the trial judge's effort and v is the probability that the defendant is a violator, i.e. that a crime was committed by the defendant. In our experiment we restrict attention to two scenarios: $v = 0$ (i.e. the defendant has not committed the crime) and $v = 1$ (i.e. the defendant has committed the crime). Thus, we are able to compare aversions by always keeping one type of error fixed on 0%. Namely, if the

defendant has not committed the crime, there is 0% for type II error; and if the defendant has committed the crime, there is 0% for type I error.

Finally, we assume that the market for errors is perfectly competitive, such that a trial judge cannot change the equilibrium probability of error. In other words, we assume that a single judge's effort does not have any effect on overall deterrence, since it is merely a "drop in the sea". Thus, a judge need not consider dynamic equilibrium effects of his effort but can instead assume that v is unaffected by effort choice.

In the absence of an appeal system, a trial judge's then maximizes as follows:[7]

$$\max_e u_{NA} = w - c_e - s(1 - p_e) \tag{1}$$

Taking the first derivative with respect to e, leads to the following FOC:

$$\frac{\partial u_{NA}}{\partial e} = -c'_e + sp'_e = 0 \Rightarrow c'_e = sp'_e \tag{2}$$

Equation (2) intuitively means that where there is no appeal system, a judge invests effort up to the point where his marginal effort cost is equal to his marginal benefit from improving his accuracy. We denote this effort level as e^*_{NA} where our standard assumptions ensure an interior solution.

3.2 A judge's utility with an appeal system

With an appeal system, several new variables must be added which we denote as follows:

- a_e - the accuracy of the appeal court, i.e. the appeal court's probability of reaching a correct result as a function of e. We assume $a_e \geq \bar{p} \geq p_e \geq 0.5$, where $\bar{p}$ is the maximal accuracy that the trial court judge can reach. This means that the appeal court's accuracy is at least as good as the trial court's accuracy.[8] More importantly, we assume $a'_e \geq 0$, which means that a_e could be either exogenous (in which case $a'_e = 0$) or increasing in the trial judge's effort.
- f_e - probability that an appeal is filed conditional on e. For the purpose of our model, we assume that this probability is endogenous and decreasing in (or unaffected by) effort ($f'_e \leq 0, f''_e \geq 0$) and is convex. In the basic model, we do not separate between the probability of opportunistic and non-opportunistic appeals, and assume that both are decreasing in (or unaffected by) the trial judge's effort. [9]
- $q_w = p_e(1 - a_e)$ - the probability of a "wrongful reversal", i.e the trial judge's decision was correct but was nonetheless mistakenly reversed by the appeal court.
- $q_r = (1 - p_e)a_e$ - the probability of a "right reversal", i.e. the trial judge's decision was wrong and was rightfully reversed by the appeal court.
- z - the trial judge's disutility from reversal. We assume $z \geq 0$.

Given the existence of an appeal system, the trial judge maximizes as follows:

$$\max_e u_{AP} = w - c_e - s[\underbrace{(1 - p_e)(1 - f_e)}_{\text{error but no appeal}} + \underbrace{f_e(1 - a_e)}_{\text{appeal court errs}}] - z\underbrace{(f_e(q_w + q_r))}_{\text{decision reversed}} \tag{3}$$

We simplify the utility function in Equation (3) and get Equation (4):

$$u_{AP} = w - c_e - s[(1 - p_e) - \underbrace{f_e(a_e - p_e)}_{b_e}] - z\underbrace{[f_e(q_w + q_r)]}_{d_e} \tag{4}$$

For brevity, we thus define two additional components:

- $b_e = f_e(a_e - p_e)$ - the net change in the probabilistic effect on s due to the appeal system. For $b_e = 0$, we get the same social utility component as in the case of no appeal system (equation (1)). For better understanding of the intuition, it is worthwhile mentioning that this net change can be rewritten as $b_e = f_e(q_r - q_w)$ (see proof 1 in the appendix). The opposite signs of q_r and q_w then reflect the simple facts that correct reversals are good news and wrongful reversal are bad news with respect to social concerns.
- $d_e = f_e(q_w + q_r)$ - the overall probability that a reputation loss is suffered.

As in the case of no appeal, we take the first derivative with respect to e and get the following FOC:

$$\frac{\partial u_{AP}}{\partial e} = -c'_e + sp'_e + sb'_e - zd'_e = 0 \Rightarrow c'_e = sp'_e + sb'_e - zd'_e. \tag{5}$$

We denote the effort level which fulfills Equation (5) as e^*_{AP}, where our standard assumptions ensure an interior solution.

3.3 The appeal system's countervailing effects on trial judges' effort

Comparing the cases with and without appeal shows that $e^*_{AP} \geq e^*_{NA}$ if and only if $s[p'_e + b'_e] - zd'_e - c'_e \geq sp'_e - c'_e$, which can be simplified to:

$$sb'_e \geq zd'_e \tag{6}$$

We thus get

Proposition 1 *A trial judge exerts (weakly) more effort given an appeal system if and only if* $sb'_e \geq zd'_e$.

A rather salient first result which arises from equation (6) is that the critical condition for exerting effort given an appeal system is unaffected by the effort cost function. Since the cost function does not change when an appeal system is introduced, this result is not very surprising. However, the effort cost does have a direct effect on the base level of effort (without appeals), namely if a judge already exerts very high effort due to a high s (relative to c_e) or very little effort due to a low s, then introducing an appeal system may not make much difference (unless reputational concerns become significant). The difference in effort then arises only when the level of s is moderate.

Effort cost aside, when looking at inequality (6), it is easy to see how one might think that s and z always pull in opposite directions. However, this of course depends on the sign (negative or positive) of both b'_e and d'_e, which in turn depends on the combination of variables that compose these elements, i.e. on f_e, q_r, q_w, a_e, p_e. Nonetheless, since we assumed $s, z \geq 0$ there are some conditions under which inequality (6) always holds, namely when $b'_e \geq 0, d'_e \leq 0$. In such cases, all judges will exert more effort

given an appeal system regardless of their orientation, i.e. irrespective of the size of s and z. Our second proposition is thus:

Proposition 2 *All judges exert more effort given an appeal system, irrespective of their orientation if (1)* $d'_e \leq 0$ *and (2)* $b'_e \geq 0$

We now explore these conditions. The reputation orientation probability element d'_e, is always non-positive (see proof 2 in the appendix), which implies that *effort always increases in* z. This result is relatively straightforward - a judge with higher reputation concerns is more careful not to err, since committing an error is detrimental. In preceding models in which no opportunistic appeals existed and effort reduced the probability of reversal (e.g. Shavell 1995) this result was trivial, since avoiding an error (by exerting effort) prevented an appeal (alongside the risk of reversal) from being filed in the first place. This however also holds when opportunistic appeals are allowed, as long as accuracy is sufficient. In our model, accuracy (in both courts) is higher than 50% and thus implies that on average (more often than not) errors are reversed and correct decisions are not, such that effort still minimizes overall reversals.

The effect of the social orientation probability element b'_e is a bit more illusive, as it is not strictly positive or negative but rather depends on the variables in play - the probability of an appeal f_e, the accuracy of the trial court p_e and the accuracy of the appeal court a_e. Namely, two effects take place - on one hand, when effort increases, the probability of an appeal being filed decreases. On the other hand, accuracy in both courts changes. If the accuracy of the appeal court insufficiently increases in comparison to the trial court's accuracy, i.e. if $a'_e < p'_e$, then the social orientation probability will be non-positive, in which case effort will always decrease in s.

More generally, a trial judge exerts more effort irrespective of s and z only if the following condition holds (see proof 3 in the appendix):

$$f_e(a'_e - p'_e) \geq f'_e(p_e - a_e) \tag{7}$$

The intuition behind this equation is as follows: the RHS is the loss of potential additional accuracy of the appeal court by exerting effort. A trial judge gains utility from appeal through the additional accuracy over his own, which is $a_e - p_e$. By exerting effort he reduces the probability of appeal thereby decreasing this benefit on expectation. The LHS is respectively the gain of accuracy in case an appeal does take place. Exerting effort therefore involves a trade-off: on one hand, effort may improve the accuracy in both courts (or at least in the trial court). On the other hand, effort reduces the chance that an appeal is filed and thus the appeal court - who is more accurate - will get to review the case. The optimal effort level will then depend on the size of each effect.

A special case of particular interest is when f_e is exogenous, i.e. when the judge can't affect the probability that an appeal is filed. Realistically, a judge can try to prevent (or encourage) appeals, but his ability to do so is limited. Presumably, financially constrained litigants may never file appeals due to the high cost of litigation while wealthy litigants will file opportunistic appeals to try and pressure their rivals into settlement. Some litigants may also miscalculate their chances on appeal due to asymmetric information (e.g. Cross 2007, pp. 125) irrespective of judicial effort. Assuming an (almost) exogenous f seems plausible.[10] When f is exogenous, then $f'_e = 0$ and inequality (7) becomes $f_e(a'_e - p'_e) \geq 0$, which always holds if $a'_e \geq p'_e$, i.e. when exerting effort marginally increases the accuracy of the appeal court more than it marginally increases the accuracy of the trial court. A similar result is achieved in the special case where $a_e = p_e$ (implying also $a'_e = p'_e$), such that effort equally increases the accuracy in both courts and thereby cancels out any disutility from reducing the probability of an appeal (inequality

(7) becomes $0 \geq 0$ which always holds).

Another special case of interest occurs when a_e is exogenous, i.e. when a trial judge has no effect on the accuracy of the appeal court. In such a case, $a'_e = 0$ and $p'_e > a'_e$ (since $p'_e > 0$), thus effort decreases in s. Our results in this section are summarized in Proposition 3 and corollaries 1-3.

Proposition 3 *The effect of an appeal system on a trial judge's incentives to exert effort, considering the judge's degree of social orientation s and reputation concerns z, is as follows:*

1. *When z increases, ceteris paribus, a trial judge will (weakly) exert more effort*
2. *When s increases, ceteris paribus, a trial judge will either exert less or more effort:*
   (a) *Case 1: if $a'_e < p'_e$, an increase in s will cause the trial judge to exert less effort*
   (b) *Case 2: if $a'_e \geq p'_e$, an increase in s will cause the trial judge to exert more effort if $f'_e(a_e - p_e) \leq f_e(a'_e - p'_e)$ and less effort otherwise*

Corollary 1 *in the special cases (of case 2) where $f'_e = 0$ (f is exogenous), a trial judge will always exert more effort if $a'_e \geq p'_e$ given an increase in s*

Corollary 2 *in the special case (of case 1) where $a'_e = 0$ (a is exogenous), a trial judge will always exert less effort given an increase in s*

Corollary 3 *in the special case (of case 2) where $a_e = p_e$, a trial judge will always exert effort (irrespective of f_e) given an increase in s*

Our theoretical results are discussed in the following section.

4 Theoretical discussion

4.1 Explaining the results of recent conflicting papers

At this point we are ready to return to the two recent papers by Chopard et al. (2014) and Christmann (2013) and explain their polar arguments by our results. Chopard et al. (2014)'s model is the special case mentioned in corrolary (2), in which a is exogenous, thus leading to the conclusion that socially oriented judges always exert less effort. On the contrary, Christmann (2013) is an example of the specific case laid out in corollary 3. He argues that the appeal court is at least as accurate as the trial court, but assumes additionally that they are "prone to the same probability of error" (pp. 16), which leads him to conclude that socially oriented judges ("resolving judges", who have zero reputation concerns) always exert more effort given an appeal system (pp. 21). His other opposite result - that reputation concerned judges exert less effort - is driven by an additional assumption that no-reversal leads to "extra reputation" and therefore such judges provoke appeal, similarly to Levy (2005). In our model, we assume that such utility, if it exists, is equal to 0. A similar result can nonetheless be achieved also under our model, by defining $z^{new} = K - z^{old}$ where K is the extra reputation gained, and assuming $z^{new} < 0$. This would lead to the exact opposite of proposition 3.1, i.e. that increases in z^{new} induce less effort.

Our paper differs from these two papers in several key issues. First, we relax the assumption of exogeneity of appeal accuracy but without assuming any special cases of an endogenous appeal accuracy. Instead, we explore the borderline conditions which determine when effort is exerted, which allows us to explain the existing polar results. Second, while our basic model allows for endogenous probability of an appeal being filed (i.e. f_e) we focus on the accuracy functions. Therefore, we adopt a simplified example

for our experiment, with exogenously given chance of appeal. Third, we take the next step of conducting an experiment to test our model's predictions.

4.2 The role of opportunistic appeals

We complete our theoretical discussion by considering the role of opportunistic appeals. Formally, excluding opportunistic appeals or alternatively assuming perfect accuracy of the appeal court, sets the probability of a wrongful reversal to 0. To see how this would affect our model, we can first rearrange inequality (6) ($sb'_e \geq zd'_e$) by substituting for d_e and b_e: $s[(f_e q_r)' - (f_e q_w)'] \geq z[(f_e q_r)' + (f_e q_w)']$, or:

$$\underbrace{(f_e * q_r)'(s-z)}_{\text{marginal effect on right reversals}} \geq \underbrace{(f_e * q_w)'(s+z)}_{\text{marginal effect on wrongful reversals}} \tag{8}$$

The LHS of inequality (8) expresses the effect of marginally increasing the probability of a right reversal by exerting effort. Note that s and z have opposite signs, since a judge gains social utility when his errors are corrected, but at the same time (unjustifiably, but nonetheless) loses reputation. Conversely, on the RHS, s and z have identical signs, since a wrong reversal causes a loss in both social and reputation elements.

If we set $q_w = 0$, inequality (8) then becomes: $(f_e q_r)'(s-z) \geq 0$, which implies that s and z equally affect the effort incentives and in opposite directions, as suggested by Chopard et al. (2014). The intuition behind this result seems clear - when only right reversals exist, any reversal is necessarily "good news" for socially oriented judges, as it implies error correction, while still being "bad news" for reputationally oriented judges. However, what determines whether exerting effort then increases or decreases the probability of reversal depends, once again, on the difference between a'_e and p'_e. If exerting effort insufficiently improves the accuracy of the appeal court (i.e. $a'_e < p'_e$, as in Chopard et al. (2014)), effort indeed reduces (right) reversal probability (i.e. then $q'_r < 0$, see proof 4 in the appendix). Contrarily, if effort does sufficiently improve the accuracy of the appeal court (i.e. $a'_e \geq p'_e$), then exerting effort may actually increase the reversal probability. This occurs since a higher accuracy implies, in this case, that more decisions (which can only be erroneous) are reversed (i.e. the proportion of reversed appeals then converges with accuracy). Thus, excluding opportunistic appeals, by itself, is insufficient to conclude that socially oriented judges always decrease their effort in response to an appeal system. It is rather the conjunction with the assumption of an exogenous appeal accuracy which brings about the extreme nature of Chopard et al. (2014)'s result.

Note, however, that excluding opportunistic appeals while assuming instead $a'_e \geq p'_e$ would lead to an opposite extreme result, in which reputation oriented judges are the ones who always exert less effort (i.e the opposite of proposition 3.1). To avoid extreme results to either direction, the (very restrictive) assumption of no opportunistic appeals must also be relaxed.

5 Experimental design

We design a laboratory experiment to test our model's predictions on a behavioral level. The experiment will be conducted in the Frankfurt School of Finance & Management experimental economics lab, using the software z-tree (Fischbacher 2007). Participants will include around 150 bachelor or master students and will receive a show up fee of 4 ECU[11] in addition to the potential earnings in the experiment itself.

Before the experiment begins, subjects must answer a series of control questions to ensure that they have understood the instructions. Then, subjects proceed to make decisions.

5.1 Decisions on violation or effort

We follow a similar path as the existing literature (e.g. Feess et al. 2014, Rizzolli & Stanca 2012, Lewisch et al. 2015), by using theft as the potential crime that subjects can commit. Each participant is randomly assigned one of two possible roles: A (judge) or B (potential violator). This role stays the same throughout the experiment. To avoid framing effects, we do not make use of any words that are related to a legal process (e.g. "judge", "crime", "court"). Participants receive an initial endowment of 1200 Experimental Currency Units (ECU), which they can convert to cash at the end of the experiment (the exchange rate is 500 ECU = 1 Euro). Each participant in role of violator is randomly and anonymously matched with another participant in the role of a judge. The participants then interact through a series of decisions.

Violators make one binary decision: whether to take for themselves another 1200 ECU, which are otherwise donated to a charity,[12] similar to the procedure used by Feess et al. (2014). Thus, if a violator takes the money, she will have 2400 ECU and if she does not take the money, she will be left with her initial endowment of 1200 ECU. Post-trial, however, a violator may be punished, in which case she loses all her remaining ECU (but not the show up fee, to avoid bankruptcy problems). Any ECU that the violator loses as a result of punishment are donated to the charity as well. This ensures that a punishment does not entail an efficiency loss (which could confound the judge's motivation), and also has the advantage of replicating the practice of compensating the victim for the crime. The decision of whether the violator is punished or not is taken by the computer (and not directly by a participant in the role of judge). However, the probability of punishment depends on the decision of the judge whom the violator is matched with.

Judges also make binary decisions: whether to invest part of their initial endowment (invest 400 ECU) in order to increase the probability of a correct result. If the judge chooses not to invest (=no effort), the computer will have an accuracy of 50%, such that a potential violator has equal probability of being punished or unpunished. However, if the judge chooses to invest, the accuracy will increase to 70% (a marginal increase of 20 percentage points). Following our model, we allow judges to condition their effort choice on the probability of theft but limit their choice to either 100% or 0%, which means that judges make two binary decision: one for the case where the violator has stolen and one where he has not stolen. Judges thus choose effort via the strategy method (but without knowing the actual choice of the violator they are matched with). This approach may seem unnatural, since judges usually do not know for certain whether the defendant has stolen, but it is actually crucial for a proper test of our model and also holds many advantages. First, to ensure no deterrence concerns, judges must perceive the decision of violators as given, which is achieved by letting judges choose conditionally. Second, focusing on the cases of stealing vs. no-stealing is merely a simplification of conditioning effort on probability of theft, which is necessary in order to ensure that observed variation in effort stems from the different treatments rather than different estimations of probabilities across subjects. Third, the strategic method allows to compare type I vs. type II aversions, by keeping the probability for one of the two types of error always fixed. Fourth, it resembles a judge's task of verifying the evidence while still keeping the aspect of legal uncertainty, which is expressed by the probabilistic nature of final result. Note that participants in the role of judges do not receive any monetary incentives to reach the correct result, but instead we rely on the actual concerns and intrinsic motivation of participants. Since we also use a setting without appeal

(see below), we can observe the base effort level (which can then be attributed to social concerns) and conduct subsequent comparisons. The decisions of judges and potential violators are subjected to varying treatments, in order to analyze two main effects:

1. *Enodegeneity effect* - do trial judges exert more effort when the appeal accuracy is (sufficiently) endogenous in comparison to exogenous? We analyze this effect by using between-subject variation.
2. *Reversal loss effect* - do trial judges exert more effort when a reversal entails a loss? We analyze this effect by using within-subject variation.

5.2 Within subject variation - no appeal vs. appeal vs. reversal loss

Each pair of judge-violator makes three sets of decisions, under three different treatments, labeled "NA", "AP" and "AR".

- **No appeal ("NA")**: In the NA treatment, the result of the case is decided by a simple matching of the violator's decision with the respective effort of the judge (i.e. the chosen effort for stealing if the violator stole, and chosen effort for no-stealing otherwise). Then, the result (punishment or no-punishment) is chosen by the computer according to the relevant probability which corresponds with the judge's choice.
- **Appeal without a reversal fine ("AP")**: In the AP treatment, the result of the case is first calculated similarly to the NA treatment. Then, the computer decides whether this result is final (i.e. there is no appeal filed) or proceeds to appeal, in which case it may be reversed by the computer according to a commonly known reversal probability (the accuracy of the appeal court varies only between subject groups - see below). We fix the probability that the case proceeds to appeal (f in our model) on 50%, in order to remove the theoretical trade-off with improving accuracy (we are thus only testing the special case of corrollary 1). In this treatment, a reversal by the computer does not entail any monetary loss to the judge, but only affects the result of the case (whether the potential violator is punished).
- **Appeal with a reversal fine ("AR")**: The AR treatment is identical to the AP treatment, except for the following difference: if the computer reverses the first result, the judge suffers a monetary reversal fine of 500 ECU.

5.3 Between subject variation - exogenous v. endogeouns appeal accuracy and order effects

Each pair of judge-violator is randomly assigned to only one of two groups, labeled "EX" or "EN":

- **Exogenous appeal accuracy ("EX"):** in the EX group, the appeal accuracy is always 70% and is unaffected by the choice of effort, which reflects an exogenous appeal accuracy.
- **Endogenous appeal accuracy ("EN"):** in the EN group, the appeal accuracy changes with the choice of effort. If the judge chooses not to invest, the appeal accuracy will be 70%. If the judge chooses to invest, the appeal accuracy increases to 90% (a marginal increase of 20 percentage points, identical to the marginal increase of the trial accuracy).

  The parameters of the EX and EN groups are summarized in table 1 below.

The "EX" and "EN" groups are then also divided to two subgroups to control for order effects of the treatments with appeal (AP and AR). We do not control for order effect of appeal vs. no appeal, since

	Trial accuracy (both groups)	Appeal accuracy - EX group	Appeal accuracy - EN group
No effort exerted	50%	70%	70%
effort exerted	70%	70%	90%

Table 1: trial and appeal accuracy by group

the setting of no appeal is actually identical across all groups. Therefore, any differences in behavior can still be attributed to the treatments. The different treatments are summarized in figure 1.

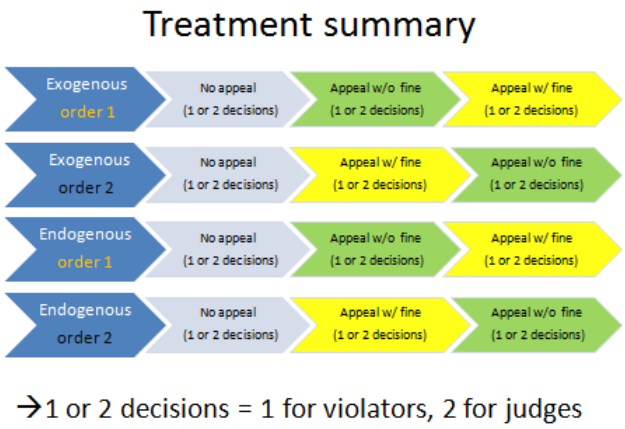

Figure 1: treatment overview

After all decisions are made, the computer chooses randomly which decision is paid out for each pair of judge-violator, i.e. only one of the three decisions (in NA, AP or AR) is paid out. Subjects are informed on the results, but not on each other's choices. This ensures that neither judges nor violators are concerned with social desirability towards the other participant when choosing whether to exert effort or whether to steal. After the second round is also complete, subjects proceed to some additional measurements.

5.4 Additional measurements and controls

We control for several additional factors:

- **Elicitation of beliefs:** we elicit beliefs using a strategy method from both violators and judges on two issues: (1) the decisions of the participant they are matched with, and (2) the percentage of judges/violators who invested/stole within the session. Note that the first point should matter for violators but theoretically should not matter for judges, who can condition their effort anyway. We incentivize honest disclosure of beliefs by using the method suggested by Schlag et al. (2013), which facilitates reports of true beliefs without making assumptions on risk preferences of subjects. Namely, participants can win an additional amount of 2000 ECU, where the probability of winning increases in the accuracy of the reported beliefs.
- **Risk attitude**: we measure risk aversion of subject by employing a binary lottery scheme, similar

to the method proposed by Holt & Laury (2002). Participants are asked to make a series of choices (10 altogether) on whether they prefer to receive 1200 ECU for certain or instead a lottery ticket which pays 3000 with probability p or 0 ECU with probability $(1-p)$. One of the decision is then randomly drawn and paid out.

- **Preferences regarding type I and II errors** - subjects answer a series of questions which reflect their preferences for type I and II errors. For example, subjects will be asked to evaluate the severity of a wrongful conviction and a wrongful acquittal.
- **Reported motivation for choices**: A set of open questions in which subjects self-report on their reasons for choosing the level of effort in each treatment. Such self-reported answers may be unreliable (Nisbett & Wilson 1977), but are nonetheless helpful as a complementary tool.
- **General questionnaire** - to control for addition effects (e.g. gender, age, education, economic status).

When the additional measurements are complete, the experiment ends.

6 Experimental results - not yet available

7 Theoretical Implications

Our (theoretical) results encompass an important implication for the design of an effective appeal system and its indirect effects on error reduction by altering trial judges' incentives. Since the impact of an appeal system critically depends on the difference between the trial effort's marginal contributions to trial and appeal accuracy, this difference should be taken into account. Namely, policy makers can reduce the problem of potential free-riding of socially oriented trial judges on appeal accuracy in one of two ways: (1) ensuring a sufficient positive externality of effort at trial to the appeal court's accuracy (for example, by limiting the scope of appeals to facts found in the trial), or (2) creating serious reputation concerns to crowd-out social orientation (for example, by making the promotion of judges highly contingent on having a very low reversal record).

These implications seem especially relevant when considering the differences between common law and civil law appeal systems (for comprehensive reviews on the many differences between these legal systems, see for example: Koetz & Weigert (1977), David (1978), Merryman & Pérez-Perdomo (2007), Shapiro (2013)). In common law countries, trial courts and appeal courts have very different roles. The trial courts conduct a fact-finding process by allowing litigants to submit evidence, question witnesses and present factual and legal arguments. Within this process, which follows the adversarial tradition, the responsibility to present the facts lies with the litigants, who must "battle" in court to convince the court of their arguments. A trial court then considers all the facts, interprets the relevant legal rules and applies them to the found facts to determine the result. When litigants are allowed to appeal the decisions, they cannot (as a rule)[13] challenge factual questions but only legal questions. An appeal court accordingly reviews only the trial court's interpretation and application of the legal rules to the facts as they were found, and does not reconsider the facts themselves (for reviews and historical developments see Grunewald (2013, pp.1180-1191) and Kunsch (1994)). The limits of the common law appeal courts scope of review presumably create a high inter-dependency between the accuracy of the trial court and the appeal court, since a trial judge who did not invest effort in fact-finding and erred as a result will limit the appeal court's ability to reach the correct result as well. Subsequently, free-riding of social judges in common law systems seems unlikely.

Conversely, in civil law systems, the distinction between the roles of each court in regards to fact finding is more vague. Both trial court judges and appeal court judges follow the inquisitorial tradition, in which the responsibility to collect all the facts lies with the judge rather than the litigants (for a comparison of adversarial and inquisitorial systems, see for example: Parisi (2002) and Grunewald (2013)). Litigants are then free to appeal questions of both fact and law, and the appeal court can conduct a 'trial *de novo*'' and fully reconsider the case (e.g. Shapiro 2013, pp. 149).[14] Presumably, this may lead to more free-riding through two channels, since a trial judge will have (1) higher influence on his own accuracy, but (2) less influence on the accuracy of the appeal court. Civil law systems may therefore be more exposed to the problem of free-riding by socially-oriented judges. To mitigate this problem without changing the basis of the legal system, civil law countries should therefore ensure that reputation concerns sufficiently crowd-out social concerns. Two "bright spots" should, however, be considered. First, the problem may be mitigated by the difference in the structure of judicial careers between civil and common law countries. In civil law system, judges usually embark on a judicial career directly after graduation from law school, while common law judges are usually appointed after accumulating several years of experience in the legal field (Zajc 2011). This may imply that civil law judges will generally have higher reputation concerns, either due to the structure of their career or to self-selection of career oriented people into the judicial profession. Second, it seems that some european civil-law countries are already taking special measurements to increase judicial effort (see Zajc 2011, pp.13), perhaps following a similar concern as the one expressed in this paper.

8 Conclusion - not yet available

Appendix

Proof 1 *We show that* $b_e = f_e(q_r - q_w)$ *as follows:* $f_e(q_r - q_w) = f_e((1-p_e)a_e - p_e(1-a_e) = f_e(a_e - a_e * p_e - p_e + a_e * p_e) = f_e(a_e - p_e) = b_e$

Proof 2 *We show that* $d'_e < 0$*, as follows: Since* $d_e = f_e(q_r + q_w)$*, its derivate with respect to e leads to* $d'_e = f'_e(q_r + q_w) + f_e(q'_r + q'_w)$ *which is always non-positive: The first term* $f'_e(q_w + q_r)$ *is non-positive, because* $f'_e \leq 0$ *by definition and* $q_w + q_r$ *is a combination of probabilities, which must be non-negative. The second term* $f_e(q'_r + q'_w)$ *is non-positive as well: we assumed* $f_e \geq 0$ *(a probability), and* $q'_w + q'_r = a'_e + p'_e - 2(p_e a_e)' = a'_e + p'_e - 2(p'_e * a_e + a'_e * p_e) = a'_e * (1 - 2p_e) + p'_e * (1 - 2a_e) \leq 0$ *since we assumed* $a_e \geq p_e \geq 0.5$ [15] *Since* $d'_e \leq 0$ *always holds, effort always increases in z (the condition* $sb'_e \geq zd'_e$ *is more likely to hold when z increases).*

Proof 3 *Since* $b_e = f_e(a_e - p_e)$*, its derivative with respect to e leads to* $b'_e = f'_e(a_e - p_e) + f_e(a'_e - p'_e)$*. Unlike the reputation components discussed above in proof 2, the sign of* b'_e *depends on the variables. The first term* $f'_e(a_e - p_e)$ *is non-positive:* $f'_e \leq 0$ *and* $a_e - p_e \geq 0$ *(we assumed* $a_e \geq p_e$*). The second term* $f_e(a'_e - p'_e)$ *depends on the slope of* p'_e *and* a'_e *(i.e. how fast each of them increase in e), since* $f_e \geq 0$ *(a probability) by definition. If* $a'_e < p'_e$*, the second term is non-positive, leading to* $b'_e \leq 0$*. In such a case, effort always decreases in s (the condition* $sb'_e \geq zd'_e$ *is less likely to hold when s increases). On the contrary, if* $a'_e \geq p'_e$ *then the first and second terms of* b'_e *have opposite signs (first term is always non-positive as said, second term is non-negative). Therefore, for* $b'_e \geq 0$ *to hold, the necessary condition is:* $f_e(a'_e - p'_e) \geq f'_e(p_e - a_e)$

Proof 4 *Since* $q_r = (1-p_e)a_e$ *its derivative with respect to e leads to* $q_r = a'_e(1-p_e) - p'_e a$*. We assumed* $a_e \geq p_e$*, therefore for* $q_r \leq 0$ *a sufficient condition is* $a'_e < p'_e$*.*

Notes

[1]Shavell (1995) models the appeal court's accuracy as derived from the level of resources invested by the state. This accuracy is therefore unaffected by the trial judge's effort.

[2]In his comprehensive paper, Shavell (1995) does consider and discusses the possibility of opportunistic appeals. However, in the relevant extension on reversal, he first concludes that all such appeals have been discouraged.

[3]Survey data may be especially biased. For example, in Bar Niv & Lachman (2009), judges were asked to answer whether they believe that writing longer verdicts will reduce error. Judges then may have an incentive to misrepresent their true belief in order to justify low effort

[4]A similar view has been taken, for example, in Drahozal (1998, pp. 493) who states: "At a minimum, the judge could spend no time at all on the case-simply guessing at the outcome- and still have some probability of getting it right."

[5]While it is sufficient for theoretical purposes that *either* p_e is concave or that c_e is convex, we set both in this way for experimental convenience, in order to differentiate between the effects of each element.

[6]For recent discussions on the Blackstone formula, see (Rizzolli & Saraceno 2013, Epps 2015)

[7]In order to make the comparison of our results to those of Chopard et al. (2014) easier, we use similar notation for most elements.

[8]The assumption that the appeal court is at least as accurate as the trial court reflects the positive added value of an appeal court.

[9]Note that the probability that an appeal is filed is independent of the correctness of the result but depends only on effort. This reflects the idea that when the judge exerts more effort, there is more information available, which reduces the probability of appeal irrespective of the trial result

[10](some models go so far as to assume that all cases proceed to appeal Randazzo 2008, pp. 675)

[11]Since the experiment has yet to be conducted, this number of subjects, as well as the exchange rate etc. might change.

[12]The use of a charity as the potential victim is presumed to increase the judge's preference for correcting a wrong. This presumption is intuitive, since stealing from a charity is likely to be perceived as worse than a theft from another participant. However, such behavior has also support by the results of previous papers, which find a higher tendency to donate where the recipient is a charity (e.g. Eckel & Grossman 1996, Konow 2010) or is otherwise deemed worthy (Fong 2007). It is worthwhile noting, however, the some papers have challenged the external validity of charitable behavior in the lab (see Levitt & List 2007)

[13]for example, Rule 52(6) of the US Federal Rules of appellate procedure states that facts cannot be set aside by the appeal court unless they are "clearly erroneous".

[14]While we consider in our model only a two-tier court system, in most countries litigants can under some conditions appeal the case to a third higher court (e.g. a supreme court). Supreme courts in both legal systems (common law and civil law) usually do not reconsider facts but only questions of law. See for example Shapiro (2013, pp. 149) and Merryman & Pérez-Perdomo (2007, pp. 85-86)

[15]This result is independent of whether a_e is exogenous or endogenous, and requires only that $a_e \geq p_e \geq 0.5$ as assumed.

References

Baker, T., Harel, A. & Kugler, T. (2003), 'Virtues of uncertainty in law: An experimental approach, the', *Iowa L. Rev.* **89**, 443.

Bar Niv, M. & Lachman, R. (2009), 'Do judges act against their own judgment? (in hebrew)', *Hapraklit (The Journal of the Israeli Bar-Association)* **50**, 71–93.

Baumann, F. & Friehe, T. (2015), *Proof beyond a reasonable doubt: Laboratory evidence*, number 181, DICE Discussion Paper.

Blackstone, W. (1765), 'Commentaries on the laws of england, 4 vols', *Oxford* **4**.

Bordalo, P., Gennaioli, N. & Shleifer, A. (2013), Salience theory of judicial decisions, Technical report, National Bureau of Economic Research.

Bowie, J. B. & Songer, D. R. (2008), 'Assessing the applicability of strategic theory to explain decision making on the courts of appeals', *Political Research Quarterly* .

Cameron, C. M. & Kornhauser, L. A. (2005), 'Decision rules in a judicial hierarchy', *Journal of Institutional and Theoretical Economics JITE* **161**(2), 264–292.

Cameron, C. M. & Kornhauser, L. A. (2006), 'Appeals mechanisms, litigant selection, and the structure of judicial hierarchies', *Institutional Games and the US Supreme Court* pp. 173–204.

Choi, S. J., Gulati, M. & Posner, E. A. (2012), 'What do federal district judges want? an analysis of publications, citations, and reversals', *Journal of Law, Economics, and Organization* **28**(3), 518–549.

Chopard, B., Marion, E. & Roussey, L. (2014), Does the appeals process lower the occurrence of legal errors?, Technical report, University of Paris West-Nanterre la Défense, EconomiX.

Christmann, R. (2013), Tipping the scales: Conciliation, appeal and the relevance of judicial ambition, Technical report, Diskussionspapier, Helmut-Schmidt-Universität, Fächergruppe Volkswirtschaftslehre.

Crockett, M. J., Özdemir, Y. & Fehr, E. (2014), 'The value of vengeance and the demand for deterrence.', *Journal of Experimental Psychology: General* **143**(6), 2279.

Cross, F. B. (2007), *Decision making in the US Courts of Appeals*, Stanford University Press.

Daughety, A. F. & Reinganum, J. F. (2000), 'Appealing judgments', *The Rand Journal of Economics* pp. 502–525.

David, R. (1978), *Major legal systems in the world today: an introduction to the comparative study of law*, Simon and Schuster.

Drahozal, C. R. (1998), 'Judicial incentives and the appeals process', *SMU Law Review* **51**, 469.

Eckel, C. C. & Grossman, P. J. (1996), 'Altruism in anonymous dictator games', *Games and economic behavior* **16**(2), 181–191.

Epps, D. (2015), 'Consequences of error in criminal justice, the', *Harv. L. Rev.* **128**, 1065.

Epstein, L., Landes, W. M. & Posner, R. A. (2013), *The behavior of federal judges: a theoretical and empirical study of rational choice*, Harvard University Press.

Feess, E., Schramm, M. & Wohlschlegel, A. (2014), 'The impact of fine size and uncertainty on punishment and deterrence: Evidence from the laboratory', *Available at SSRN 2464937* .

Fischbacher, U. (2007), 'z-tree: Zurich toolbox for ready-made economic experiments', *Experimental economics* **10**(2), 171–178.

Fischman, J. B. (2006), Uniformity of interpretation in a hierarchical court, Technical report, Working Paper. 14 Jan.

Fong, C. M. (2007), 'Evidence from an experiment on charity to welfare recipients: Reciprocity, altruism and the empathic responsiveness hypothesis*', *The Economic Journal* **117**(522), 1008–1024.

Friesen, L. (2012), 'Certainty of punishment versus severity of punishment: An experimental investigation', *Southern Economic Journal* **79**(2), 399–421.

Grunewald, R. (2013), 'Comparing injustices: Truth, justice, and the system', *Alb. L. Rev.* **77**, 1139–1201.

Guthrie, C., Rachlinski, J. J. & Wistrich, A. J. (2000), 'Inside the judicial mind', *Cornell L. Rev.* **86**, 777.

Guthrie, C., Rachlinski, J. J. & Wistrich, A. J. (2007), 'Blinking on the bench: How judges decide cases', *Cornell L. Rev.* **93**, 1.

Holt, C. A. & Laury, S. K. (2002), 'Risk aversion and incentive effects', *American economic review* **92**(5), 1644–1655.

Iossa, E. (2007), 'Reputational concerns in arbitration: Decision bias and information acquisition', *Available at SSRN 1018202* .

Iossa, E. & Palumbo, G. (2007), 'Information provision and monitoring of the decision-maker in the presence of an appeal process', *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft* pp. 657–682.

Koetz, H. & Weigert, K. (1977), *An introduction to comparative law*, North-holland Publishing Company.

Konow, J. (2010), 'Mixed feelings: Theories of and evidence on giving', *Journal of Public Economics* **94**(3), 279–297.

Kornhauser, L. A. (2012), '2 appeal and supreme courts', *Procedural Law and Economics* **8**, 19.

Kunsch, K. (1994), 'Standard of review (state and federal): A primer', *Seattle UL Rev.* **18**, 11.

Lax, J. R. (2003), 'Certiorari and compliance in the judicial hierarchy discretion, reputation and the rule of four', *Journal of Theoretical Politics* **15**(1), 61–86.

Levitt, S. D. & List, J. A. (2007), 'What do laboratory experiments measuring social preferences reveal about the real world?', *The journal of economic perspectives* pp. 153–174.

Levy, G. (2005), 'Careerist judges and the appeals process', *RAND Journal of Economics* pp. 275–297.

Lewisch, P., Ottone, S. & Ponzano, F. (2015), 'Third-party punishment under judicial review: An economic experiment on the effects of a two-tier punishment system', *Review of Law & Economics* .

Lim, C. S. (2013), 'Preferences and incentives of appointed and elected public officials: evidence from state trial court judges', *The American Economic Review* **103**(4), 1360–1397.

Merryman, J. H. & Pérez-Perdomo, R. (2007), *The civil law tradition: an introduction to the legal systems of Europe and Latin America*, Stanford University Press.

Nisbett, R. E. & Wilson, T. D. (1977), 'Telling more than we can know: Verbal reports on mental processes.', *Psychological review* **84**(3), 231.

Ouss, A. & Peysakhovich, A. (2013), 'When punishment doesn't pay:'cold glow'and decisions to punish', *Available at SSRN 2247446* .

Parisi, F. (2002), 'Rent-seeking through litigation: adversarial and inquisitorial systems compared', *International Review of Law and Economics* **22**(2), 193–216.

Posner, R. A. (1993), 'What do judges and justices maximize?(the same thing everybody else does)', *Supreme Court Economic Review* pp. 1–41.

Randazzo, K. A. (2008), 'Strategic anticipation and the hierarchy of justice in us district courts', *American Politics Research* .

Rizzolli, M. & Saraceno, M. (2013), 'Better that ten guilty persons escape: punishment costs explain the standard of evidence', *Public choice* **155**(3-4), 395–411.

Rizzolli, M. & Stanca, L. (2012), 'Judicial errors and crime deterrence: Theory and experimental evidence', *Journal of Law and Economics* **55**(2), 311–338.

Schlag, K. H. et al. (2013), 'Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality'.

Scott, K. M. (2006), 'Understanding judicial hierarchy: Reversals and the behavior of intermediate appellate judges', *Law & society review* **40**(1), 163–192.

Shapiro, M. (2013), *Courts: a comparative and political analysis*, University of Chicago Press.

Shavell, S. (1995), 'The appeals process as a means of error correction', *Journal of Legal Studies* **24**, 379–426.

Shavell, S. (2006), 'The appeals process and adjudicator incentives', *Journal of Legal Studies* **35**, 1–29.

Shavell, S. (2007), 'Optimal discretion in the application of rules', *American Law and Economics Review* **9**(1), 175–194.

Shavell, S. (2010), 'On the design of the appeals process: The optimal use of discretionary review versus direct appeal', *The Journal of Legal Studies* **39**(1), 63–108.

Sonnemans, J. & van Dijk, F. (2012), 'Errors in judicial decisions: experimental results', *Journal of Law, Economics, and Organization* **28**(4), 687–716.

Spitzer, M. & Talley, E. (2000), 'Judicial auditing', *The Journal of Legal Studies* **29**(2), 649–683.

Zajc, K. (2011), 'What do the european judges strive for-an empirical assessment', *International Journal for Court Administration* **3**(2), 3–19.

Zamir, E. & Ritov, I. (2012), 'Loss aversion, omission bias, and the burden of proof in civil litigation', *The Journal of Legal Studies* **41**(1), 165–207.

Zorn, C. J. & Bowie, J. B. (2007), 'An empirical analysis of hierarchy effects in judicial decision making', *Available at SSRN 987862* .