Soc Just Res (2010) 23:290–307
DOI 10.1007/s11211-010-0120-5

Automatic Judgment and Reasoning About Punishment

Margit E. Oswald · Ingrid Stucki

Published online: 3 December 2010

Abstract Several studies provide evidence that judgments on punishment are influenced by variables that are more or less independent of guilt considerations. It is postulated that these so called extralegal variables, such as the victim's reputation or outcome severity that occurs accidentally and without intention by the offender, in particular influence judgments that are made under restricted cognitive capacity (low processing depth). Two studies, using a vignette methodology, explore whether participants are able to correct the biasing influences of extralegal variables if they are motivated to elaborate their judgments under the most optimal conditions (high processing depth). Study 1 investigates the influence of victim's reputation, and Study 2 the combined influence of victim's reputation and accidentally occurring outcome severity under either low or high depth of information processing. Results show that the influence of extralegal variables can be corrected. However, corrections are either limited or excessive, and are sometimes even inappropriate.

Keywords Automatic judgment · Punishment · Extralegal variables · Bias correction · Overcorrection · Severity effect · Reputation of victim

Lay people's judgments about punishment are often biased. They are, for example, guided by a so called "severity effect" (Robbennolt, 2006), meaning that punishment is influenced by the severity of the outcome of an accident or an offense without sufficient considerations as to whether or not the culprit is responsible for it. Although classical theories of moral reasoning (Heider, 1958;

M. E. Oswald (✉) · I. Stucki
Department of Psychology, University of Berne, Unitobler-Muesmattstrasse 45,
3000 Berne 9, Switzerland
e-mail: margit.oswald@psy.unibe.ch

Kohlberg, 1969; Malle, 2006; Piaget, 1965; Shaver 1985; Weiner, 1995) presuppose that lay people normally distinguish effects that are produced intentionally from those that are unintended, and that their punishment is guided by considerations of culpability, such as intention and controllability by the actor, numerous studies over the last decade are challenging this view. These studies demonstrate that the amount of damage or injury resulting from a violation of norms influences the severity of punishment that is advocated even if the defendant caused the damage only indirectly (Greene & Darley, 1998), or if at least some portion of the total damage/injury was neither intended, nor could have been anticipated (Rucker, Polifroni, Tetlock, & Scott, 2004), or if its occurrence or non-occurrence was mainly accidental (Oswald, Orth, Aeberhard, & Schneider, 2005).

Lay people's judgments about punishment are, however, also biased by the influence of several other variables, such as the reputation of the victim, even if the offender could not have had any knowledge of this reputation (Mazzocco, Alicke, & Davis, 2004), the offender's ethnicity (ForsterLee, ForsterLee, Horowitz, & King, 2006; Sommers & Ellsworth, 2000, 2001), his or her gender (Rodriguez, Curry, & Lee, 2006) and attractiveness (Stewart, 1980; Zebrowitz, & McDonald, 1991). Such systematic influences have also been called "extralegal" since they put either the victim or the defendant at a disadvantage, and because they violate legal rules or ethical principles (cf. Vidmar, 2002).

Fair and just decisions about the punishment of an offender are central aims not only of the criminal court, and of penal legislation, but also of society in general (cf. De Keijser & Elffers, 2009). In order to improve decisions about punishment, it is, therefore, an important question as to whether, and under what conditions, people avoid biases, or are willing and able to correct their judgments. The research on correction of judgmental biases has been influenced crucially by Devine's (1989) observation that, in the intergroup context, prejudices and stereotyping occur automatically, and may be corrected only afterwards. Subsequently, there has been a vivid debate about the conditions and mechanisms of correction of judgmental biases, concerning both the inhibition of automatic processing and the correction of automatically triggered biases (for an overview cf. Bodenhausen, Todd, & Richeson, 2009).

Skitka, Mullen, Griffin, Hutchinson, and Chamberlin (2002) were able to show that participants' motivation to elaborate on automatically generated moral judgments varies with their value orientations. They refer to the often demonstrated result that liberally oriented individuals tend to attribute the problematic behavior of other persons (e.g., crime, poverty, and diseases for which the person is herself or himself responsible) more externally, whereas those who are more conservatively oriented tend to attribute these phenomena more internally. In their own research, Skitka et al. (2002) showed that, in their automatically generated judgments (generated under high cognitive load), both liberally and conservatively oriented participants tended to refuse payment for medical treatment of a disease if the patient her or himself is to be blamed for it. However, the liberally oriented participants seem to show a higher motivation to correct this decision than conservatives. When the additional cognitive load is absent the liberals are significantly more willing to grant payment for medical treatment. But individual

attitudes are not the only sources of differences in motivation to elaborate initial judgments. Specific situations may also induce such motivation. According to Tetlock (1989, 2002) people are much more likely to apply correcting strategies spontaneously if they expect real consequences, or if they know that they may be held responsible for their judgments on punishment. Thus, if subjects are accountable for their judgment, they might be especially motivated to correct their decision in such a way as to produce a closer correspondence to their moral values (see also Tetlock et al., 2007). Gilbert (1995) was also able to demonstrate that subjects do correct a fundamental attribution error, i.e., they more extensively take into account external influences upon behavior and they are less inclined to infer from observed behavior to basic dispositions, if they have the capacity to act on deliberative motivation. Finally, Lieberman (2002) showed that subjects who had to take the role of jurors in an action for compensation did not show more leniency as a function of the extralegal variable "attractiveness of the defendant" if they were reminded about applying a rational mode of information processing.

The empirical findings on the correction of judgment biases are not uniformly encouraging, however, since suppression of thoughts (Payne, 2005) as well as subsequent correction of biases (Fleming, Wegener, & Petty, 1999; Wegener & Petty, 1997; Wegener, Kerr, Fleming, & Petty, 2000) may ironically lead to undesired results. Thus, self-awareness, ethical standards, and social pressure may well motivate people to correct possible judgmental biases. In addition to a motivation to elaborate one's own judgments, for a successful correction of biases it seems to be necessary that persons do have knowledge of the fact that their judgment has been influenced by extralegal variables (Cacioppo, Petty, Feinstein, & Jarvis, 1996). However, individuals may not always have a clear idea or picture of what has caused their judgment, and they have to rely on their naive theories to determine the direction and magnitude of any potential bias (Bargh, 1999; Nisbett & Wilson, 1977). Thus, unless perceivers are skillful enough to identify precisely the kind and magnitude of any bias, it may happen that they are mistaken, and correct for the wrong bias on the wrong dimension, or undercorrect or even overcorrect their initial decision (Wegener & Petty, 1995, 1997). Hence, it is not surprising if the simple instruction in court that lay judges should judge in a fashion as fair and unbiased as possible is very often not sufficient to ensure avoidance of judgmental biases (Lieberman & Sales, 1997; Lord, Lepper, & Preston, 1984; Steblay, Hosch, Culhane, & Mc Wethy, 2006; Tanford & Cox, 1988). Judicial judgments may be especially problematic because there are in general not one but several sources of judgmental biases. Here, it may happen that persons are well able to avoid a stereotyping judgment based on the most salient cues, but they could still be influenced by other biasing variables. Blair, Judd, and Chapleau (2004) demonstrated, for example, that nowadays the skin color of the defendant barely influences degree of punishment, but that raters are still liable to unconscious race stereotyping. Racial stereotyping in sentencing is now based more on the facial appearance of offenders. Be they White or Black, offenders who possess more Afrocentric features (e.g., dark skin, wide nose, and full lips) receive harsher sentences for the same crimes than offenders less Afrocentric in appearance (Blair et al., 2004, p. 678).

Overview of the Present Research

Two studies reported here examine how and under what conditions extralegal variables will influence either automatically formed or more elaborated punishment decisions. We examine whether the influence of variables, such as the victim's reputation, or the severity of an outcome that was unintended by the offender, are corrected if participants are motivated to elaborate their judgment on punishment. It is the central aim of both studies to determine whether participants will successfully correct the biasing influence of extralegal variables when they have *optimal conditions* to elaborate their judgment. Thus, we compare, on the one hand, judgments on punishment generated under restricted cognitive capacity and, on the other hand, judgments generated not only with sufficient cognitive capacity but also under conditions of self-awareness and accountability.

In the first study, only one extralegal variable is manipulated. Thus, it should be relatively easy for participants to identify the biasing influence. Even if the unwarranted influence occurs unconsciously it should be relatively easy to access a correct naive theory of what might possibly have had a biasing influence. It is assumed that the extralegal variable will bias the judgment if participants have to judge automatically because they are doing so under conditions of restricted cognitive capacity, but that this biasing influence will be corrected if participants are motivated to come up with a correct judgment. In the second study, the judgment is more complicated because the combination of two extralegal variables is manipulated. Up to now, it is an unanswered question as to whether participants will correct any biasing influence if more biasing variables are involved, or whether they will focus only on one of them. In cases where participants are motivated to elaborate their judgment, it is, therefore, of additional interest whether people will spontaneously correct only for the most obvious influences of extralegal variables, or whether they will exhaustively scrutinize all possible biases, and correct them accordingly.

Among extralegal variables, we manipulated first the reputation of the victim (Study 1), and second both the reputation of the victim and the extent of injury that was not intended by the offender but happened more or less accidentally (Study 2). Severity of injury probably has considerable significance for an intuitively formed moral judgment since it triggers intuitions of injustice and threat by others (Mazzocco et al., 2004).

Study 1

In Study 1, we examined whether judgments about punishment are influenced by the victim's reputation. Although it is often considered morally unjustified to punish an offender on the basis of the victim's reputation (Mazzocco et al., 2004), we anticipated that participants would nevertheless be influenced by the victim's reputation if their judgment was made automatically, under restricted cognitive capacity (low processing depth). If automatic judgments are guided by emotional reactions, then victims with a good rather than a poor reputation should trigger more

empathy and a greater desire to defend or seek redress for the victim. However, correction of an intuitive judgment was predicted when participants were motivated to elaborate their judgment, e.g., under conditions of accountability (high processing depth). Consequently, we contrasted judgments about punishment under conditions of low information processing depth and conditions of high information processing depth. We chose a crime (presented in a vignette) in which the reputation of the victim was varied, but would be unknown to the offender.

Method

Participants

Students of psychology ($N = 77$) at the University of Berne participated for credits in their introductory psychology courses. The sample consisted of 60 women (78%) and 17 men (22%) who were between 18 and 41 years old ($M = 21.97$, $SD = 3.63$).

Scenarios

A vignette was presented to the participants in which a victim is assaulted. The victim is on his way home from work when a young man attacks him, hitting him, and stealing his wallet. The victim has to go to hospital with concussion and spends the night there (see Appendix for the vignettes).

Design and Measures

The study has a 2 (victim reputation: good vs. bad) $\times$ 2 (processing depth: low vs. high) between-subjects factorial design. The victim's reputation was manipulated as follows. In the good reputation condition, the victim is a physician, described as a decent person with basically positive attributes. In the bad reputation condition, the victim is a small-time criminal, described as having basically negative attributes. All other information about the offender and the offense was kept identical. Participants listened to the case history over headphones.

Depth of processing was manipulated by means of cognitive capacity, self-awareness, and accountability. Participants in the low processing depth condition had to solve a dual-processing task while they listened to the case story. Those in the deep processing condition had no dual task to solve and were additionally made more self-aware and accountable for their judgment. They were informed that they would have to justify their final decision about the punishment, and that their final statement would be videotaped.

Punishment was measured with five items incorporating 7-point Likert rating scales (Cronbach's alpha: .80). In the first item, participants were asked about how severely the offender should be punished for his offence (1 = not at all, 7 = very severely). Then, they were asked to give their views of a relatively harsh punishment (2 years imprisonment), and of a relatively lenient punishment (4 weeks work of benefit to the public), on a 7 point rating scale (1 = much too

lenient, 7 = much too harsh), and of their perception of the adequacy of the respective punishments (1 = not at all adequate, 7 = completely adequate).

The perceived severity of the victim's injury was assessed by means of a three-item scale (Cronbach's alpha: .69). Participants were asked about how severe they estimated the material, the physical, and the psychological damage suffered by the victim as a result of the offence. Quality of recall was measured by means of a recognition test. Forty items (statements about what happened in the story) were presented to the participants, only half of the statements being true. True statements asked, for example, whether the offender stole a wallet, or whether the victim had to spend a night in the hospital. False statements asked, for example, whether the offender extorted money, whether he was married, or whether the victim had to undergo surgery.

Procedure

Instructions to the participants were given by computer. They were all told that they would hear a story over headphones. Participants in the low processing depth condition pressed the left key every time the letter "K" or a red letter appeared on the screen, and pressed the right key if any other letter appeared. Letters remained on the screen for 1.5 s. Participants in the high processing depth condition listened undisturbed to the story and were additionally instructed that they would have to account for their judgments about the offence, and that their arguments would be recorded on a videotape, and would be analyzed later on by experts. After the participants had heard the story, they completed a questionnaire on the computer. Finally, they performed the recognition test on the computer. Statements about the story appeared on the screen, and they had to decide as quickly as possible whether the statements described something mentioned in the story, or not.

Results

The manipulation of processing depth was successful. The number of correctly identified items in the recognition test was lower in the dual-task condition ($M = 31.76$, $SD = 3.20$) than the accountability condition ($M = 35.85$, $SD = 2.58$), as indicated by a significant main effect of processing depth, $F\ [1, 75] = 42.21$, $p < .001$, $\eta^2 = .36$. This effect was moderated neither by the gender of participants $F\ [1, 73] = .34$, $p = .56$, $\eta^2 = .01$, nor by the victim's reputation $F\ [1, 72] = .358$, $p = .55$, $\eta^2 = .01$. Additionally, we were able to verify the success of this manipulation by comparing reaction times for responses to the questions assessing the dependent variables; these also differed significantly ($M_{\text{low processing depth}} = 10.90$ s.; $M_{\text{high processing depth}} = 13.70$ s.), $F\ [1, 75] = 6.09$, $p < .05$, $\eta^2 = .08$. Again, this effect was not moderated by the gender of participants or victim reputation.

Table 1 displays means and standard deviations of punishment in the different conditions.

Table 1 Mean judgments of punishment as a function of the reputation of the victim and processing depth

Victim's reputation	Low processing depth			High processing depth		
	Mean	*SD*	*n*	Mean	*SD*	*n*
Good	5.02	.77	18	3.98	.97	20
Bad	4.57	1.07	19	4.75	.89	20

Note: Scores range from 1 to 7

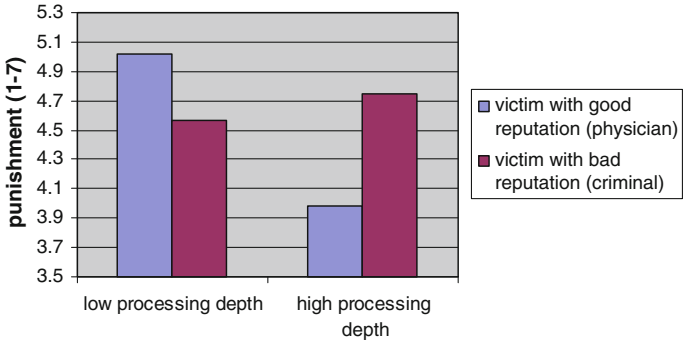

Fig. 1 Mean punishment for an offender as function of victim's reputation and depth of information processing

Recommended punishment was more severe in the dual-task condition than in the accountability condition, as indicated by a significant main effect of *processing depth,* F [1, 73] = 4.07, $p < .05$, $\eta^2 = .05$. Additionally, as shown in Fig. 1, there was a significant interaction between *victim reputation* and *processing depth,* F [1, 73] = 8.30, $p < .01$, $\eta^2 = .10$. The punishment recommended for the "good" victim was harsher in the low-depth condition than the high-depth condition F [1, 36] = 13.32, $p < .01$, $\eta^2 = .27$, whereas recommended punishment did not differ significantly for the "bad" victim F [1, 37] = 3.71, $p < .55$, $\eta^2 = .01$. Under conditions of low processing depth, harsher punishments were recommended in case of the "good" victim than of the "bad" victim, F [1, 36] = 4.02, $p < .05$, $\eta^2 = .07$, but under conditions of high processing depth harsher punishments were recommended in case of the "bad" victim than of the "good" victim, F [1, 36] = 6.88, $p < .05$, $\eta^2 = .15$.

Perceived injury to the physician ($M = 5.05$; $SD = 1.01$), and to the small-time criminal ($M = 4.72$; $SD = 1.05$) did not differ significantly. Thus, the reputation of the victim did not affect perceived injury, F [1, 73] = 2.01, $p = .15$. Interestingly, the injury perceived to be suffered by the victim was significantly more severe in the dual-task condition ($M = 4.86$; $SD = .94$) than in the accountability condition ($M = 4.11$; $SD = .93$), F [1, 73] = 4.37, $p < .001$, $\eta^2 = .14$. This influence was not moderated by the gender of the participants nor by the victim's reputation.

Discussion

In this study, we found that only under conditions of low processing depth (dual-task condition) would participants recommend harsher punishment if the victim had a good rather than a bad reputation. If subjects experienced no additional cognitive load, and if they expected they would have to justify their judgment in front of a video camera, their punishment decisions were actually even more lenient if the victim had a good rather than a bad reputation. Thus, we can conclude that people automatically prefer harsher punishment for an offender if the victim has a good rather than a bad reputation, but that this preference can be changed if an opportunity to elaborate the judgment is provided. But, in the latter case it is not easy to explain why the condition enabling higher depth of processing did not really have a de-biasing effect. What we actually found was a reverse influence rather than the disappearance of any influence of the extralegal factor *victim reputation*. Interestingly, the punishment was not harsher for an offender who had attacked a small-time criminal but was more lenient for an offender who had attacked a physician. Thus, participants in the high depth of processing condition seemed to be more concerned about punishing too harshly than about punishing too leniently. But why then do participants recommend even more lenient punishment if the victim had a good rather than a bad reputation? According to Wegener and Petty (1995, 1997) participants probably realized the nature of the likely bias in their judgment but had difficulties calibrating its magnitude, leading them to overcorrect their initial decision.

Furthermore, we found that level of processing depth influenced the punishment recommended. Participants preferred harsher punishments under restricted (dual task) cognitive capacity conditions than under conditions of high-depth processing. One reason for the effect of processing depth on recommended punishment might be that the injury suffered by the victim was perceived to be less serious under greater processing depth conditions than with lesser depth of processing. The perceived severity of the assault was, however, not influenced by the reputation of the victim.

Study 2

It was the aim of this second study to gain somewhat more insight into the ability of people to correct judgmental biases if they are motivated to come up with an unbiased judgment. In this study, we increased the difficulty of the task by manipulating not just one extralegal variable but the combined influence of two of them. In this case, it should be more difficult for participants to have exact knowledge about the biasing influence, and we wondered whether the motivation to elaborate their judgment would be sufficient to correct the influence of both extralegal variables. We manipulated the victim's reputation but also outcome severity that occurred unintended by the offender, and examined their effects on judgments of punishment under different processing depth conditions.

Method

Participants

Students of psychology at the University of Berne participated for credits in their introductory psychology courses. The sample consisted of 106 women (73.6%) and 38 men (26.4%) who were between 18 and 48 years old ($M = 22.42$; $SD = 4.60$).

Scenario

A similar vignette to that used in Study 1 was presented to the participants: the victim is again on his way home from work when a young man attacks him, hitting him and stealing his wallet. As in Study 1 section, the victim has to go to hospital with concussion and spend the night there. Additionally, this time the victim goes to the toilet (during the night in the hospital), stumbles and falls down with consequences that vary in severity across conditions (see Appendix for the vignettes).

Design and Measures

The Study has a 2 (outcome severity: low vs. high) × 2 (processing depth: low vs. high) × 2 (victim reputation: good vs. bad) between-subjects factorial design. The severity of the outcome that was unintended by the offender varied in terms of the consequences of the victim's fall in the hospital: either the victim was unhurt, or broke a leg and had to endure a complicated operation because of the fracture. Thus, the manipulated severity of the outcome was absolutely accidental and had nothing to do with the offender's attack. As in Study 1, processing depth was varied by cognitive capacity, self-awareness, and the motivation to make a just decision. The victim's reputation was varied in a manner analogous to Study 1. In one condition, the victim is a physician, and in the other a small-time criminal. All other information about the offender and the offense was identical across conditions. Punishment was again measured with a five-item rating scale (Cronbach's alpha: .75). Additionally, we assessed the probability attributed by subjects to different possible outcomes of the offense (fall without harm, and fall with fracture of the leg) on a 7-point Likert scale. The perceived harm to the victim (Cronbach's alpha: .62) as well as the accuracy of memory were measured identically to Study 1. However, the memory test included four additional statements to capture in particular the new content of the scenarios describing the variations in *outcome severity*. The whole procedure of the Study remained entirely the same as in Study 1.

Results

The manipulation of *unintended outcome severity* was successful. The injury was judged to be significantly more severe if the fall resulted in a broken leg ($M = 5.00$; $SD = .76$) than if it did not ($M = 4.42$; $SD = .68$), F [1, 143] = 23.02, $p < .001$,

$\eta^2 = .14$. This effect was moderated neither by gender of participants nor by victim reputation. The effect of processing depth was verified on the basis of correctly identified items in the recognition test; the scores differed significantly between the levels of processing, F [1, 144] = 33.70, $p < .001$, $\eta^2 = .19$. Out of a total of 44 items, participants responded correctly on average to 35.91 items ($SD = 3.77$) in the dual-task condition, and to 39.09 items ($SD = 2.80$) in the accountability condition. This effect was moderated neither by participants' gender, nor by victim's reputation, but was moderated by outcome severity. However, after excluding the four recognition items that refer to the different outcomes of the victim's fall in the hospital, the influence of *outcome severity* on the recognition test disappeared, F [1, 143] = .311, $p = .58$), while the effect of *processing depth* remained significant, F [1, 143] = 75.19, $p < .001$, $\eta^2 = .34$. Thus, outcome severity does not influence recognition in general. Only statements about the victim's broken leg were more accurately recalled in the high severity condition than were the items about the victim's fall that had no consequences in the low severity condition. Additionally, we found that participants perceived the high and low severity outcomes as equally probable. Thus, we can exclude the possibility that punishment recommendations are influenced by any tendency for participants to attribute different probabilities of occurrence to low versus high outcome severity.

Table 2 displays means and standard deviations for punishment.

There was no three-way interaction. Recommended punishment was harsher in the "good" victim condition ($M = 4.31$, $SD = .83$) than in the "bad" victim condition ($M = 3.95$, $SD = .89$), as indicated by a significant main effect of *victim reputation*, F [1, 137] = 6.81, $p < .05$, $\eta^2 = .05$. Although we could not replicate the interaction between victim's reputation and processing depth found in Study 1, an additional analysis of contrasts did show that the main effect of victim reputation is mainly associated with low processing depth. In the low-depth condition, the punishment recommended for the offender who assaulted a "good" victim was harsher ($M = 4.32$; $SD = .82$) than for the offender who assaulted a "bad" victim ($M = 3.84$; $SD = .65$), as indicated by a significant main effect, F [1, 68] = 6.02, $p < .05$, $\eta^2 = .08$, but in the high-depth condition the recommended punishment did not differ significantly as between offenders who had assaulted either a "good" victim ($M = 4.30$; $SD = .84$) or a "bad" victim ($M = 4.05$; $SD = 1.07$). Thus, in the high-depth condition participants seemed to avoid the biasing influence of *victim*

Table 2 Mean judgments of punishment as a function of unintended outcome severity, processing depth, and reputation of the victim

Victim's reputation	Outcome severity	Low processing depth			High processing depth		
		Mean	*SD*	*n*	Mean	*SD*	*n*
Good	High	4.49	.91	17	4.12	.91	19
	Low	4.12	.71	20	4.52	.76	20
Bad	High	4.14	.38	17	4.19	1.13	17
	Low	3.55	.74	16	3.90	1.03	19

Note: Scores range from 1 to 7

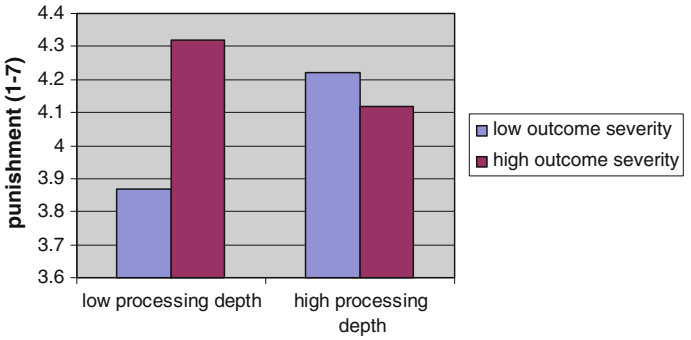

Fig. 2 Mean punishment recommended for an offender as a function of unintended outcome severity and depth of information processing

reputation more than in the low-depth condition. But this time the correction was more modest, and we could not find a tendency to overcorrect the bias, as we did in Study 1. Furthermore, we found a marginally significant interaction between *outcome severity* and *processing depth* (F [1, 137] = 3.62, $p = .059$, $\eta^2 = .03$). In the low-depth condition, the suggested punishment was harsher if the victim suffered more harm because of his fall in the hospital, F [1, 69] = 6.88, $p < .05$, $\eta^2 = .09$, whereas in the high-depth condition the participants recommended the same punishment, independent of whether the victim suffered more or less severely as a consequence of his fall in the hospital. However, this convergence in recommended punishment occurred not only because participants reduced their punishment in the high severity condition, but also because they increased their recommended punishment in the low severity condition.

Taken as a whole, results of Study 2 show that the influence of the victim's reputation as well as that of unintended outcome severity biased the judgment in the low-depth condition as expected, and that this influence is largely absent in the high-depth condition. Whether the recommended punishment in the high-depth condition is still biased or not seems to be a separate question, however. The results displayed in Fig. 2 are combined across the levels of victim's reputation.

Discussion

In Study 2, we manipulated not only the reputation of the victim, as in Study 1, but also the outcome severity of the offense, to make the judgment more complex. The manipulation of outcome severity was devised to exclude direct responsibility on the part of the offender. The victim, who had been assaulted by the offender, suffered additionally, or not, as a result of a fall in hospital (a broken leg vs. no consequences). Thus, we manipulated two extralegal factors such that a punishment recommendation should depend neither on the specific reputation of the victim nor on the degree of injury to the victim that could not be controlled by the offender. The results show that generally harsher punishment was recommended if the offender's victim had a good rather than a bad reputation. However, we did not find, as in Study 1, any

significant interaction between victim's reputation and processing depth, although an additional analysis of contrasts did show that the main effect of the victim's reputation was mainly associated with low-depth processing. On the other hand, both studies show that under conditions of higher information processing depth participants do not punish an offender significantly more harshly simply because he had assaulted a victim with a good rather than with a bad reputation. But in Study 1, avoidance of such a bias actually resulted in an overcorrection, while no such overcorrection could be found in Study 2. Furthermore, we were able to corroborate a marginally significant interaction between outcome severity and processing depth in Study 2. In the low-depth condition, we found some proof of an extralegal severity effect, because the offender was punished more harshly if the victim had additionally suffered because of a broken leg than if there was no additional injury. Under conditions of higher processing depth, however, the different outcome severities attracted almost equivalent punishment recommendations. Interestingly, the convergence of punishment in the high depth as compared to the low-depth condition does not result entirely from more lenient punishment recommendations. When participants have the opportunity to elaborate their judgment, they recommend greater leniency in the case of a significant but unintended injury to the victim, but also greater severity, if the victim's fall results in no further injury.

The increase in punishment in the high-depth condition with a low severity outcome is not easily explained since elaboration of the judgment should clarify that the victim's fall in the hospital was not intended by the offender, and thus should not influence the punishment at all. However, it is possible that the elaboration also leads the participants to think about what could have happened to the victim as a result of his fall, even though it had no consequences in this case. Study 2 also shows that participants do take care to make certain corrections of their automatic judgments, but that these corrections do not entirely comply with the normative rules for the avoidance of judgmental biases.

General Discussion

Considering the results of both studies, they clearly confirm that judgments about punishment differ considerably between conditions of low (restricted cognitive capacity) and high processing depth (accountability and self-awareness). Both studies successfully replicated previous findings that harsher punishment will be advocated for an offender if his victim has a good rather than a poor reputation. This is, however, primarily true under conditions of limited processing depth. If people are motivated to elaborate their punishment recommendations (high processing depth condition) they are, in principle, willing to correct their automatic judgment. The nature and magnitude of the correction of the extralegal influence of victim reputation under conditions of greater information processing, however, is not necessarily successful. In Study 1, the bias was overcorrected such that even harsher punishment was recommended if the offender had assaulted a petty criminal (poor reputation), and not an honorable physician (good reputation). But in Study 2, the influence of reputation was merely reduced as indicated by the main effect of

victim's reputation that was mainly associated with the low-depth condition. Nevertheless, in Study 2 the effect of the interaction between victim reputation and processing depth on recommended punishment was not significant, in contrast to Study 1. How can we explain the different degrees of correction of the influence of reputation across the two studies? One possible explanation could be that in Study 2 it was not only victim reputation that was manipulated as an extralegal variable but also injury to the victim that was unintended by the offender (outcome severity). Therefore, it may be that in Study 2 participants' efforts to correct their judgments were divided between the two extralegal variables. The marginally significant interaction between processing depth and outcome severity seems consistent with this possibility. With respect to outcome severity, we found that participants in the low-depth condition proposed harsher punishment the more severe, albeit unintended, the outcome of his offense was (severity effect), but that in the high-depth condition this effect was attenuated. Thus, participants tried to avoid the biasing influence of both extralegal variables, namely unintended outcome severity and reputation. However, correction of the severity effect was again not what one should expect if normative standards were being respected. Since the victim's fall in the hospital was neither intended by nor under the control of the offender, it should not significantly influence the recommended punishment one way or the other. But what we found was that the correction was not purely due to a more lenient punishment in the high-depth condition. While a decrease in punishment was observed in the high outcome severity condition, the recommended punishment in the low outcome severity condition was now even harsher in the high-depth processing condition than in the low-depth processing condition. One can only speculate why participants in the high-depth condition increased their punishment when the victim had no additional suffering to bear. They might have thought about what else could have happened to the victim and these conditional scenarios might have aggravated their recommended punishment.

In summary, participants can be motivated to correct biasing influences on their judgment. However, these corrections seem to be quite limited. They either go too far, or not far enough, and are partly even erroneous. Although it is quite easy for people to become aware of the fact that an offender should neither be punished too severely because the victim happened to be an honorable physician, nor too leniently because the victim happened to be a small-time criminal, as in Study 1 where only the victim's reputation was manipulated, they nonetheless lack a frame of reference for the appropriate amount of punishment to be meted out (cf. Wegener & Petty, 1995, 1997). In this case, their correction of the initial inclination may go beyond the target, as the overcorrections indicate. If the detection of biasing influences becomes more complex, as in Study 2, where two biasing influences were manipulated, participants correct insufficiently for each of the two variables, and to some extent not in the right direction. They may either focus only on the most salient influence and conclude their critical reasoning process as soon as one correction is accomplished, or they may mistakenly correct on the wrong dimension, as when recommending harsher punishment when the unintended outcome severity was low. Taken together, cognitive capacity to act on accountability does not always de-bias judgments or elicit fairer judgments as e.g. Tetlock et al. (2007) assume.

Further studies on de-biasing conditions of punishment decisions are important, and in the interests of greater ecological validity should continue the approach of expanding on the number of biasing influences from just one to more extralegal variables. They should also, however, focus more on the meta-cognitions of the participants, in order to uncover what they are thinking while elaborating their judgments. If judgments about appropriate punishment are to be improved, it is important to determine whether participants are aware of extralegal factors at all, and why they are correcting their judgments in one and not another direction. Additionally, although it was the central aim of both studies to discover whether people are able to successfully correct for the influence of extralegal variables under *optimal conditions*, future studies should differentiate more between effects of an elaboration that are either due to sufficient cognitive capacity, to self-awareness, or to accountability.

Acknowledgment The research was part of a project supported by the Swiss National Science Foundation (No. 101411-101758).

Appendix: Vignettes

Vignettes in Study 1:

Condition 1: "good" Victim

Thomas R is a heart surgeon. He works in a large university hospital and conducts several complicated operations every day. He is married and has three children. In his free time he likes to hike with his family, and is member of the Swiss Alps Club. He lives in a nice house at the periphery. His neighbors describe him as a person who works a lot. They say he is not a great talker but always helpful. The evening of September 22nd, he has had a busy day during which he had to conduct two operations and to attend several emergencies. So he is quite tired when he goes home, and looks forward to going to the cinema with his family.

As usual, he walks from the bus stop to his house, and he is glad that the sun is shining this day. Shortly before arriving at his house, he is suddenly attacked by a young man. The man knocks him down and steals him his wallet. The offender flees. Thomas R. remains lying injured on the floor until some passers-by help him. He has to go to the hospital with a concussion and some bruises, and has to spend the night there. A little later the offender is caught because of the descriptions of the passers-by. He is a 25 year old Swiss who confesses the offence shortly after his detention. He says that he wanted to make money quickly and that he chose his victim at random.

Condition 2 "bad" Victim

Lukas K is a small-time criminal who earns his living by means of frauds. His favored method is to sell insurances which don't exist. To do so, he

impersonates an insurance agent and promises his victims cheap insurances. The premiums and advance payments end up in his private account. His victims are mostly older ladies who live in modest financial circumstances. They lose all their savings because of the fraud. Lukas lives in a nice house at the periphery. His neighbors describe him as communicative but not very cooperative when help is needed. He has a son with his ex-girlfriend but doesn't care much about him. On the evening of September 22nd, he is once again on his way home from a successful fraud.

As always he walks from the bus stop to his house and he is glad that the sun is shining this day. Shortly before arriving at his house he is suddenly attacked by a young man. The man knocks him down and steals him his wallet. The offender flees. Lukas K. remains lying injured on the floor until some passers-by help him. He has to go to the hospital with a concussion and some bruises and has to spend the night there. A little later, the offender is caught because of the descriptions of the passers-by. He is a 25 year old Swiss who confesses the offence shortly after his detention. He says that he wanted to make money quickly and that he chose his victim at random.

Vignettes in Study 2:

Condition 1: High Outcome Severity (Good Victim)

Thomas R is a heart surgeon. He works in a large university hospital and conducts several complicated operations every day. He is married and has three children. In his free time he likes to hike with his family, and is member of the Swiss Alps Club. He lives in a nice house at the periphery. His neighbors describe him as a person who works a lot. They say he is not a great talker but always helpful. The evening of September 22nd, he has had a busy day during which he had to conduct two operations and to attend several emergencies. So he is quite tired when he goes home, and looks forward to going to the cinema with his family.

As usual, he walks from the bus stop to his house, and he is glad that the sun is shining this day. Shortly before arriving at his house, he is suddenly attacked by a young man. The man knocks him down and steals him his wallet. The offender flees. Thomas R remains lying injured on the floor until some passers-by help him. He has to go to the hospital with a concussion and some bruises, and has to spend the night there. During the night Thomas R. wakes up because he has to go urgently to the toilet. Because he is still quite dizzy as a consequence of the concussion, he stumbles and falls on the hard floor. He falls so badly that he gets a complicated fracture on his left thigh bone. He will be handicapped for several weeks and has to undergo a surgery. A few days later, the offender is caught because of the descriptions of the passers-by. He is a 25 year old Swiss who confesses the offence shortly after his detention. He says that he wanted to make money quickly and that he chose his victim at random.

Condition 2: Low Outcome Severity (Good Victim)

Thomas R is a heart surgeon. He works in a large university hospital and conducts several complicated operations every day. He is married and has three children. In his free time he likes to hike with his family, and is member of the Swiss Alps Club. He lives in a nice house at the periphery. His neighbors describe him as a person who works a lot. They say he is not a great talker but always helpful. The evening of September 22nd, he has had a busy day during which he had to conduct two operations and to attend several emergencies. So he is quite tired when he goes home and looks forward to going to the cinema with his family.

As usual, he walks from the bus stop to his house, and he is glad that the sun is shining this day. Shortly before arriving at his house, he is suddenly attacked by a young man. The man knocks him down and steals him his wallet. The offender flees. Thomas R remains lying injured on the floor until some passers-by help him. He has to go to the hospital with a concussion and some bruises, and has to spend the night there. During the night Thomas R wakes up because he has to go urgently to the toilet. Because he is still quite dizzy in consequence of the concussion, he stumbles and falls on the hard floor. Fortunately the fall has no further consequences for him. A few days later, the offender is caught because of the descriptions of the passers-by. He is a 25 year old Swiss who confesses the offence shortly after his detention. He says that he wanted to make money quickly and that he chose his victim at random.

Conditions 3 and 4: High and Low Outcome Severity (Bad Victim)

The same manipulations of high and low outcome severity were made for the vignettes with the small-time criminal (compare vignettes of Study 1).

References

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*(19), 674–679.

Bodenhausen, G. V., Todd, A. R., & Richeson, J. A. (2009). Controlling prejudice and stereotyping : Antecedents, mechanisms, and contexts. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 111–135). New York: Psychology Press.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in the need for cognition. *Psychological Bulletin, 119*(2), 197–253.

De Keijser, J. W., & Elffers, H. (2009). Public punitive attitudes: A threat to the legitimacy of the criminal justice system? In M. E. Oswald, S. Bieneck, & J. Hupfeld-Heinemann (Eds.), *Social psychology of punishment of crime* (pp. 55–74). Chichester: Wiley-Blackwell.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18.

Fleming, M. A., Wegener, D. T., & Petty, R. E. (1999). Procedural and legal motivations to correct for perceived judicial biases. *Journal of Experimental Social Psychology, 35*, 186–203.

ForsterLee, R., ForsterLee, L., Horowitz, I. A., & King, E. (2006). The effects of defendant race, victim race, and juror gender on evidence processing in a murder trial. *Behavioral Sciences & The Law, 24*, 179–198.

Gilbert, D. T. (1995). Attribution and interpersonal perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 99–147). New York: McGraw-Hill.

Greene, E. J., & Darley, J. M. (1998). Effects of necessary, sufficient, and indirect causation on judgments of criminal liability. *Law and Human Behavior, 22*, 429–451.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago: Rand McNally.

Lieberman, J. D. (2002). Head over the heart or heart over the head? Cognitive-experiential self-theory and extra-legal heuristics in juror decision-making. *Journal of Applied Social Psychology, 32*, 2526–2553.

Lieberman, J. D., & Sales, B. (1997). What social sciences teaches us about the jury instruction process. *Psychology Public Policy and Law, 3*(4), 589–644.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*(6), 1231–1243.

Malle, B. F. (2006). The relation between judgments of intentionality and morality. *Journal of Cognition and Culture, 6*, 61–86.

Mazzocco, P. J., Alicke, M. D., & Davis, T. L. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology, 26*, 131–146.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Oswald, M. E., Orth, U., Aeberhard, M., & Schneider, E. (2005). Punitive reactions to completed crimes versus accidentally uncompleted crimes. *Journal of Applied Social Psychology, 35*, 718–731.

Payne, B. K. (2005). Conceptualizing control in social cognition: How executive control modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology, 89*, 488–503.

Piaget, J. (1965). *The moral judgment of the child*. New York: Free Press.

Robbennolt, J. K. (2006). Outcome severity and judgments of "responsibility": A meta-analytic review. *Journal of Applied Social Psychology, 30*, 2575–2609.

Rodriguez, S. F., Curry, T. R., & Lee, G. (2006). Gender differences in criminal sentencing: Do effects vary across violent, property, and drug offenses? *Social Science Quarterly, 87*, 318–340.

Rucker, D. D., Polifroni, M., Tetlock, P. E., & Scott, A. L. (2004). On the assignment of punishment: The impact of general-societal threat and the moderating role of severity. *Personality and Social Psychology Bulletin, 30*, 673–684.

Shaver, K. G. (1985). *The attribution of blame*. New York: Springer.

Skitka, L. J., Mullen, E., Griffin, T., Hutchinson, S., & Chamberlin, B. (2002). Dispositions, scripts, or motivated correction? Understanding ideological differences in explanations for social problems. *Journal of Personality and Social Psychology, 83*(2), 470–497.

Sommers, S. R., & Ellsworth, P. C. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin, 26*, 1367–1379.

Sommers, S. R., & Ellsworth, P. C. (2001). White juror bias: An investigation of racial prejudice against Black defendants in the American courtroom. *Psychology, Public Policy, and Law, 7*, 201–229.

Steblay, N., Hosch, H. M., Culhane, S. E., & Mc Wethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis. *Law and Human Behavior, 30*, 469–492.

Stewart, J. E. (1980). Defendant's attractiveness as a factor in the outcome of criminal trials: An observational study. *Journal of Applied Social Psychology, 10*(4), 348–361.

Tanford, S., & Cox, M. (1988). The effects of impeachment evidence and limiting instructions on individual and group decision making. *Law and Human Behavior, 12*, 477–497.

Tetlock, P. E. (1989). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45*, 74–83.

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109*(3), 451–471.

Tetlock, P. E., Visser, P. S., Singh, R., Polifroni, M., Scott, A., Elson, B., et al. (2007). People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility. *Journal of Experimental Social Psychology, 43*(2), 195–209.

Vidmar, N. (2002). Case studies of pre- and midtrial prejudice in criminal and civil litigation. *Law and Human Behavior, 26*(1), 73–105.

Wegener, D. T., Kerr, N. L., Fleming, M. A., & Petty, R. E. (2000). Flexible corrections of juror judgments: Implications for jury instructions. *Psychology, Public Policy, and Law, 6*(3), 629–654.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology, 68*, 36–51.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. *Advances in Experimental Social Psychology, 29*, 141–208.

Weiner, B. (1995). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*, 548–573.

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' babyfacedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior, 15*, 603–623.