von Helversen, Bettina ; Rieskamp, Jörg

Predicting sentencing for low-level crimes: Comparing models of human judgment

Journal Article as: peer-reviewed accepted version (Postprint)

DOI of this document*(secondary publication): https://doi.org/10.26092/elib/3454

Publication date of this document: 08/11/2024

* for better findability or for reliable citation

Recommended Citation (primary publication/Version of Record) incl. DOI:

von Helversen, B., & Rieskamp, J. (2009). Predicting sentencing for low-level crimes: Comparing models of human judgment. Journal of Experimental Psychology: Applied, 15(4), 375–395. https://doi.org/10.1037/a0018024

Running head: COGNITIVE MODELING OF SENTENCING DECISIONS

Predicting Sentencing for Low-Level Crimes:

Comparing Models of Human Judgment

Bettina von Helversen

Max Planck Institute for Human Development; University of Basel

Jörg Rieskamp

University of Basel

Contact: Bettina von Helversen

University of Basel, Department of Psychology,

Missionsstr.62a, 4055 Basel, Switzerland

Phone: +41 61 2673528, Fax: +41 61 2670613

Email: Bettina.vonhelversen@unibas.ch

Authors' Note

Bettina von Helversen, Max Planck Institute for Human Development, Berlin, Germany; University of Basel, Switzerland. Jörg Rieskamp, University of Basel, Switzerland. We would like to thank Attorney M. Neff and the prosecution authority of Eberswalde for providing us access to the trial records. We are very grateful to Christoph Engel, Stefan Bechthold, Stefan Tontrupp, Andreas van den Eikel, and Tobias Lubitz for their help and advice in devising the coding system. We gratefully acknowledge Patrizia Ianiro, Daria Antonenko, and Cornelia Büchling's commitment and helpful ideas in coding and analyzing the data. We would like to thank Anita Todd for editing a draft of this manuscript. This work has been supported by a doctoral fellowship of the International Max Planck Research School LIFE to the first author and by a research grant (RI 1226/5) of the German Research Foundation to both authors. Correspondence concerning this article should be addressed to Bettina von Helversen.

Abstract

Laws and guidelines regulating legal decision making are often imposed without taking the cognitive processes of the legal decision maker into account. In the case of sentencing, this raises the question whether the sentencing decisions of prosecutors and judges are consistent with legal policy. Especially in handling low-level crimes, legal personnel suffer from high case loads and time pressure, which can make it difficult to comply with the often complex rulings of the law. To understand the cognitive processes underlying sentencing decisions, an analysis of trial records in cases of larceny, fraud, and forgery was conducted. Applying a Bayesian approach, five models of human judgment were tested against each other to predict the sentencing recommendations of the prosecution and to identify the crucial factors influencing sentencing decisions. The factors influencing sentencing were broadly consistent with the penal code. However, the prosecutors considered only a limited number of factors and neglected factors that were legally relevant and rated as highly important. Furthermore, testing the various cognitive judgment models against each other revealed that the sentencing process was apparently not consistent with the judgment policy recommended by the legal literature. Instead the results show that prosecutors' sentencing recommendations were best described by the mapping model, a heuristic model of quantitative estimation. According to this model, sentencing recommendations rely on a categorization of cases based on the cases' characteristics.

Keywords: legal decision making; heuristics; sentencing; quantitative estimation; multiple-cue judgment

Predicting Sentencing for Low-Level Crimes: Comparing Models of Human Judgment

How are criminal sentences determined? Although legal systems differ from country to country, judges worldwide struggle with the problem of determining which factors should be considered and how they should be combined to form appropriate and just sentences. Even if the legal system provides guidelines to regulate the sentencing process, the question remains how well judges and other legal personnel follow the prescribed policies (Ruback & Wroblewski, 2001). Research on sentencing has a long tradition of identifying deviations from legal policy: Extralegal factors such as ethnic background and gender have been found to influence sentencing, and in some cases legal factors are ignored (e.g., Davis, Severy, Kraus, & Whitaker, 1993; ForsterLee, ForsterLee, Horowitz, & King, 2006; Henning & Feder, 2005; Johnson, 2006; Ojmarrh, 2005). Thus, the cognitive processes of legal professionals can lead to decisions that are inconsistent with the policy specified by the law, be it for bail setting (Dhami & Ayton, 2001; Ebbesen & Konečni, 1975) or sentencing (Konečni & Ebbesen, 1982; 1984; Tata 1997; Van Duyne, 1987). Legal decision making might not live up to legal policy because human decisions are often guided by heuristics (Gigerenzer & Todd, 1999; Guthrie, Rachlinski & Wistrich, 2001). The goal of the present article is to examine the extent to which sentencing decisions can be explained by heuristic models and how this leads to deviations from legal regulations. To this end, we computationally model prosecutors' sentence recommendations, comparing heuristic models against traditional models of judgment and decision making. Based on the modeling we investigate the extent to which the sentences are consistent with legal policy.

In the following, we describe the legal approach to sentencing and then discuss how different cognitive approaches could be applied to explain the sentencing process.

Sentencing Decisions by the Prosecution

In the American system, the vast majority of cases are closed by plea-bargaining and thus crucially depend on the prosecution. In plea-bargaining, the prosecution and defense negotiate a sentence, which then is ratified by a judge. In contrast, in many other countries such as Germany plea-bargaining before a case goes to court is an exception to the rule, although informal arrangements prior to trial are possible, and in minor cases the prosecution can issue a penalty order, which may be considered a take-it-or-leave-it offer. However, in Germany the prosecution also plays an important role in determining the sentence magnitude in cases that go to court. As in most legal systems, a sentence is determined by a judge. However, the judge makes this decision after hearing sentencing recommendations from both the prosecution and the defense. Research has shown that the sentencing recommendation of the prosecution in Germany (Schünemann, 1988) or in similar decisions in the US by the probation officer (Konečni & Ebbesen, 1982; 1984) is the single most important factor influencing the judge's decision. Similarly, Englich and Mussweiler (2001) found that, all things being equal, the recommendation of the prosecution significantly influenced the sentence imposed. Further, Dhami and Ayton (2001) showed that in bail decisions, British magistrates followed the recommendation of the prosecution almost without exception (see also Ebbesen & Konečni, 1975). These findings indicate that our understanding of factors influencing a sentence's magnitude relies critically on first investigating the process by which the prosecution determines the sentence recommendation.

How should the prosecution go about making its recommendation? While the American system is an adversarial system with the prosecution representing one side and the defense the other, Germany's legal system is inquisitorial. In an adversarial system the prosecution and the defense try to convince a third impartial party, for example the jury, of their position. The jury judges the case, but is not actively involved in the investigation. In an inquisitorial system, the judge actively investigates the case drawing upon information from the prosecution and the defense. In this system, the prosecutions' task is not only to ensure a conviction, but also to objectively investigate the crime and to present the complete case in court. This also includes considering mitigating or exonerating information.

German sentencing is regulated by the German penal code (*Strafgesetzbuch*, *StGB*; Tröndle & Fischer, 2007), more specifically by Articles 21, 23, 46, 47, and 49, and by decisions of the German Federal Court of Justice. Both judge and prosecution are bound by the same legal regulations. The general goal is to determine an appropriate sentence that is proportional to the guilt of the offender. For each offense there exists a sentencing range that establishes the minimum and maximum sentence that can be imposed. Within this often rather broad sentencing range, the determination of the sentence depends on the seriousness of the case and is largely left to the judge's discretion. The judge's task, as well as the prosecution's, is to evaluate factors mitigating or aggravating guilt of the offender and to determine the sentence accordingly. Factors that can be considered mitigating or aggravating are specified in the penal code. Article 46 of the *StGB* alone lists over 20 factors that can be considered in the sentencing decision, although it cautions that this is not an exhaustive list.

The German penal code (§ 46) does not provide explicit guidelines on how the factors should be combined. However, the German Federal Court of Justice recommends that mitigating and aggravating factors be balanced in an integrative evaluation of the overall picture (Schäfer, 2001). According to the predominant opinion in the legal literature, this is best accomplished with a three-step sentencing process: All relevant factors are evaluated according to the direction of their effect on the sentence (aggravating or mitigating), then weighted by their importance, and finally added up to form the sentence (Bruns, 1985, 1988; Foth, 1985; Schäfer, 2001; but see Mösl, 1981, 1983; and Theune, 1985a, 1985b). This approach is consistent with a linear additive judgment strategy and thus with a long tradition of modeling human judgment in psychology (Anderson, 1981; Brehmer, 1994; Juslin, Olsson & Olsson, 2003).

Modeling Decision Making

Social Judgment Theory

Since the 1950s, the idea that human judgment follows a linear additive strategy has been a prominent approach in research on judgment and decision making (Hammond, 1955), culminating in the establishment of "social judgment theory" (SJT; for a review, see Doherty & Kurz, 1996). Based on the idea that judgments are linear additive, SJT has proposed that multiple linear regression can be employed to capture judgment policies and to determine which information has been utilized (e.g., Cooksey, 1996; Doherty & Brehmer, 1997). The SJT view has been broadly supported by research (for a review, see Brehmer, 1994) and has led to a proliferation of regression analysis to capture human judgment policies in many areas of applied research such as medical decision making (e.g., Harries & Harries, 2001), personnel psychology (e.g., Zedeck & Kafry, 1977) and educational psychology (e.g., Cooksey, Freebody & Davidson, 1986). Also in legal decision making, regression analysis has become the standard procedure for analyzing sentencing policies (Engen & Gainy, 2000; Johnson, 2006; Kautt, 2002; Kautt & Spohn, 2002) and a benchmark for all models of judgments.

However, there is evidence that the cognitive process underlying human judgment does not always follow a linear additive judgment strategy. Instead, sometimes, judges rely on nonlinear and noncompensatory judgment strategies (Einhorn, 1970, 71, 72; Mellers, 1980; Ganzach, 1997). In particular, disjunctive and conjunctive strategies have been identified to predict judgments well. These strategies imply a noncompensatory cognitive process, making a decision primarily a function of poor attributes (conjunctive model) or good attributes (disjunctive model). The descriptive power of disjunctive or conjunctive judgment strategies has been widely acknowledged in the literature (e.g., Dhar, 1996; Payne, Bettman & Johnson, 1993; Svenson, 1979), making them alternative, promising models to describe sentencing. Although disjunctive and conjunctive strategies assume a nonlinear and noncompensatory judgment process, Einhorn, (1970) and Komorita (1964) have illustrated how the decision outcomes of both strategies can technically be modeled with a regression framework.

Heuristic Approach to Decision Making

Despite the success of regression type models in describing the outcome of a cognitive process (i.e., the final estimation), they have been criticized for not capturing the process itself, but providing only a paramorphic description (Brehmer, 1994; Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Hoffman, 1960; for a review, see Doherty & Brehmer, 1997). To attain deeper insight into human decision processes, Simon (1956) had argued for the development of process models that take the limitations of human cognitive capacities as well as the constraints of decision situations such as limited time and information into account. In this vein, Gigerenzer, Todd, and the ABC research group (1999) have argued that humans rely on heuristics, simple strategies that use a small amount of information that is processed in a simple fashion to make a quick decision. Following this argument, a large amount of evidence has been accumulated illustrating that people indeed apply heuristics, in particular in situations with high search costs, time pressure, or when heuristics perform better than alternative models (e.g., Bröder, 2000; Bröder & Schiffer, 2003; Dieckmann & Rieskamp, 2007; Mata, Schooler, & Rieskamp, 2007; Newell, Weston, & Shanks, 2003; Rieskamp, 2006, 2008; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006). Likewise, heuristics often do well in predicting people's decisions in many real-world tasks such as sports (Johnson & Raab, 2003), medicine (Green & Mehr, 1997), or product evaluations (Astebro & Elhedhli, 2006).

Using heuristics appears very reasonable, particularly when they take the structure of the environment into account: Here, they have proven to perform as well as, or even better than information-intense strategies (Todd & Gigerenzer, 2007). Accordingly, the reliance on heuristics in legal decision making should not come as a surprise. The legal decision environment is highly complex and the workload of legal personnel heavy; decisions need to be made under time pressure, often with little or no feedback regarding the quality of the decision (Konečni & Ebbesen, 1984; Gigerenzer, 2006). Even if specific rules exist to guide the decision process, they are often too complex to be executed in the allotted time (Ruback & Wroblewski, 2001). Following this hypothesis, Dhami and Ayton (2001) illustrated that British magistrates' bail decisions were best described by the matching heuristic, a simple decision tree that predicted more than 90% of the bail decisions based on three pieces of information (Dhami, 2003). Likewise, it has been discussed that heuristics could underlie sentencing decisions (Englich, Mussweiler, & Strack, 2006; for an overview see Colwell, 2005). To our knowledge, we are the first to propose a heuristic model to predict prosecutors' sentencing decisions, to compare this model rigorously to traditional regression models, and finally, to investigate the extent to which the model accurately represents legal policy.

Heuristics for Sentencing

Recently, von Helversen & Rieskamp (2008) introduced the mapping model that presents a heuristic approach of quantitative estimations (see also von Helversen & Rieskamp, 2009). The mapping model seems to be a promising model to predict sentencing: In a series of experiments, von Helversen and Rieskamp (2008) showed that it captured participants' judgments well, outperforming a linear regression and an exemplar model in a situation in which the cases' criterion values followed a skewed distribution. However, the original mapping model could not be applied to sentencing decisions, as it had only been defined for binary cues. Therefore, in the following we propose an extension of the mapping model making it applicable to situations with continuous cue values. This extension is built on Parduci's work (1974) on how people subjectively perceive and categorize continuous distributions. Furthermore, the present work tests the mapping model not only against the regression analysis, but also against various alternative models of human judgment (Einhorn, 1970).

The mapping model assumes that the psychological process underlying quantitative estimation consists, first, of a categorization process and, second, of an estimation based on the categorization. More specifically, when people make a judgment about a case or object, they assign the object to a category and use a typical criterion value for this category as an estimate. With binary cue values, categories are formed on the basis of the cues' sum, that is, the sum of an object's characteristics or features indicating a high criterion value. The typical criterion value of a category is represented by the median criterion value of all cases belonging to this category. For example, to estimate the selling price of a house, the mapping model assumes that one would consider the house's features that speak in favor of a high price (e.g., great location, a deck, a swimming pool), categorize the house according to its average value on theses features into a certain price class, and estimate a price that is typical for houses within this price class, that is, the median price of the houses in this category.

The mapping model makes the assumption that all factors considered are weighted equally. Although this assumption may initially sound counterintuitive, there is ample evidence for the plausibility of equal weighting. For one, many studies have shown that differential weighting of cues often does not lead to higher accuracy (Dawes, 1979; Hogarth & Karelaia, 2005; 2007). Furthermore, equal weighting is a common building block for heuristics and has been supported broadly by behavioral research (e.g., Bröder & Gaissmeier, 2007; Gigerenzer & Goldstein, 1996; Hertwig, Davis, & Sulloway, 2002). Because of its success in predicting judgments, we consider a unit weight linear model as a second heuristic model (Dawes, 1979; Gigerenzer & Goldstein, 1996; Hogarth & Karelaia, 2007). The unit weight linear model is an improper linear regression model that gives the same weight to every factor (Dawes & Corrigan, 1974). Although the mapping model and unit weight linear model share the assumption of equal weighting and thus overlap in their predictions, they differ in their assumptions about the estimation process with the mapping model allowing for nonlinear estimations.

In summary, we identified the standard linear regression model, the disjunctive and the conjunctive models, the unit weight model, and the mapping model as promising candidates for predicting human judgments. We test these models against each other to find which one best describes the sentencing process to gain a better understanding of the cognitive processes underlying sentencing and the extent to which they are consistent with legal requirements. In the following, we describe how the models can be adapted to predict sentencing judgments.

Multiple linear regression.

According to social judgment theory, sentencing should be considered as a process of weighting factors according to their importance and summing factor values to determine the sentence (Doherty & Brehmer, 1997; Cooksey, 1996). The weights that best characterize the sentencing process are found by a regression analysis, minimizing the squared deviation between the judgment and model prediction (cf. Cohen, Cohen, West, & Aiken, 2003; Cooksey, 1996):

$$Y = \sum_{i=1}^{I} a_i X_i + a_0 , \qquad (1)$$

where Y is the predicted judgment, I is the number of factors i, and the judgment is determined by the sum of the product of the factor values, X_i with their respective weights, a_i, for each factor plus an intercept, a_0. To illustrate the judgment process predicted by this model we consider a case of a defendant convicted for theft (see Figure 1). The defendant confessed to five cases of theft and has three prior convictions; the net worth of stolen goods is 30 dollars. If the prosecutor's judgment process for sentencing was described by a linear regression, he or she would first determine factors important for sentencing, that is, presence of a confession, number of charges, prior record and net worth of property. Then the prosecutor would weigh each factor's value according to the importance assigned to the factor, resulting in the number of days of fine the sentence would be increased or decreased in response to this factor. For example, in our case the sentence would by increased by12 days because the defendant has three prior convictions compared to a defendant with no prior convictions. The final sentence consists then of the sum of the weighted cue values resulting in this example in a sentence of 26 days of payment. Thus, psychologically, a linear additive judgment model assumes that each factor can be judged independent of the values of the other factors and the final sentence hinges on the weight that is given to each factor.

If prosecutors and judges follow legal policy in sentencing and thus employ a linear additive judgment process, sentencing should be captured well by a linear regression model. Furthermore, if the sentencing policy corresponds to the law, all legally relevant factors should make a significant contribution, whereas extralegal factors should not be considered.

Disjunctive & conjunctive models.

Both the disjunctive and the conjunctive models assume a nonlinear and noncompensatory judgment process. More specifically, the conjunctive model requires that a case pass a minimum score for each factor so that a high evaluation is reached. This implies that an evaluation is primarily based on the factor with the lowest score. For sentencing decisions, the model suggests a benign judge, sentencing the offender according to the factor that suggests the least severe offense. In the above-mentioned example, this could result in a judge focusing on the fact that the defendant confessed, and thus assigning an overall less severe sentence. According to Einhorn (1970; 1972), the model can be approximated by a parabolic response surface, with the general model for this surface:

$$Y = \prod_{i=1}^{I} X_i a_i \ , \tag{2}$$

where Y is the predicted judgment, I is the number of factors i, and a_i is the importance weight of each factor value X_i. This model can be easily estimated when taking the logarithm on both sides of Equation 2, so that the following regression equation results:

$$\ln Y = \sum_{i=1}^{I} a_i \ln X_i \ . \tag{3}$$

The disjunctive model assumes that a case is evaluated primarily according to the factor with the highest score. This model posits a tough judge sentencing an offender based on the factor suggesting the harshest sentence, e.g., focusing on the prior record in our example. According to Einhorn (1970) and Komorita (1964), the model can be approximated by a hyperbolic response surface, with the general model for this surface:

$$Y = \prod_{i=1}^{I} (1/b_i - X_i)^{a_i} \ , \tag{4}$$

where Y is the predicted judgment, I is the number of factors i, b_i is a constant arbitrarily set above the highest factor value X_i, and a_i is the importance weight for each factor. This model can easily be estimated by a multiple linear regression after a logarithmic transformation of Equation 4:

$$\ln Y = \sum_{i=1}^{I} -a_i \ln(b_i - X_i) \ . \tag{5}$$

Both models capture a configural judgment process that does not correspond to the legal policy that requires an independent judgment of each factor.

The mapping model.

How can the mapping model be applied to sentencing decisions? In the binary version, the mapping model categorizes objects according to their cue sums. However, to apply the mapping model to continuous cues, it is necessary to first normalize factors in order to allow comparisons of factors with different dispersions. Instead of using a purely statistical technique (i.e., z-transformation), we followed a psychological approach by applying range-frequency theory (Parducci, 1974) to normalize factors. Range-frequency theory holds that human judgments of magnitude and size are context-dependent, that is, they depend on the range of stimulus values as well as on the frequency with which a stimulus value appears (for details, see Appendix A). Thus, our normalization of factors relies on a psychological theory of how humans subjectively perceive magnitudes of factor values.

After normalizing all factors, we determined the mean factor value for all encountered cases. To simplify the statistical analysis, we inverted all factors that were negatively correlated with sentence magnitude, so that after inversion all factors were positively correlated with sentence magnitude. Please note that this is only a statistical simplification; alternatively, the difference between the mean score on aggravating factors and the mean score on mitigating factors could be taken.

Secondly, generalizing the mapping model to continuous cues requires that category number and category are determined, whereas in the binary case, the number of categories and their boundaries are defined by the number of cues and cue sums, respectively. We assumed a default value of seven categories with equally spaced category boundaries. First, humans' limited cognitive capacities make it reasonable to assume a limited number of categories (see also Miller, 1956). Second, research on judgment scales suggest that scales with 5 to 9 options usually capture participants' responses well (Alwin & Krosnick, 1991).[1] Accordingly, we formed seven equally spaced categories by determining the minimum and maximum value of cases' seriousness and dividing the range into equally sized categories. Category boundaries were chosen so that all categories featured the same distance between boundaries. Next, the typical sentence for each category was computed by taking the median sentence of all cases that fell into the same category.

Lastly, the sentence for a new case is determined by establishing its category membership and then using the typical sentence of that category as a sentence for the new case. One goal of this article was to test whether this generalization of the mapping model to continuous cues, with its simplifying assumptions, can provide a good account of judgment processes, in general, and of sentencing decisions, in particular.

In Figure 1 we give an example how a prosecutor would determine the sentence when following a judgment process that is best predicted by the mapping model. Similar to a linear additive judgment process, a prosecutor would first consider the factors of the case relevant for sentencing. Next, he or she would standardize the factors according to range-frequency theory, that is, rating the severity of each factor value on the same scale, for example, a scale ranging from 1 to 10, resulting in a low value for the amount of money stolen, but a high value of 7 for the three prior convictions. Then the prosecutor would form an overall impression of the case's severity by taking the mean of the standardized factor values, resulting in an average value of three. Based on this average score, the category into which this case falls would be determined by comparing the average score with the category boundaries, and the typical sentence for that category retrieved. In this case, the resulting sentence would be 20 days of fine. The mapping model differs from a linear additive model, because it does not assume that factors are differentially weighted. Further, for the mapping model the sentence magnitude depends crucially on the typical sentences that go with each category. These depend on the experience and knowledge the prosecutor has gained with similar cases.

Unit weight linear model.

As a second heuristic model, we implemented a unit weight linear model, which is an improper version of a linear regression model, with the simplifying assumption that all factors receive the same weight. Thus, Equation 1 is reduced to:

$$Y = a \cdot \sum_{i=1}^{I} X_i + b, \tag{6}$$

where Y is the judgment predicted, I is the number of factors i, a is a constant weight that is given to each factor value X_i, and b is an intercept. Thus, the unit weight linear model corresponds to a sentencing process that determines an average sentence and then adjusts this sentence up or down depending on the number of mitigating and aggravating factors. More specifically, the prosecutor in our example (see Figure 1) would determine the factors and factor values and standardize them to make them comparable. In the next step, he or she would weight them all by the same weight and then adjust an average sentence, which corresponds to the intercept in the regression equation by the weighted factor values; e.g., the presence of a confession would lead to a reduction of the 30-day average sentence by 6 days, but the prior record would increase the sentence by 4 days, and so on, resulting in a final sentence of 21 days of fine. The basic logic of the unit weight model and linear regression is the same, however, assuming for simplicity that all factors are equally important. This assumption of equally weighted factors is shared by the unit weight model and the mapping model. However, the two models differ because the mapping model determines the sentence with respect to the categories a case is classified into; thus, the mapping model allows nonlinear judgments, whereas the unit weight linear model does not.

Study: Analysis of Trial Records

To study the cognitive processes underlying sentencing in common cases and to determine how well these processes correspond to legal policy, we adopted an actuarial approach by conducting an analysis of trial records. We chose this approach because it is based on real cases and thus reinforces the external validity of our findings. Furthermore, the complexity of real cases and contextual factors, such as the time pressure of daily case loads, could be decisive in the cognitive process underlying sentencing decisions, favoring the analysis of real case data (Konečni & Ebbesen, 1979).

Method

We focused on three common property offenses, namely, theft, fraud, and forgery. This allowed us to include different offenses while measuring the severity of each offense on a common scale—money—and keeping the sentencing range equal (0–5 years for a common case and 3–6 months to 10 years for an aggravated case). To investigate the sentencing process we collected trial records from a small Brandenburg Court (the *Amtsgericht Bad Freienwalde*), for the years 2003 to 2005. All records with a main charge of theft, forgery, or fraud (§§ 242, 243, 244, 248, 263, and 267) were included in the analysis. Trial records included the indictment, the transcript of the trial, orders by the prosecution, and the verdict. Based on these documents we identified offense and offender characteristics relevant for sentencing, the sentencing range, and the recommendations of the prosecution and the defense.

Coding system. To extract the factors that might influence the magnitude of the sentence from the trial records, we developed a coding system to classify the offense and offender characteristics that could be relevant for sentencing. We decided which factors to include in the coding system based on the German penal code (§§ 46, 47, 52, 53, 242, 243, 244, 248, 263, and 267) in close cooperation with legal experts in the area of sentencing. Besides the factors stated in §46, German law allows sentence adjustments to achieve general as well as specific prevention objectives (Meier, 2001; Schäfer, 2001). Furthermore, the sentencing range can be lowered if mitigating reasons as specified in articles 21, 23, and 49, exist. As our sample did not included mitigated sentencing ranges according to these articles, we relied on the sentencing ranges as specified for common and aggravated cases of theft (§242 ff.), fraud (§263), and forgery (§267).

The coding of a factor rested upon the indictment, the trial transcripts, and the verdict. Besides the legal factors, the coding system also included extralegal factors that have been found to affect sentencing (e.g., Konečni & Ebbesen, 1982; ForsterLee et al., 2006). Table 1 provides an overview and a description of the factors.

The coding system included personal information on the offender, as well as legally relevant factors concerning the offender's criminal and personal history. To capture the severity of the crime, several characteristics of the offense were coded, such as the number of charges and the net worth of property violated. The presence of mitigating and aggravating factors concerning the conduct of the crime was coded in two summary factors capturing the amount of mitigating and aggravating evidence. If the description of a case in the indictment and the trial protocols left doubt about the presence of a mitigating or aggravating factor, the verdict was used as a reference. Only if the behavior in question was mentioned in the rationale of the verdict was it considered as mitigating or aggravating evidence. Additionally, the presence of a confession and mitigating behavior after the crime, such as remorse, were coded as two separate factors. A further mitigating summary factor coded whether the net worth of property violated was low enough to count as a less severe case (§ 248) and whether the offender had no prior record; these are two characteristics specifically identified by the German penal code as mitigating the sentence regardless of the overall impact of property violated or of any prior record. Additionally, we included three factors concerning legal regulations, such as, for instance, the sentence range applied. We did not include the recommendation of the defense in the analysis, because in most cases, the defendant did not have a defense attorney present during the trial. The role of the defense in the German legal system is less pronounced than in the American adversarial system. Because of the obligation of the prosecution to objectivity, the role of the defense at least in minor cases is often restricted to giving legal advice and help in dealing with the legal system.

For most variables, a nominal or ordinal level of measurement was assumed. Nominal variables were binary coded, indicating the presence or absence of a factor; ordinal variables were dichotomized by a median split.[2] For the variables number of charges, offenders, and prior convictions, amount of mitigating or aggravating evidence, and net worth of property, an interval scale was assumed. Two independent raters coded the cases. The raters' agreement was satisfactory on all subjectively rated factors (r = .77, SD = .12). Nonrandom missing data were analyzed and missing values substituted with the mean of the variable, because no effect on the dependent variable was found and the overall number of cases was rather small.

Dependent variables. Dependent variables were the number and magnitude of daily payments (for fines) and the length of a prison term in months (for incarceration) as recommended by the prosecution and the verdict. According to the German legal system, a fine is constructed as a number of daily payments of a certain magnitude. The number is determined in correspondence to the severity of the crime, whereas the magnitude depends on the income of the defendant. As the aim of this study was to compare sentencing for prison terms and fines, we focused on the number of daily payments as the dependent variable for fines corresponding to length of prison sentence. The number of daily payments can vary between 5 and 365; more severe offenses are sentenced by incarceration. The dependent variable for incarceration length was number of months sentenced to prison, irrespective of whether the offender was let off with probation.[3] We analyzed the sentences for fines and incarceration separately for several reasons. Even though according to the law one day of fine should correspond to one day in prison, research has indicated that incarceration sentences are perceived as more severe (e.g. Oswald, 1994), making it difficult to judge the two punishments on a single scale. Even more important, it has been argued that fines and incarceration sentences serve different sentencing goals, with fines giving more weight to rehabilitation goals and imprisonment emphasizing protection and prevention goals (Schäfer, 2001). This suggests that fining and incarceration decisions may be based on different factors and that underlying cognitive processes may differ. Therefore, we analyzed fines and incarceration separately to explore potential differences in cognitive process.

Description of the court, the offenses, and the offenders. The Amtsgericht Bad Freienwalde is a small court in the Brandenburg district of Märkisch-Oderland, close to the Polish border, and under the jurisdiction of the Frankfurt (Oder) district attorney's office. The city of Bad Freienwalde has a population of 13,000 with an unemployment rate of 12%. Overall, 99 cases of theft, fraud, and forgery were tried in this court in the years 2003 and 2004. Of the 99 cases, 15 were excluded because the major charge was not an offense under consideration in this study, juvenile law was applied, or the case did not lead to a conviction. Of the remaining 84 cases, 82% were tried by the same judge. The 84 cases were prosecuted by 45 different attorneys with a maximum of 5 cases handled by the same attorney. In 49 cases, the main charge was theft, in 20 it was fraud, and in 15, forgery. On average, property worth €2,497 was violated (SD = €8,826). The offenders were predominantly German males; 69 were men and 15 women. Eight offenders did not have German citizenship. The mean age of the offender was 36 years, ranging from 20 to 80 years. About half of the offenders were sentenced to a fine (M = 48 days; SD = 27) and half to a prison term (M = 8 months; SD = 6).

Model selection. To identify the cognitive process underlying sentencing, we compared how well several models of human judgment were suited to predict sentence magnitude. More specifically we tested a linear regression model (e.g., Cooksey, 1996), a disjunctive model and a conjunctive model (Einhorn, 1970; Einhorn, 1972), a unit weight model (Dawes, 1979), and the mapping model (von Helversen & Rieskamp, 2008).

Testing these models on the data of real cases raised two crucial methodological problems: First, real cases involve an enormous number of factors that could potentially predict the sentence. In our cases, we recorded 22 factors that could influence the sentencing decision. A common technique to find out which factors have an effect when using regression analysis relies on standard null hypothesis significance testing (NHST). Often procedures are performed where factors are step-wise either included in or excluded from the regression equation (cf., Cohen et al., 2003). However, when considering a larger number of factors, this procedure is very unsatisfying, because factors that were added to the equation at the beginning of a step-wise forward procedure might not have been added had other factors already been included. Therefore, different statistical procedures applied to the same original set of factors often lead to inconsistent results (i.e., different regression equations), which can lead to very different conclusions. Furthermore, the mapping and the unit linear weight model do not include a mechanism to select which factors are relevant but assume that this knowledge is available. Thus, we needed a methodology to test the impact of the different factors for these models.

The second methodological problem we faced concerns the models' complexity, that is, their flexibility in describing different results. In particular, we were interested in testing models with different numbers of free parameters and therefore differences in their potential to describe different processes. Therefore, we required a methodology that takes the models' complexity into account when testing them against each other.

To tackle these two methodological problems, we followed a Bayesian approach, specifically the Bayesian model averaging (BMA) method (see Raftery, 1995, Raftery, Madigan, & Hoeting, 1997; Wagenmakers, 2007). In general, the Bayesian approach allows estimating the probability of a model given a set of data and prior beliefs about the probability of the model. By providing an estimate of the probability of a model given the data, Bayesian methods may be preferable to null hypothesis testing, which only considers the probability of the data given a null hypothesis. In contrast, the Bayesian method identifies the most probable model(s) given the data. Furthermore, BMA provides reliable estimates of the predictors' influence on the dependent variable and allows comparison of models with different complexities by taking the models' free parameters into account. BMA was proposed especially for model selection and to examine the uncertainty of parameter estimates. To identify the most probable models, the Bayesian method calculates the posterior probability of a model given the observed data. Pragmatically this is performed by determining the Bayesian information criterion (BIC), which approximates the so-called Bayes factor (Raftery, 1995; Schwarz, 1978). The method additionally allows one to specify the probability that a factor will have an impact on the dependent variable: Taking model uncertainty fully into account, the average amount of evidence speaking for an effect of a factor is determined by summing the posterior probabilities of all models that include this factor (for details, see Appendix B).

Factor selection. The most reliable method for model selection, according to Raftery (1995), is to construct all possible models that can be built with the available factors and then select the models with the highest posterior probability given the data. However, including all candidate predictor variables would result in an enormous number of possible models, as 22 predictor variables amount to 4,194,304 models. Thus, to ensure that we considered all relevant factors and to check for quadratic effects and interactions—but at the same time keep the number of models in the comparison to a manageable number—we took a two-fold approach. First, we selected the factors where an effect on sentence length was most probable. Then we additionally tested all further factors, quadratic effects, and two-way interactions by adding each of them to the factors selected previously, but only including them if they led to an improved model measured by a BIC improvement of at least a value of 5. The first set of factors consisted of all factors that substantially correlated with the dependent variable (i.e., showed a value of $r > .3$) and additionally the factors confession and remorse that were not necessarily correlated with sentence magnitude but are of special theoretical importance because they frequently appear as mitigating reasons in the rationale of the verdict. For the magnitude of fines, the recommended number of daily payments correlated substantially negatively with the presence of a confession and mitigating evidence II, and substantially positively with the net worth of the property violated, number of prior convictions, number of charges, the amount of aggravating evidence, summary penalty, and penalty order. For incarceration length, the factors net worth of property violated, summary penalty, aggravating evidence, number of charges, and number of offenders correlated significantly positively with the sentence length, while the second mitigating factor (coding a low worth of property violated and no prior record) correlated negatively with recommended sentence length (see Table 3). The factor penalty order was not applicable, as a sentence by penalty order is not allowed for prison sentences. Somewhat unexpectedly, the presence of a confession and special circumstances leading to diminished capacity correlated positively with incarceration length. This effect, however, is probably due to the comparatively serious nature of these cases and does not reflect a negative evaluation of these factors for sentencing.

Based on these factors and remorse, we calculated the BIC values for all models resulting from all possible combinations of the factors. Then, step by step, we added all further factors, quadratic terms for all factors, and all two-way interactions to examine if they improved the predictive accuracy of the models. For the fines, this was not the case, so that we kept the 11 factors stated above for the main analysis (see Table 2). For the incarceration data set, three more factors—probation status, type of last sentence (dummy coded), and remorse—increased accuracy, leading to a total of 12 factors that we considered (see Table 3). The intercorrelation matrices are reported in Appendix C. This amounted to 2,048 models in the case of fining decisions and 4,096 models in the case of incarceration decisions for each model class.

Model comparison. For all of the models, we calculated the BIC′ values based on the models' amount of variance explained (R^2) as a measure of goodness-of-fit and the models' number of free parameters (see Raftery, 1995). Details on the computation and the equations can be found in Appendix B. The BIC′ value gives the odds with which a specific model is preferred to a baseline model. In the case of regression, usually a null model is chosen as a baseline model. The null model only includes an intercept (i.e., estimates the mean criterion value for all objects) and no further predictor (i.e., free parameter). It explains none of the variance in the data and its BIC′ is zero (see Equation 2); the BIC_k' of a specific model M_k is defined so that if the BIC_k' value is positive, the null model is preferred, while a negative BIC_k' value provides evidence for the model M_k under consideration. The lower the BIC_k' value, the more the model is supported by the data:

$$BIC_k' = n\ln\left(1 - R_k^2\right) + q_k \ln n \ , \qquad (7)$$

where R_k^2 is the value of R^2 for model M_k, q_k is the model's free number of parameters, and n is the number of data points.

Model fitting. For linear regression, the disjunctive model, and the conjunctive model, a least squares regression was run with the factors as predictor variables and the sentence recommendation of the prosecution as the dependent variable. Because the distribution of the factor values and the criteria was not normal for all factors, we additionally tested whether a logarithmic transformation of the cues or the criterion would lead to a better performance for the linear regression. Please note that a model with a logarithmic transformation of the cues and the criterion is equivalent to the conjunctive model (Einhorn, 1970). Additionally we tested whether the models would also benefit from a range-frequency transformation of the cue values in comparison to a logarithmic transformation. Because this was not the case, we only report the results based on the logarithmic transformation.

For the mapping models, the category borders and the typical sentence for each category were estimated from the data. First, the perceived factor score was calculated based on range-frequency theory, which comprises a free parameter *w* that captures the relative importance of range and frequency information (for details, see Appendix A). We assumed that *w* is constant for all factors and determined the best-fitting parameter value by a grid search. Then, case seriousness was computed by averaging the factor scores over all factors, the minimum and maximum case seriousness was determined, and the range was divided into seven equally sized categories. For each category, the typical sentence was calculated by taking the median of all cases that fell into this category. The typical sentence was estimated for all cases falling into one category and the amount of variance in the sentence recommendation of the prosecution explained (R^2) was computed.

For the unit weight linear model, we normalized the factors using a z-transformation and then estimated the parameter of Equation 6 using a least square fit criterion. Again we tested whether a normalization based on range-frequency theory led to a better predictive accuracy in comparison to a z-transformation. This was not the case, so we do not report the results.

For the purpose of model selection we first ran separate BMA analyses for all five models to identify the model that reached the lowest BIC value for each model class. In the next step, we used the BMA method to compare the best two or three models against each other in a joint analysis, and to determine the posterior probability of the models and stable estimates for the impact of each factor. Additionally, we computed an approximation of a Bayesian point estimator of beta weights and standard errors for each factor (see Appendix B).

Lastly, we included a test of an ordinal regression model. Multiple linear regression procedures could be at a disadvantage because they assume that cue values as well as criterion values are metric. If this assumption is violated, this could boost the performance of models with less stringent requirements such as the mapping model. To determine whether assumptions about the measurement scale limited the accuracy of the linear regression model, we ran an adjacent category logit model based on factors included in the model test (11 for fines and 12 for incarceration) and compared it to the linear regression model. However, because the adjacent category logit model for fines as well as for incarceration resulted in a smaller R^2 than the linear regression [fines: R^2 (*Nagelkerke*) = .63 and incarceration: R^2 (*Nagelkerke*) = .78] we did not consider it further.

Results

Overall, the class of mapping models offered the most probable description of the data for fines and incarceration sentences. For fines, the second-best model class was conjunctive models, and for incarceration, the unit weight model class was followed closely by the linear regression model class. Furthermore, our analysis showed that fining and incarceration decisions were only influenced by a small number of factors. In the following, we report the results of the analysis for fines and incarceration separately.

Magnitude of fines. To identify the model that best describes the fining decisions, we tested the five models against each other. We followed the BMA approach on the basis of the 11 factors identified during factor selection (see Table 2). This resulted in 2,048 possible factor combinations for each of the five model classes that we tested. In a first step, we identified the best of the 2,048 possible models for each of the five model classes. The two best models belonged to the class of mapping models reaching a BIC′ value of -56 and -55 and explaining 74% of variance in the sentence recommendations. The class of conjunctive models was second-best, with the two best conjunctive models reaching a BIC′ value of -38 and -37 and explaining 70% and 72% of the variance in the sentencing recommendations, respectively (see Appendix D for the accuracy of the other models).

In summary, the mapping model class provided the best description of the fining decisions. For a further model comparison test, we compared the mapping model class with the second-best model class of conjunctive models. In a joint analysis, we determined the posterior probability of all mapping and conjunctive models and determined stable estimates of the probability with which each factor was used, resulting in a comparison of 4,096 models with prior probabilities of .0002. Again the class of mapping models performed best (see Table 2 for the five best models). Figure 2 shows that the mapping model class reached an impressive overall posterior probability of .99999 compared to .00001 for the conjunctive model class.

Critical factors. To measure the importance of factors calculated based on the joint BMA analysis, we determined the probability of each factor affecting sentencing decisions. The most important factors were net worth of property, whether the sentence was recommended by a penalty order or after trial, and the presence of aggravating evidence. Additionally, the number of prior convictions, the number of charges, and whether the sentence was a summary penalty affected sentence magnitude; all of these variables were positively correlated with sentence magnitude. However, age, nationality, and a confession or remorse did not influence the sentence recommendation. These findings provide clear evidence that the judgment process of fining decisions is inconsistent with the legal requirement of accounting for all legally relevant factors.

Incarceration length. To examine whether the mapping model class was also best in comparison to the four alternative model classes in describing incarceration decisions, we again applied BMA analysis. For incarceration decisions, BMA analysis was based on 12 factors, corresponding to 4,096 different models and a prior probability of .0002 (see Table 3). According to the BIC′ criterion, the two best models again belonged to the mapping model class; the best two mapping models achieved BIC′ values of -74 and -73. The second-best model class was the unit linear weight model class, with the best model achieving a BIC′ value of -60, followed closely by the two best linear regression models with BIC′ values of -59 and -58, respectively (see Appendix D for the performance of the other models). Please note that the BIC difference of 14 between the best mapping model and the best alternative model provides substantial empirical evidence for the mapping model (with a BIC difference of more than 6 providing strong evidence for the better model, Raftery, 1995). However, the two best linear regression models explained a similar amount of variance in sentence length, as the two best mapping models (mapping model: 86% and 85% vs. linear regression: 88% and 87 %), and more than the best unit weight linear model (82%).

To determine which of the three model classes was the most likely to underlie decisions of the incarceration sentences, we compared the mapping model class to the unit linear weight model and the linear regression model class in a joint analysis. This analysis compared 12,288 models and provided overall supported for the mapping model as the superior type of model. The five best models belonged to the class of mapping models (see Table 3), with the posterior probabilities of these models adding up to an impressive joint probability of .80, compared to a posterior probability of only .20 for the remaining 12,283 models. Similarly, the joint posterior probability of models belonging to the class of mapping models was much higher than for the class of regression models or unit linear weight models, adding up to .999 for the mapping model class, compared to .001 for the regression model class combined with the unit linear weight model class. Figure 2 depicts the posterior probabilities of the best 1,000 models. Again, similar to fine sentencing, the empirical evidence provides strong support for the mapping model class in best describing incarceration sentencing.

Critical factors. Overall, there were more factors influencing sentencing in incarceration than in fines. The final model comparison showed strong evidence for the factors' number of charges, diminished capacity, mitigating evidence (II), probation status, and prison sentence as last sentence, and some support for summary penalty, aggravating evidence, and remorse. Of these factors, number of charges, probation status, prison sentence as last sentence, summary penalty, aggravating evidence and, surprisingly, diminished capacity correlated positively with sentence magnitude, while mitigating evidence and remorse correlated negatively with sentence magnitude. Again, confession did not affect sentencing.

Questionnaire Study

Our results suggested that only a small number of factors were relevant for both fines and incarceration sentencing. To test the extent to which our modeling results can be confirmed by prosecutors' independent judgments, we conducted a survey questionnaire. We asked prosecutors of two public prosecution offices in Berlin to rate the importance of factors included in our coding system on a scale from 1 (not at all important) to 7 (very important). Berlin has two major public prosecution offices, the Amtsanwaltschaft and Staatsanwaltschaft. While the Staatsanwaltschaft prosecutes all offenses and at all court levels, the Amtsanwaltschaft is restricted to prosecuting minor offenses at the municipal court level. Because our cases could be prosecuted by both prosecution offices, we recruited prosecutors from both.

Overall 54 prosecutors (48% males) participated in the study with a mean age of 44 years (*SD* = 5.77) and an average of 14 years of professional experience (*SD* = 5.27). The ratings of the prosecutors were very much in line with those expected by the penal code. Most factors relevant for sentencing according to the penal code received high ratings, while factors such as gender and nationality were rated as unimportant (see Figure 3). Some factors such as received mixed support, reflecting that they may play a role in some situations. Particularly, the factors summary penalty and penalty order received very mixed ratings. This is probably due to differences in the legal standing between the factors that caused some confusion for the prosecutors. While factors such as prior convictions or mitigating evidence are evaluated by the prosecution and the judge, whether or not a summary penalty or penalty order applies is not under the discretion of the sentencing personnel and thus some prosecutors were not sure how to rate their importance. We included them in the questionnaire because although their effect should be standardized, they have been found to influence the sentencing process.

Overall, prosecutors were fairly consistent in their ratings and no interoffice differences were found, except for mitigating circumstances, which were rated as significantly more important by the Staatsanwaltschaft ($M = 5.13$, $SD = .98$) than by the Amtsanwaltschaft, $M = 4.38$, $SD = 1.04$, $t(52) = 2.74$, $p < .01$).

We compared the prosecutors' importance ratings with the probability of each factor affecting sentencing as determined by our modeling analysis to examine whether the prosecutors would place importance on factors that are relevant for predicting sentencing decisions. As illustrated in Figure 3, for both fines and incarceration sentences, the modeling and ratings matched for the most and least important factors. This was true in particular for incarceration, where we found a significant rank correlation between the modeling of the importance and the ratings of the importance; $r(10) = .77$, $p = .01$. However, for incarceration and fining decisions, more factors received high ratings than were attributed importance by the modeling analysis, such as confession or prior conviction in the incarceration case.

Discussion

There are two ways in which sentencing decisions can deviate from the law: First, the decision can be based on a different set of factors than required by the law; second, the way these factors lead to a sentence can be inconsistent with the prescribed legal policy. The present article examined both routes by taking a cognitive approach. Testing five models of decision making against each other using a Bayesian approach, we identified the crucial factors influencing sentencing and the processes underlying sentencing and compared how well they matched with legal policy. Our results show that, although the factors influencing sentencing were broadly consistent with the penal code, the prosecutors considered only a limited number of factors and neglected factors that are legally relevant and rated as highly important. Furthermore, the sentencing process was apparently not consistent with the judgment policy assumed by the legal literature. Instead, the mapping model (von Helversen & Rieskamp, 2008; 2009) described the prosecutors' decisions best, suggesting that heuristics are an important tool to improve our understanding of sentencing decisions. In the following, we will first discuss which factors predicted sentence recommendations and highlight differences between fines and incarceration sentences. Then we will turn to the model comparison, the role of heuristics for legal decision making and its implications for sentencing. Finally we will discuss the significance of the BMA method and limitations of the current study.

Predictors of Sentencing Decisions

Overall, our analysis revealed a substantial match between the factors prosecutors considered important, the penal code, and the factors they actually used. Consistent with the law and self-reports of prosecutors, we found that extralegal factors such as sex, age, or nationality did not influence sentencing in fining and incarceration decisions. Furthermore, comparing the rated importance of the predictors with the importance of the factors according to the BMA analysis showed that self-reports matched with modeling results for the most important and least important factors. However, the comparison also showed that some legally relevant factors, such as the presence of a confession or prior record, were neglected even though they received high ratings of importance and were routinely mentioned in the rationale of the verdict. This suggests that prosecutors had some insight into which factors affected their sentencing. However, it also shows that they apparently overestimated the number of factors they considered in the directions of the penal code requirements. Similar results have been reported by Konečni and Ebbesen (1984), who found that judges reported using more factors for bail setting or sentencing than they actually did. This finding also resonates with a broad body of literature on judgment and decision making, which has shown that judges often have limited insight into their judgment policies (e.g., Brehmer & Brehmer, 1988) and rely on few cues for their judgments (Brehmer, 1994). This apparent mismatch between the self-reported use of case information and the information actually used could be fostered by cognitive constraints and time limitations under which sentencing decision are made, as has been stated by Konečni and Ebbesen (1984, pp. 15-16):

> "Faced with the need to make these decisions, but lacking the necessary cognitive capacities, it seems reasonable that decision makers would be aware of the complexity of the task (…), but be capable of using only the simplest strategies."

On the other hand, our results for incarceration sentences suggest that prosecutors are capable of achieving a good match between the factors they consider important and the ones they use for their sentence recommendations. The discrepancy between the rated importance of factors and the modeling was stronger for offenses punished by a fine than offenses punished by incarceration. This better match for incarceration than for fines might be explained by the fact that cases punished by incarceration are usually more serious than cases punished by fines. More serious cases are less frequent, arouse more public interest, have a higher probability of appeal, and thus are likely to be allotted more time and processed more systematically (Smith & DeCoster, 2000). This suggests that if prosecutors were given more time to make decisions, all relevant factors would be taken into consideration in their decision making (Payne et al., 1993).

Our results regarding the influence of confession on sentence magnitude warrant a separate discussion. Confession is routinely mentioned as a mitigating factor in judgment rationale and receives high ratings of importance by prosecutors. While confession correlated negatively with sentence magnitude in fines, it was positively correlated with a higher incarceration sentence and aggravating circumstances. However, the effect of confession on incarceration length disappeared after controlling for aggravating circumstances. This finding can be explained in two ways. First, confession—particularly in more serious cases punished by incarceration—could lead to more incriminating details becoming known and thus cause a higher number of aggravating circumstances and higher sentences. Second, defendants accused of more serious crimes may confess more often, hoping for a milder sentence. In any case, our analyses suggest that prosecutors were not fully able to discount confession context but disregarded confession as a mitigating factor. Interestingly, this was also true for fines, where confession initially predicted lower sentences. The modeling analysis also revealed that confession did not seem to play a role in determining a fine's magnitude. This suggests that confession may be less important for sentencing than one might initially think. However, this finding could be exaggerated by the nature of our data. It could be that judges and prosecutors make more fine-grained decisions about what constitutes a confession than mentioned in the trial records. For instance, they could differentiate between "real" confessions incited by remorse and "strategic" confessions aiming for a lower sentence after guilt has been proven.

In summary, our study indicated that the role of confession in sentencing is complex and hinges on variables such as the type of sentence as well as interdependent third variables such as aggravating information. Furthermore, the extent to which a confession influences a sentencing decision may depend on the perception of prosecutors and judge. Further research is necessary to clarify how confession influences sentencing.

Lastly, incarceration and fining decisions differed according to which factors influenced sentence magnitude. Fining decisions were more influenced by aggravating evidence and number of prior convictions, while incarceration length was more affected by mitigating evidence (II) and remorse. One reason could be that fines are often considered a less severe punishment than incarceration, even though they are equivalent according to German law (e.g., Oswald, 1994). A reanalysis of the data showed that, in fact, getting an incarceration sentence instead of a fine correlated significantly positively with prior record, probation status, aggravating information, and prison sentence as last sentence, but negatively with mitigating information. This suggests that the decision regarding type of sentence (fine or incarceration) could serve as a gateway that then influences which variables are employed for determining sentence length. Thus, in cases punished by fines, the prosecution might have already “used up” the influence of any mitigating information by sparing the offender an incarceration sentence, while giving more weight to aggravating information when the milder sentence was chosen. In principle this could be due to a reduced variance of these variables in the sub-data sets. However, this explanation does not seem to hold for our data sets, because variance in the complete set and subsets was the same for prior record as well as for mitigating information and remorse. This suggests that discounting factors in determining sentence magnitude could have been caused by making an earlier related decision (e.g. Russo, Medvec & Meloy, 1996; Svenson, 1979; 1996).

Heuristics in Legal Decision Making

Although we found that factors influencing sentencing were broadly consistent with legal policy, the process underlying sentencing decisions clearly deviated from the judgment process proposed in the legal literature (Schüneman, 1988). For both types of sentencing decisions, our analyses illustrated that a cognitive heuristic, the mapping model, provided a much better explanation for the sentencing process than a linear regression model, a disjunctive, a conjunctive, or a unit weight linear model. This provides further evidence that legal decision makers rely heavily on simple decision heuristics (Dhami & Ayton, 2001; Konečni & Ebbesen, 1984), and suggests that eliciting these employed heuristics is an important step in understanding and improving legal decision making.

The heuristics used by legal decision makers, however, seem to differ according to task. For instance, the mapping model differs from the matching heuristic (Dhami & Ayton, 2001; Dhami, 2003) in terms of how the decision process is conceptualized. The matching heuristic is a noncompensatory decision strategy describing how magistrates decide whether an accused person is released on bail or remains in custody. According to Dhami, magistrates made a punitive decision if the prosecution, the police, or other benches had kept the defendant in custody or denied bail, but granted it otherwise. Thus, a punitive decision by the prosecution could not be compensated for by a nonpunitive decision by the police and another bench. In contrast, the mapping model is a compensatory strategy, assuming that high values on one factor can compensate for low values on another. One important task feature that could trigger the use of either a compensatory or noncompensatory decision strategy could be the position of the decision maker in the sequence of decisions (i.e. decision hierarchy): If the decision maker is the first to make a decision or is building on decisions already made by others, different strategies might be applied. In Dhami's study, magistrates were the last instance to make a decision and could thus rely on previous decisions by other legal personnel, such as recommendations of the prosecution, police, or another bench. In the present study, however, prosecutors could not follow someone else's evaluation in determining the sentence recommendations. Looking at the judges' sentencing decisions, we found that they were almost perfectly predicted by the prosecutors' recommendations (Fines: r (42) = .99; Incarceration: r (38) = .96). Similarly, Konečni and Ebbesen (1982) reported that the probation officers' recommendation was the most important factor influencing judges sentencing decisions. This suggests that the position of the decision maker in the decision hierarchy has an important influence on the strategies people apply: Decision makers who cannot rely on previous decisions of others might select compensatory strategies that trade off different factors more often in comparison to decision makers who can also rely on the decisions of others.

Implications for Legal Decision Making

When examining the general question of whether decision makers deviate from legal policy (e.g., Dhami & Ayton, 2001; Ebbesen & Konečni, 1975; 1981; Hertwig, 2006; Van Duyne, 1987), our research leads to an important conclusion: It is not sufficient to consider which factors play a role under a single judgment model. Instead, it is important to take into account that the judgment process might not follow the judgment method recommended by legal policy. We showed that for both incarceration decisions and fines, the decision process was best described by the mapping model and that it deviated from legal policy. This suggests that the current legal regulations in Germany are not successful in leading to the desired integration of the information in the judgment. Furthermore it indicates that although it might be helpful to provide more time per case so that more relevant factors are taken into account when determining a judgment, this might not change the cognitive process underlying the judgment. To change the decision process, it appears necessary to additionally formulate "procedural" rules (in addition to "factual" rules) that specify how factors should be used in arriving at a decision. For instance, in the German legal system, fines are separated into number and amount of daily payments. While the amount reflects the financial situation of the person convicted, the number of payments reflects the seriousness of the crime. This separation was introduced to ensure that punishments for similarly serious crimes are equally hard on rich and poor offenders (Hillsman, 1990). However, it also ensures that both factors are considered in sentencing and prevents their being ignored or undervalued by the person determining the sentence. Thus, procedural rules explicitly formulating intermediate steps in the sentencing decision could bring the sentencing process into line with the broadly agreed policy, without requiring complex, rigid sentencing guidelines (Ruback & Wroblewski, 2001).

Finally, the results also resonate with an ongoing discussion in the legal literature: Whether a linear additive policy is the best and fairest, or whether there exist alternative processes that would ensure fair punishment while being more in line with human cognitive processing. In England, for example, similarity-based decision aids for sentencing are under discussion (e.g., Tata, 1998), and in Germany the relevance of “normal” and “average” cases as reference points for sentencing has been hotly discussed (see Bruns, 1988; Mösl, 1981, 1983; Theune, 1985a, 1985b). This suggests that integration processes relying on measures of typicality, such as the mapping model, could also be considered as fair and just as long as the important features of the case are taken into account. Here, psychological research could inform legal institutions to achieve a better match between decision processes and legal requirements.

Bayesian Approach

When examining how decision makers deviate from the legal policy, we have argued that it is important to infer which model describes the decision process best and to infer the importance of factors on the basis of the best empirical model. In the present study, we have argued against the standard methodological approach in the policy-capturing literature (e.g., Cooksey, 1996). According to the standard approach, a single regression model is estimated by applying a specific statistical test procedure. This approach has the disadvantage that it can lead to somewhat different results and conclusions depending on the statistical procedure chosen. Moreover, the interpretation of the influence of single factors is rather complicated, because the influence can depend on the other factors included in the model.

In contrast, we have argued and applied a Bayesian approach to compare different models and to infer the importance of various factors in the decision process. Bayesian theory describes how beliefs about hypotheses such as which model best describes a decision process are updated in the light of new evidence according to the laws of probability theory. Thereby a Bayesian approach allows direct estimation of the probability of a model or a hypothesis given a set of data and prior beliefs about the probability of the model or hypothesis. By providing an estimate of the probability of a hypothesis given the data, Bayesian methods may be preferable to standard null hypothesis significance testing (NHST) as a method of scientific inference in general, because NHST can only give insight into the probability of the data if the null hypothesis is true (for a discussion of NHST and Bayesian alternatives see Wagenmakers, 2007).

In our case, the Bayesian approach led us to estimate the posterior probability of each model and thus to directly answer the question of which model class was more likely to account for the data. Furthermore, by considering all models of each class and not just one best model, we could take model uncertainty into account and achieved more reliable conclusions about which model predicted the sentencing decision best and which factors are important for sentencing decisions.

Limitations of the Study

Our study focused on a single German court. This naturally raises the question of how well the results generalize. Many studies have shown the importance of location and the legal culture of a jurisdictional district (e.g., Johnson, 2006; Kautt, 2002; Langer 1994). Especially the factors influencing sentence magnitude could differ between districts, and thus our results concerning the importance of factors should be treated with caution. Furthermore, our results were based on a fairly small sample, which could reduce the generalizability of the results even within the jurisdictional district. Nevertheless, for the restricted data set, we have illustrated the benefits of a cognitively inspired approach to legal decision making. Future research is necessary to test whether these results can be replicated with larger samples for a wider range of jurisdictional districts.

Generalizing further, similar results might even be anticipated outside of Germany, given that the general features of the task remain the same. That is, as long as the prosecutor or the judge has to integrate several factors to determine a final sentence, the mapping model could offer a valid description of the process. However, it is important to consider that this generalization needs to be limited to legal systems with similar structures. In contrast, in legal systems where sentencing is strictly regulated by sentencing guidelines—as, for instance, in the United States—different processes might apply.

Besides how far our results can be generalized to sentencing in other systems, a second question concerns in how far plea-bargaining offers by the prosecution could rely on similar cognitive processes as those outlined in this manuscript. Naturally, plea-bargaining is a negotiation process and thus often depends on many other variables, such as probability of a conviction, related cases of the prosecution, strength of the defense, among others. However, the special type of sentence bargaining where the prosecution makes an offer after the defendant confessed, in particular if it is a take-it-or-leave-it offer may follow similar processes as discussed in the present article. However, further studies investigating the generalizability of the mapping model to explain sentencing are necessary. In a similar vein, it is important to note that this study focused on low-level offenses. It remains open whether the same cognitive processes underlie the sentencing of more severe cases, such as capital crimes.

Lastly, it should be noted that we relied in this study on an analysis of trial data and did not conduct an experimental study. Thus our data provides insight into which factors play a role in the sentencing process of typical cases; however, our data does not allow conclusions about causality, which would require testing in an experimental setting.

Conclusion and Outlook

This article provides evidence that a cognitive approach to decision making can lead to a better understanding of the cognitive processes underlying sentencing decisions. Our results suggest that the sentence recommendations of prosecutors were not consistent with the legal requirements; instead, sentence recommendations were well described by the mapping model, a heuristic model for quantitative estimation (von Helversen & Rieskamp, 2008). This study joins a growing body of research questioning the ability of decision makers to comply with legal regulations and emphasizes the importance of understanding cognitive processes for the development of institutions.

References

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Alwin, D. F., & Krosnick , J. A. (1991). The reliability of survey attitude measurement: The influence of question and responedent attributes. *Sociological Methods & Research, 20*(1), 139-181.

Astebro, T., & Elhedhli, S. (2006). The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Science, 52(3),* 395-409.

Brehmer, A., & Brehmer, B. (1988). What have we learnt about human judgment from thirty years of policy capturing? In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75-114). Elsevier/North Holland: Amsterdam.

Brehmer, B. (1994). The psychology of linear judgment models. *Acta Psychologica, 87,* 137–154.

Bröder, A. (2000). "Assessing the empirical validity of the "Take-the-best" heuristic as a model of human probabilistic inference." *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26,* 1332-1346.

Bröder, A., & Gaissmeier, W. (2007). Sequential processing of cues in memory-based multi-attribute decisions. *Psychonomic Bulletin & Review, 14,* 895-900.

Bröder, A., & Schiffer, S. (2003). Take the best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General, 132,* 277–293.

Bruns, H-J. (1985). *Das Recht der Strafzumessung, Eine systematische Darstellung für die Praxis* (2. Aufl.) [The law on sentencing: A systematic account for legal practitioners (2nd ed.]. Köln: Heymanns.

Bruns, H-J. (1988). Die Bedeutung des Durchschnitts-, Regel- und des Normalfalles im Strafzumessungsrecht. Mögliche Orientierungspunkte für die Eingliederung der Tat und des Strafmaßes in die Stufenfolge des Rahmens? [The relevance of the average, the norm, and the normal case in criminal sentencing. Possible points of reference for the placement of the offense and the sentence in the levels of the sentencing range] *Juritische Zeitung, 43,* 1053-1058.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Colwell, L. H. (2005). Cognitive heuristics in the context of legal decision making. *American Journal of Forensic Psychology, 23,* 17-41.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods and applications.* San Diego: Academic Press.

Cooksey, R. W., P. Freebody, et al. (1986). "Teachers predictions of childrens early reading achievement: An application of social judgment theory." *American Educational Research Journal, 23,* 41-65.

Davis, T. L., Severy, L. J., Kraus, S. J., & Whitaker, J. M. (1993). Predictors of sentencing decisions: The beliefs, personality variables, and demographic factors of juvenile justice. *Journal of Applied Social Psychology, 23,* 451-476.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571-582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95-106.

Dhami, M., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making, 14,* 141–168.

Dhami, M. K. (2003). Psychological models of professional decision making. *Psychological Science, 14,* 175-180.

Dhar, R. (1996). The effect of decision strategy on deciding to defer choice. *Journal of Behavioral Decision Making, 9(4),* 265-281.

Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy or probabilistic inferences. *Memory & Cognition, 35,* 1801-1813.

Doherty, M., & Brehmer, B. (1997). The paramorphic representation of clinical judgment: A thirty-year retrospective. In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on judgment and decision making: Currents, connections and controversies* (pp 537–551). Cambridge: Cambridge University Press.

Doherty, M. E., & Kurz, E. (1996). Social judgement theory. *Thinking and Reasoning*, 2*,* 109–140.

Ebbesen, E. B., & Konečni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology, 32,* 805-821.

Ebbesen, E. B., & Konečni, V. J. (1981). The process of sentencing adult felons. In B. D. Sales (Ed.), *The trial process* (pp. 413-458). New York: Plenum.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin, 73,* 221-230.

Einhorn, H. J. (1971). “Use of nonlinear, noncompensatory models as a function of task amount of information.” *Organizational behavior and human performance*, 6, 1-27.

Einhorn, H. J. (1972). “Expert measurement and mechanical combination.” Organizational Behavior and Human Performance*, 7,* 86-106.

Einhorn, J. H., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Regression models and process tracing analysis. *Psychological Review, 86,* 465–485.

Engen, R. L., & Gainey, R. R. (2000). Modeling the effects of legally relevant and extralegal factors under sentencing guidelines: The rules have changed. *Criminology, 38,* 1207-1229.

Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology, 31,* 1535-1551.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin, 32,* 188-199.

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

Foth, E. (1985). Strafschärfung/Strafmilderung—eine noch unerledigte Frage der Strafzumessung [Aggravating and mitigating sentences—outstanding issues for sentencing]. *Juristische Rundschau, 10,* 397-399.

Ganzach, Y. (1997). Theory and configurality in clinical judgments of expert and novice psychologists. *Journal of Applied Psychology, 80,* 954-960.

Gigerenzer, G. (2006). Heuristics. In G. Gigerenzer & C. Engel (Eds.), *Heuristics and the law* (pp. 17-41). Cambridge, MA: MIT Press.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650-669.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 3–34). New York: Oxford University Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Green, L., & Mehr, D. R. (1997). What alters physicians' decisions to admit the coronary care unit? *The Journal of Family Practice, 45(3),* 219-226.

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2001). Inside the judicial mind: Heuristics and biases. *Cornell Law Review, 86,* 777–830

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62, 255-262.

Harries, P. A., & Harries, C. (2001). Studying clinical reasoning. Part 2: Applying social judgment theory. *British Journal of Occupational Therapy, 64,* 285-292.

Henning, K., & Feder, L. (2005). Criminal prosecution of domestic violence offenses: An investigation of factors predictive of court outcomes. *Criminal Justice and Behavior*, *32,* 612-642.

Hertwig, R. (2006). Do legal rules rule behavior. In C. Engel & G. Gigerenzer (Eds.), *Heuristics and the law* (pp. 391-411). Cambridge, MA: MIT Press.

Hertwig, R., Davis, J. N., & Sulloway, F. J. (2002). Parental investment: How an equity motive can produce inequality. *Psychological Bulletin*, *128(5),* 728-745.

Hillsman, S. T. (1990). Fines and day fines. *Crime and Justice***,** *12,* 49-98.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57,* 116–131.

Hogarth, R. M., & Karelaia, N. (2005). Simple models for multiattribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Management Science*, *51,* 1860-1872.

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review, 114,* 733.

Johnson, B. D. (2006). The multilevel context of criminal sentencing: Integrating judge- and county-level influences. *Criminology, 44,* 259-297.

Johnson, J. G., & Raab, M. (2003). Take the first: Option-generation and resulting choices. *Organizational Behavior and Human Decision Processes*, *91,* 215-229.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132,* 133-156.

Kautt, P., & Spohn, C. (2002). Crack-ing down on black drug offenders? Testing for interactions among offenders' race, drug type, and sentencing strategy in federal drug sentences. *Justice Quarterly*, *19,* 1-35.

Kautt, P.M. (2002). Location, location, location: Interdistrict and intercircuit variation in sentencing outcomes for federal drug-trafficking offenses. *Justice Quarterly, 19,* 633-671.

Komorita, S. S. (1964). A model for decision-making under risk. *American Journal of Psychology*, *77,* 429-436.

Konečni, V. J., & Ebbesen, E. B. (1979). External validity of research in legal psychology. *Law and Human Behavior, 3,* 39-70.

Konečni, V. J., & Ebbesen, E. B. (1982). An analysis of the sentencing system. In V. J. Konečni & E. B. Ebbesen (Eds.), *The criminal justice system: A social-psychological analysis* (pp. 293-332). San Francisco: W.H. Freeman and Company.

Konečni, V. J., & Ebbesen, E. B. (1984). The mythology of legal decision making. *International Journal of Law and Psychiatry, 7,* 5-18.

Langer, W. (1994). *Staatsanwälte und Richter, Justitielles Entscheidungsverhalten zwischen Sachzwang und lokaler Justizkultur* [Prosecutors and judges, legal decision making between necessity and local legal culture]. Stuttgart: Enke.

Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology & Aging*, 22, 796-810.

Meier, B-D. (2001). *Strafrechtliche Sanktionen* (2.Aufl.) [Criminal sanctions (2nd ed.)]. Berlin: Springer.

Mellers, B. A. (1980). Configurality in multiple-cue probability learning. *The American Journal of Psychology*, 93, 429-443.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Mösl, A. (1981). Zum Strafzumessungsrecht [On sentencing law]. *Neue Zeitung für Strafrecht*, 131-135.

Mösl, A. (1983). Zum Strafzumessungsrecht [On sentencing law], *Neue Zeitung für Strafrecht*, 160-164.

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best". *Organizational Behavior & Human Decision Processes, 91,* 82-96.

Ojmarrh, M. (2005). A meta-analysis of race and sentencing research: Explaining the inconsistencies. *Journal of Quantitative Criminology, 21,* 439-466.

Oswald, M.E. (1994). *Psychologie des richterlichen Strafens* [Psychology of judicial sentencing]. Stuttgart: Enke.

Parducci, A. (1974). Context effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (vol. 2). New York: Academic Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25,* 111-163.

Raftery, A. E., Madigan, D., & Hoeting, J. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association, 92,* 179-191.

Rieskamp, J. (2006). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32,* 1350-1371.

Rieskamp, J. (2008). The importance of learning in probabilistic inferences. *Journal for Judgment and Decision Making, 3,* 261-277.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135,* 207-236.

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica, 127,* 258-276.

Ruback, R. B., & Wroblewski, J. (2001). The federal sentencing guidelines: Psychological and policy reasons for simplification. *Psychology, Public Policy and Law, 7,* 739-775.

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The Distortion of Information during Decisions. *Organizational Behavior and Human Decision Processes, 66*, 102-110.

Schäfer, G. (2001). *Praxis der Strafzumessung* (2. Aufl.) [The practice of sentencing (2nd ed.)]. München: Beck.

Schünemann, B. (1988). Daten und Hypothesen zum Rollenspiel zwischen Richter und Staatsanwalt bei der Strafzumessung [Data and hypotheses on the role play between judges and prosecutors in sentencing]. In G. Kaiser, H. Kury, & H. J. Albrecht (Eds.), *Kriminologische Forschung in den 80er Jahren. Projektberichte aus der BRD* (pp. 265-280). Freiburg: Max Planck Institute for Foreign and International Criminal Law.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129-138.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4(2),* 108-131.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance, 23,* 86-112.

Svenson, O. (1996). Decision making and the search for fundamental psychological regularities: What can be learned from a process perspective? *Organizational Behavior & Human Decision Processes, 65*, 252-267.

Tata, C. (1997). Conceptions and representations of the sentencing decision process. *Journal of Law and Society, 24,* 395–420.

Tata, C. (1998). The application of judicial intelligence and 'rules' to systems supporting discretionary judicial decision-making. *Artificial Intelligence and Law, 6,* 203-230.

Theune, W. (1985a). Grundsätze und Einzelfragen der Strafzumessung; aus der Rechtsprechung des Bundesgerichtshofs (Teil 1) [Principles and individual issues of sentencing: From the jurisdiction of the Federal Court of Justice in Germany (part 1)]. *Strafverteidiger, 4,* 162-168.

Theune, W. (1985b). Grundsätze und Einzelfragen der Strafzumessung; aus der Rechtsprechung des Bundesgerichtshofs (Teil 2) [Principles and individual issues of sentencing: From the jurisdiction of the Federal Court of Justice in Germany (part 2)]. *Strafverteidiger, 5,* 205–210.

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science, 16,* 167-171.

Tröndle, H., & Fischer, T. (2007). Strafgesetzbuch und Nebengesetze (54. Aufl) [Penal code and comments (54th ed.)]. München: Beck.

Van Duyne, P. (1987). Simple decision making. In D. C. Pennington & S. Lloyd-Bostock (Eds.), *The psychology of sentencing: Approaches to consistency and disparity* (pp. 143-158). Oxford: Centre for Socio-Legal Studies.

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General, 137,* 73-96.

von Helversen, B., & Rieskamp, J. (2009). Models of Quantitative Estimations: Rule-based and Exemplar-Based Processes Compared. *Journal of Experimental Psychology: Learning, Memory & Cognition, 4*, 867-889.

Wagenmakers, E.J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin and Review, 14*, 779-804.

Zedeck, S., & Kafry, D. (1977). Capturing rater policies for processing evaluation data. *Organizational Behaviour and Human Performance, 18,* 269-294.

Appendix A

Range-Frequency Theory

According to range-frequency theory (Parducci, 1974), human judgments of magnitudes and size are context dependent, that is, they depend on the range of the stimulus values as well as on the frequency with which a stimulus value appears. The judged magnitude *J* of a stimulus *i* is given by the weighted sum of the range value *R* and the frequency value *F* (cf. Parducci, 1974, p. 209):

$$J_i = w\,R_i + (1 - w)\,F_i\ , \tag{A1}$$

with $0 < w < 1$. The range value *R* represents the proportion of the current range below the current stimulus S_i:

$$R_i = (S_i - S_{\mathrm{min}}) / (S_{\mathrm{max}} - S_{\mathrm{min}}), \tag{A2}$$

where S_i denotes the current stimulus value and S_{min} and S_{max} are respectively the smallest and the largest stimulus in the set. The frequency value F_i represents the proportion of all current values below the current stimulus:

$$F_i = (r_i - 1) / (N - 1), \tag{A3}$$

where F_i represent the frequency value of the stimulus *i*, r_i is the rank of stimulus *i,* and *N* the number of stimuli in the set.

Appendix B

Bayesian Model Averaging

The Bayesian information criterion (BIC) gives the odds with which a specific model is preferred to a baseline model. To calculate a model's BIC′ value we compared it with the null model (a baseline model with no independent variables), following Raftery (1995, Equation 26, p. 135):

$$BIC'_k = n\log\left(1 - R_k^{\,2}\right) + q_k \log n\,, \tag{B1}$$

where R_k^2 is the value of R^2 for model M_k, q_k is the number of free parameters for that model, and n is the number of data points. The BIC'_k gives the BIC value for the null model compared to the model M_k. The BIC' of the null model is zero. Accordingly, if the BIC'_k is positive, the null model is preferred to the model M_k. However, if the BIC'_k is negative, model M_k is preferred to the null model, and the smaller the BIC'_k the more M_k is supported by the data.

The posterior probability of a model is defined as:

$$p\left(M_k \middle| D\right) \approx \exp\left(-\tfrac{1}{2} BIC_k\right) / \sum_{l=1}^{K} \exp\left(-\tfrac{1}{2} BIC_l\right), \tag{B2}$$

(cf. Raftery, 1995, Equation 35, p. 145), where p gives the probability of model M_k given the data D in comparison with all models from set K assuming an equal prior probability of $1/k$ for all models.

The posterior probability pr that a factor B has an effect ($B \neq 0$) is given by the sum of the posterior probabilities of all models $p\left(M_k \middle| D\right)$ that include B, here referred to as model set A:

$$pr\left[B \neq 0 \middle| D\right] = \sum_{A} p\left(M_k \middle| D\right), \tag{B3}$$

(cf. Raftery, 1995, Equation 36, p. 145).

The beta weight and the standard error of the beta weights can be estimated by an approximation to a Bayesian point estimator and an analogue of the standard error. Approximations are given by

$$E[\beta_1 | D, \beta_1 \neq 0] \approx \sum_A \hat{\beta}_1(k) p'(M_k | D) \tag{B4}$$

(cf. Raftery, 1995, Equations 38 and 39, p. 146), where

$p'(M_k | D) = p(M_k | D) / pr[\beta_1 \neq 0 | D]$, E denotes the expected value of the beta weight β_1, and $\hat{\beta}_1(k)$ is the maximum likelihood estimator of β_1 under Model M_k.

Respectively, the standard error can be approximated by

$$SD^2[\beta_1 | D, \beta_1 \neq 0] \approx \sum_A \left[se_1^2(k) + \hat{\beta}_1(k)^2\right] p'(M_k | D) - E[\beta_1 | D, \beta_1 \neq 0]^2, \tag{B5}$$

where $se_1^2(k)$ is the standard error of β_1 under Model M_k (cf. Raftery, 1995, p. 146).

Appendix C

Correlation matrices of the factors considered in the analyses

Table C1 Intercorrelation between all factors considered for fines

	Age	Prior conviction	No. charges	Net worth property	Confession	Penalty Order	Summary penalty	Aggr. evidence	Mitigating evidence II	Remorse
Prior convictions	0.24									
No. Charges	0.20	-0.07								
Net worth property	0.15	0.29	0.12							
Confession	**0.31***	0.01	0.29	0.11						
Penalty Order	0.29	0.00	0.01	-0.04	**0.71***					
Summary penalty	0.19	0.02	**0.68***	0.21	**0.41***	0.26				
Agravating evidence	0.19	0.13	**0.64***	0.29	0.27	-0.16	**0.39***			
Mitigating evidence II	**0.33***	**0.50***	0.05	0.21	0.29	**0.37***	0.25	0.14		
Remorse	0.09	0.08	0.10	0.18	**0.57***	**0.40***	0.10	0.09	0.15	
Nationality	0.18	0.00	-0.04	0.00	**0.38***	**0.53***	0.07	-0.09	**0.35***	0.21

Note. N = 44; Please note that factors mitigating evidence II, confession and remorse were recoded to ensure a positive correlation with the sentence magnitude; * $p < .05$

Table C2 Intercorrelations between all factors considered for incarceration

	No. charges	Dim. capacity	Net worth property	Summary penalty	Aggr. evidence	Mit. evidence II	No. offender	Con-fession	Re-morse	Pro-bation	Last sentence: fine
Dim. Capacity	-0.03										
Net worth property	0.09	-0.03									
Summary penalty	**0.43***	**0.35***	**0.32***								
Aggr. Evidence	0.30	**0.49***	**0.47***	**0.56***							
Mitigating evidence II	0.18	0.04	0.15	**0.40***	0.20						
confession	0.16	0.21	0.15	**0.38***	**0.35***	**0.33***					
No. offender	0.29	-0.06	**0.56***	0.12	0.24	-0.08	0.09				
Remorse	-0.28	0.10	0.10	-0.30	-0.12	-0.20	**-0.50***	0.06			
Probation	0.10	**-0.42***	0.12	-0.07	-0.18	0.04	-0.06	0.24	-0.07		
Last sentence: fine	**0.34***	-0.05	0.28	0.17	0.30	0.22	0.19	**0.35***	-0.31	0.12	
Last sentence: incarceration	-0.14	**0.33***	-0.14	0.09	0.07	-0.05	0.16	**-0.44***	0.09	**-0.56***	**-0.45***

Note. N = 40; Please note that factors mitigating evidence II, probation status and remorse were recoded to ensure a positive correlation with the sentence magnitude; * $p < .05$

Appendix D

Further model tests

Besides a linear regression model we also tested a disjunctive model, a conjunctive model (Einhorn, 1970), and a unit weight linear model (Dawes, 1979). The *BIC* values and R^2 of the best models of each type are reported in Tables D1 and D2.

Table D1: Performance of the models in predicting the magnitude of fines

	Best model		*Second-best model*	
	BIC'	R^2	*BIC'*	R^2
Mapping model class	-56	.74	-55	.74
Linear regression model class	-31	.68	-30	.65
Conjunctive model class	-38	.70	-37	.72
Disjunctive model class	-23	.57	-22	.53
Unit weight linear model class	-38	.65	- 37	.64

Notes. The two best models were selected out of 2,048 models reflecting all combinations of the 11 factors considered for influencing the magnitude of the fines.

Table D2: Performance of the models in predicting the length of incarceration sentence

	Best model		*Second-best model*	
	BIC'	R^2	*BIC'*	R^2
Mapping model class	-74	.86	-73	.85
Linear regression model class	-59	.88	-58	.87
Conjunctive model class	-25	.72	-25	.72

Disjunctive model class	-28	.72	-27	.76
Unit weight linear model class	-60	.82	-59	.81

Notes. The best model was selected out of 4,096 models reflecting all combinations of the 12 factors ultimately considered for influencing the length of incarceration terms.

Footnotes

1. The number of categories could be considered a free parameter of the mapping model that could depend on the properties of the decision task as well as on the decision maker. However, for the sake of simplicity, we assumed a priori a default value of seven for number of categories. Higher numbers of categories become unreasonable due to memory limitations and lower numbers give the decision maker less flexibility in the judgments.

2. Because the dichotomization of nominal and ordinal variables might exclude information that is relevant for sentencing, we additionally created dummy variables to capture the other cue values, and in the final phase of model testing, we tested whether including these variables would improve the performance of the models. As this was not the case, we do not report the results.

3.We excluded probation as a dependent variable because the decision if probation is given or not is a secondary decision depending on variables such as the probability of a repeated offense which should not affect the judgment of sentence length.

Table 1

Overview of the categorization system

Factors	Description	Values	Mean (std)/base rate for code ‘0’
Offender information			
Gender	female vs. male	0 vs. 1	18%
Nationality	German vs. non-German	0 vs. 1	12%
Age		20–80 years	36.49 (13.84)
Family status	Married (with and without children) or single parents vs. single and no children	0 vs. 1	79%
Occupational status	Employed, apprenticed, or student vs. unemployed	0 vs. 1	71%
Economic status	Above poverty line vs. below poverty line (ca. €900 per month)	0 vs. 1	65%
Diminished capacity	No diminished capacity vs. diminished capacity (diminished capacity was assumed if the defendant had a psychological or medical diagnosis of a mental or organic	0 vs. 1	90% (just for incarceration)

Factors	Description	Values	Mean (std)/base rate for code '0'
	disorder)		
No. of prior convictions		0–14	3.51 (3.55)
Type of last sentence	Fine or incarceration	Dummy coded	$N_{\text{Fine}} = 26$ $N_{\text{Incarceration}} = 14$
Probation status	Offender was not on probation when the offense was committed vs. was on probation	0 vs. 1	78% (just for incarceration)
Offense characteristics			
Net worth of property violated		€0–80,000	2,497 (8,826)
No. of charges		1–112	4.34 (17.01)
No. of offenders		1–3	1.25 (.46)
Mitigating evidence I	Coded as a summary factor; one point was added for whether there was external pressure to commit the crime (e.g., an emergency situation or blackmail), the crime was a failed attempt, the offender's role was secondary, or the	0–2	.19 (.45)

Factors	Description	Values	Mean (std)/base rate for code ‘0’
	offender’s capacity was diminished due to alcohol		
Mitigating evidence II	One point was added for whether the offender had no prior convictions or the net worth of property violated was below €30	0–2	.55 (.59)
Remorse	Defendant showed no remorse vs. showed remorse, offered reparation or amends	0 vs. 1	61%
Confession	Defendant did not confess vs. defendant confessed	0 vs. 1	31%
Aggravating evidence	One point was added for whether any of the following conditions was fulfilled: a high number of offenses (> 5), over a long period of time (> 6 month); the offense was carefully planned; perseverance in the face of obstacles; incited others to commit the crime; used unnecessary violence	0–2	.21 (.49)

Factors	Description	Values	Mean (std)/base rate for code ‘0’
Legal regulations			
Offense type	Theft, fraud, or forgery	Dummy coded	$N_{\text{Theft}} = 49$ $N_{\text{Fraud}} = 20$ $N_{\text{Forgery}} = 15$
Summary penalty	A summary penalty was not given vs. a summary penalty was given	0 vs. 1	25 %
Penalty order	Sentencing by trial vs. sentencing by penalty order	0 vs. 1	80% (just for fines)
Sentencing range	Max. sentence 5 years vs. max. sentence 10 years	0 vs. 1	82%

Note: Mean and standard deviation were reported for variables were an interval scale was assumed. For nominal variables the percentage of cases with the code ‘0’ is reported. For the variables penalty order the base rate is limited to the cases punished with a fine, because only three penalty orders were issued in cases punished with incarceration. Similarly, the base rate for probation status and diminished capacity are only for cases punished by incarceration because all offenders that violated probation were punished by incarceration and no offender punished with a fine met the criteria for diminished capacity.

Table 2

Results of correlation analysis and model comparison for fines

	Fines (no. of days)	Five best models					Probability	Beta	*SD*
	Pearson correlation (*p* values)	Mapping model 1	Mapping model 2	Mapping model 3	Mapping model 4	Mapping model 5			
Age	.34 (.02)	○	○	○	○	○	0.08	-0.07	0.01
No. of prior convictions	.32 (.03)	●	○	●	○	○	0.57	0.14	0.02
No. of charges	.36 (.02)	●	●	○	●	○	0.64	0.19	0.03
Net worth of property	.46 (.001)	●	●	●	●	●	0.97	0.33	0.01
Confession	-.50 (.001)	○	○	○	○	○	0.15	-0.11	0.04
Penalty order	.50 (.001)	●	●	●	●	●	0.91	0.38	0.02
Summary penalty	.53 (.001)	○	○	●	○	●	0.46	0.19	0.02
Aggravating evidence	.39 (.01)	●	○	●	○	○	0.59	0.25	0.01
Mitigating evidence II	-.48 (.001)	○	●	○	○	○	0.25	0.39	0.01
Remorse	-.20 (.20)	○	○	○	●	○	0.13	-0.07	0.01

	Fines (no. of days) Pearson correlation (p values)	Five best models					Probability	Beta	*SD*
		Mapping model 1	Mapping model 2	Mapping model 3	Mapping model 4	Mapping model 5			
Nationality	.32 (-03)	○	○	○	○	○	0.14	0.02	0.01
PMP		.15	.13	.09	.07	.06			
BIC′		-56	-55	-55	-54	-54			
R^2		.74	.74	.73	.73	.73			

Note. N= 44; Probability denotes the probability that the factor had an effect and is given by Equation B3 (Appendix B). BIC′ denotes the Bayesian Information Criterion. PMP denotes posterior model probability. An open circle denotes that a factor was not included in the model; a solid circle denotes that a factor is included in the model. For the analyses, the factors confession, remorse, and mitigating evidence II were recoded so that they correlated positively with sentence magnitude. The five best models all belonged to the class of mapping models.

Table 3

Results of correlation analysis and model comparisons for incarceration

	Incarceration (no. of months) Pearson correlation (*p* value)	Five best models					Probability	Beta	*SD*
		Mapping model 1	Mapping model 2	Mapping model 3	Mapping model 4	Mapping model 5			
No. of charges	.40 (.01)	●	●	○	●	○	0.74	0.34	0.005
Diminished capacity	.41 (.01)	●	●	●	●	●	0.97	0.45	0.006
Net worth of property	.62 (.001)	○	○	○	●	○	0.26	0.58	0.006
Summary penalty	.65 (.001)	●	●	●	○	●	0.86	0.11	0.009
Aggravating evidence	.58 (.001)	●	●	●	○	●	0.81	-0.05	0.01
Mitigating evidence II	-.41 (.01)	●	●	●	●	●	0.94	0.26	0.005
No. of offenders	29 (.07)	○	●	○	○	●	0.32	-0.08	0.006
Confession	.31(.05)	○	○	○	●	○	0.23	0.036	0.009
Remorse	-.01 (.98)	●	●	●	○	●	0.80	0.036	0.006

	Incarceration (no. of months) Pearson correlation (*p* value)	Five best models					Probability	Beta	*SD*
		Mapping model 1	Mapping model 2	Mapping model 3	Mapping model 4	Mapping model 5			
Probation status	-.07 (.65)	●	●	●	●	●	0.99	0.35	0.007
Last sentence: fine	.11 (52)	○	○	○	○	○	0.01	-0.12	0.007
Last sentence: incarceration	.17 (.29)	●	●	●	●	●	0.99	0.35	0.008
PMP		0.38	0.24	0.07	0.06	0.04			
BIC′		-74	-73	-70	-70	-69			
R^2		0.86	0.85	0.84	0.84	0.84			

Note. $N = 40$; Probability denotes that the probability that the factor has an effect and is given by Equation B3 (Appendix B). BIC′ denotes the Bayesian Information Criterion PMP denotes the posterior model probability. A solid circle denotes that a factor was included in the model; an open circle denotes that a factor was not included in the model. For the analyses, the factors mitigating evidence II and remorse were recoded so that they correlated positively with sentence magnitude. The five best models all belonged to the class of mapping models.

Figure Captions

Figure 1.

This shows an example case and how it is solved according to a linear regression model, a unit weight model, and the mapping model. The linear regression model (upper right panel) assumes that the cue values are weighed by their importance, resulting in each factors contribution to the sentence. These are summed up to form the final sentence. The unit weight model (middle panel) assumes a similar process with the difference that each factor is equally important, that is, each cue value is (after a normalization procedure) multiplied by the same weight. Then the sentence is adjusted respective to each factor. Lastly, the mapping model (lower panel) assumes that a severity score is determined for each case by taking the mean of the (standardized) cue values. Sentencing then depends on a categorization processes. Based on the severity score, the case is categorized into one of seven categories by comparing it to the category boundaries. The chosen sentence depends on the typical sentence of this category.

Figure 2.

The posterior model probability of the best models. The upper panel shows the results for fines giving the 1,000 best of all 4,096 models to describe the fining process, differentiated by model class. Of the 1,000 best models, 97% belong to the class of mapping models and 3% to the class of conjunctive models. The lower panel shows the posterior model probability of the best 1,000 models describing the incarceration decisions, differentiated by model class. Of the 1,000 best models, 90% belong to the class of mapping models, 9% to the class of linear regression models, and 1% to the class of unit linear weight models.

Figure 3.

Mean ratings of the factors relevant for sentencing and the probability of each factor of having an influence on sentencing. For better comparison we transformed the mean ratings from a ratings scale from 1 to 7 to a scale from 0 to 1. The upper panel shows the comparison of the ratings with the modeling results for fines. The lower panel for incarcerations. The sample size for ratings was 54.

Figure 1.

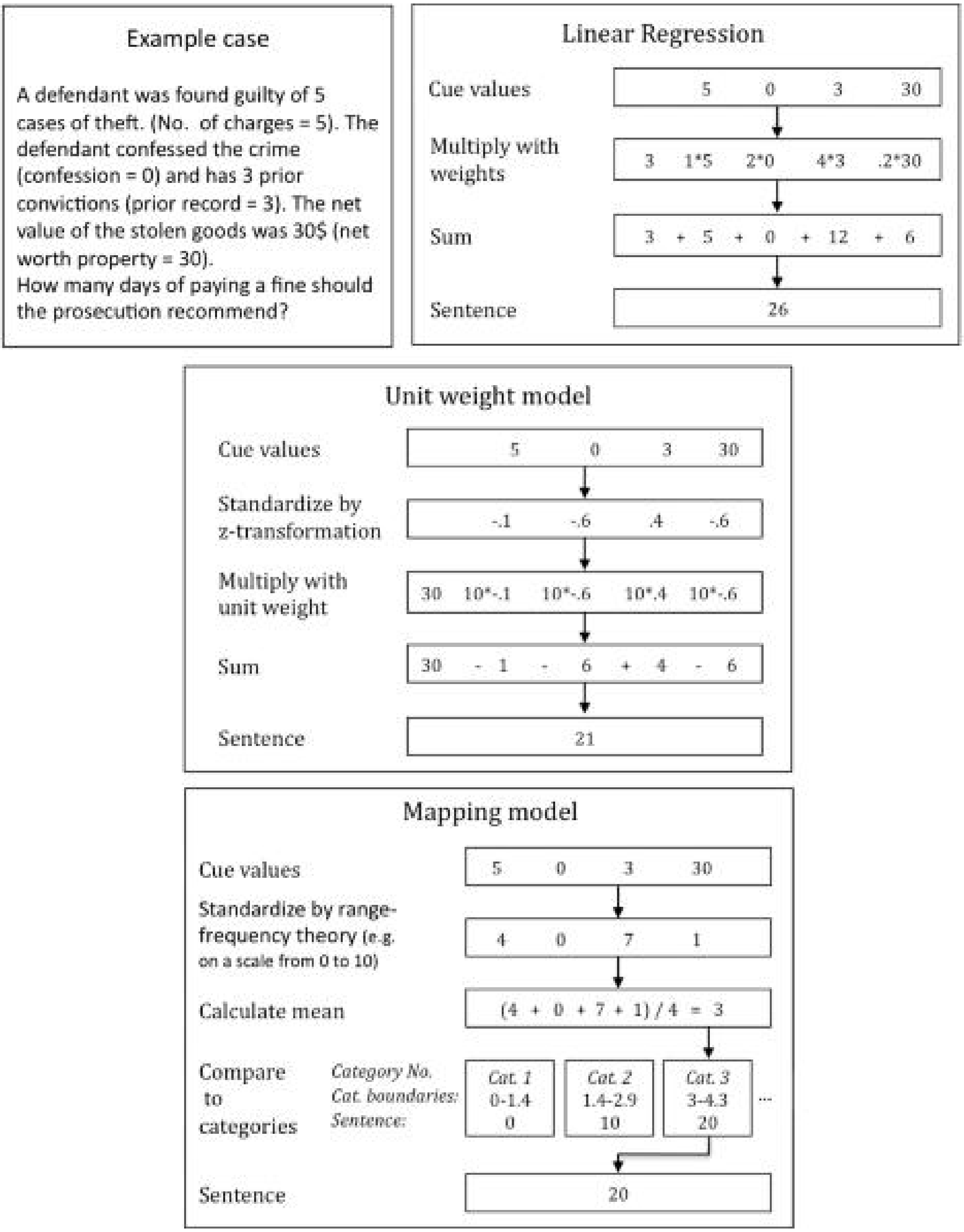
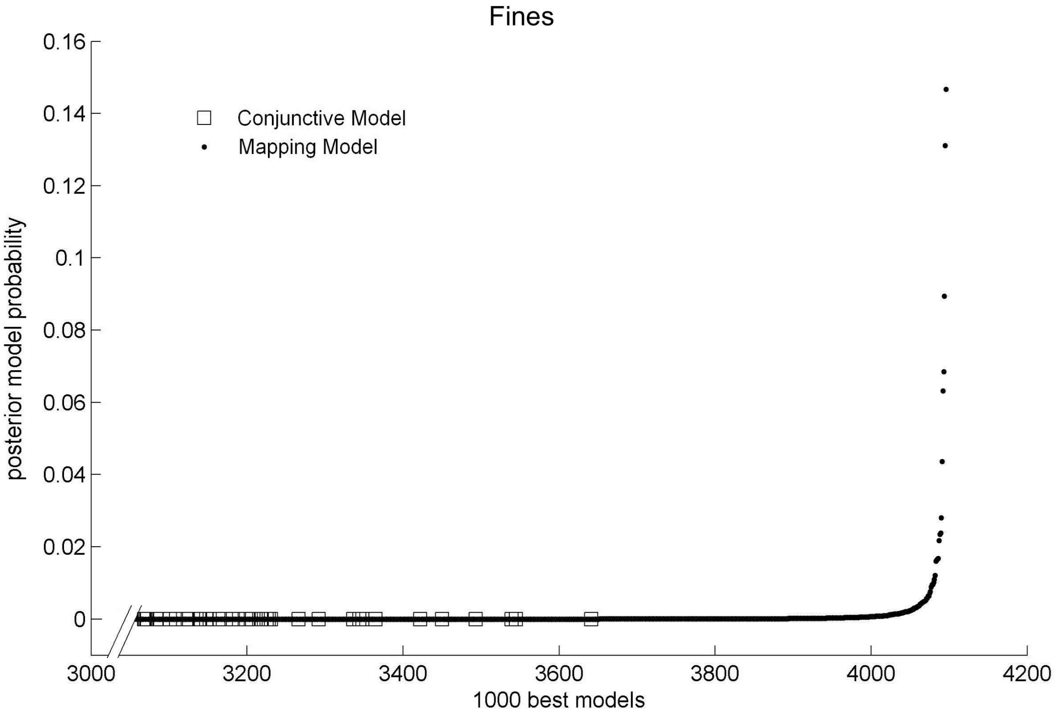

Example case
A defendant was found guilty of 5 cases of theft. (No. of charges = 5). The defendant confessed the crime (confession = 0) and has 3 prior convictions (prior record = 3). The net value of the stolen goods was 30$ (net worth property = 30).
How many days of paying a fine should the prosecution recommend?
Linear Regression
Cue values
5 0 3 30
Multiply with weights
3 1*5 2*0 4*3 .2*30
Sum
3 + 5 + 0 + 12 + 6
Sentence
26
Unit weight model
Cue values
5 0 3 30
Standardize by z-transformation
-.1 -.6 .4 -.6
Multiply with unit weight
30 10*-.1 10*-.6 10*.4 10*-.6
Sum
30 - 1 - 6 + 4 - 6
Sentence
21
Mapping model
Cue values
5 0 3 30
Standardize by range-frequency theory (e.g. on a scale from 0 to 10)
4 0 7 1
Calculate mean
(4 + 0 + 7 + 1) / 4 = 3
Compare to categories
Category No.
Cat. boundaries:
Sentence:
Cat. 1
0-1.4
0
Cat. 2
1.4-2.9
10
Cat. 3
3-4.3
20
...
Sentence
20

Figure 2


Fines
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
posterior model probability
Conjunctive Model
Mapping Model
3000
3200
3400
3600
3800
4000
4200
1000 best models

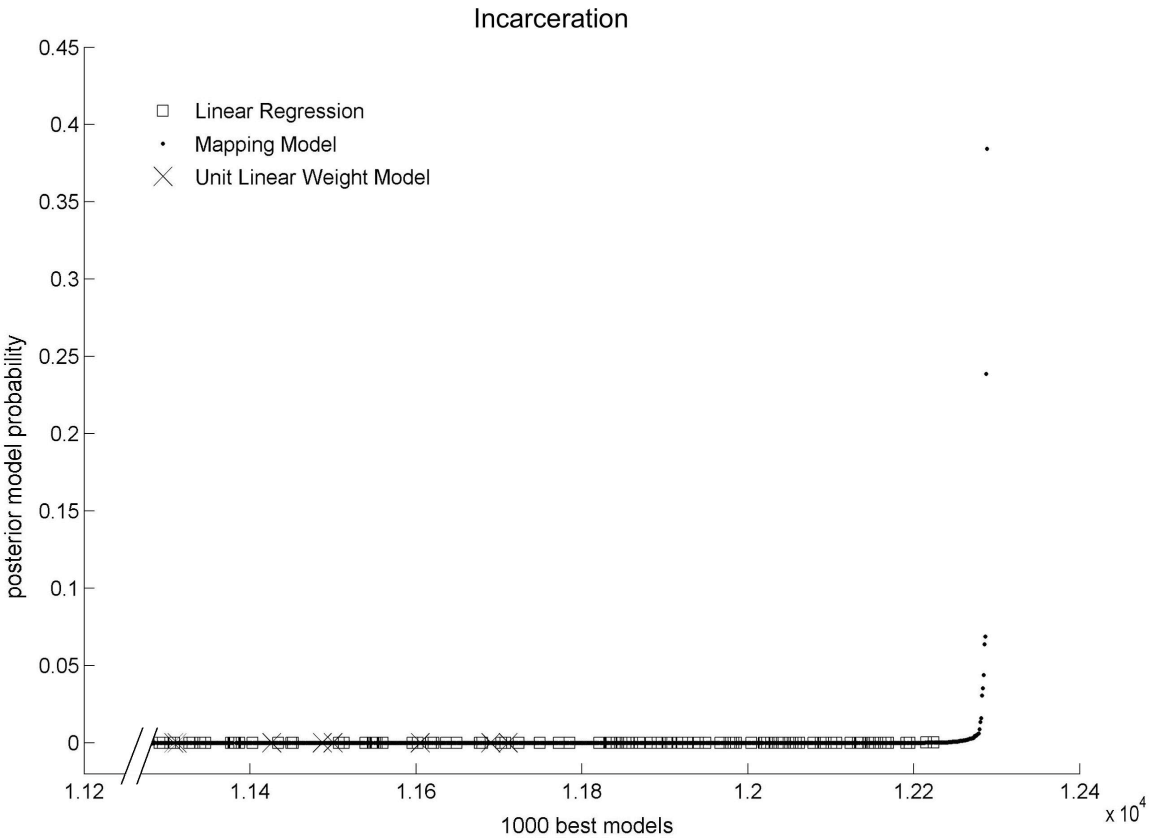
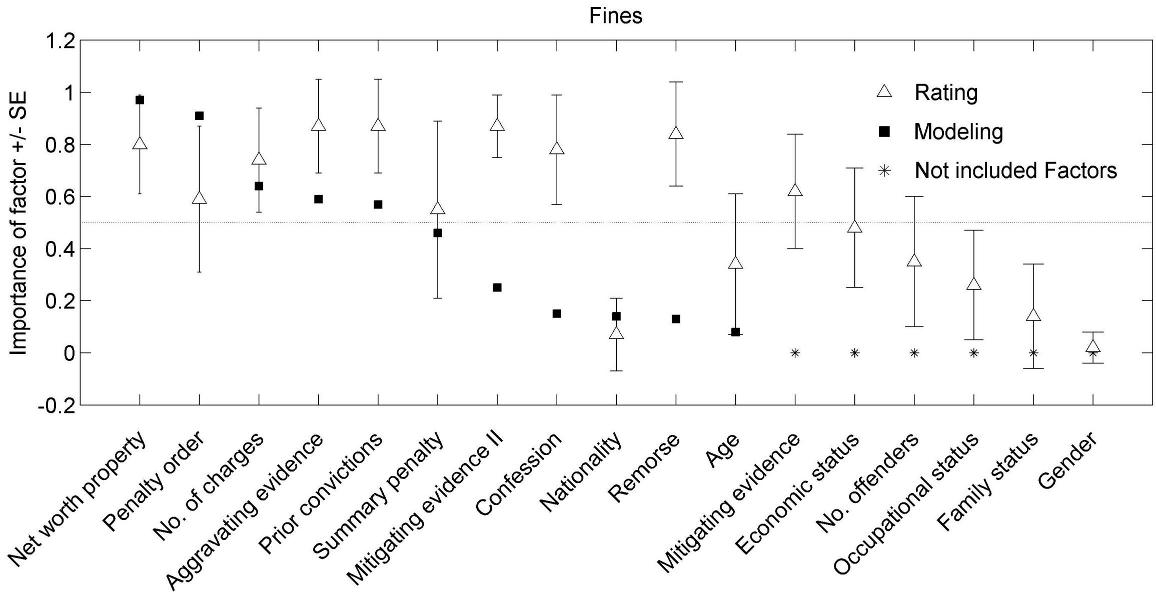
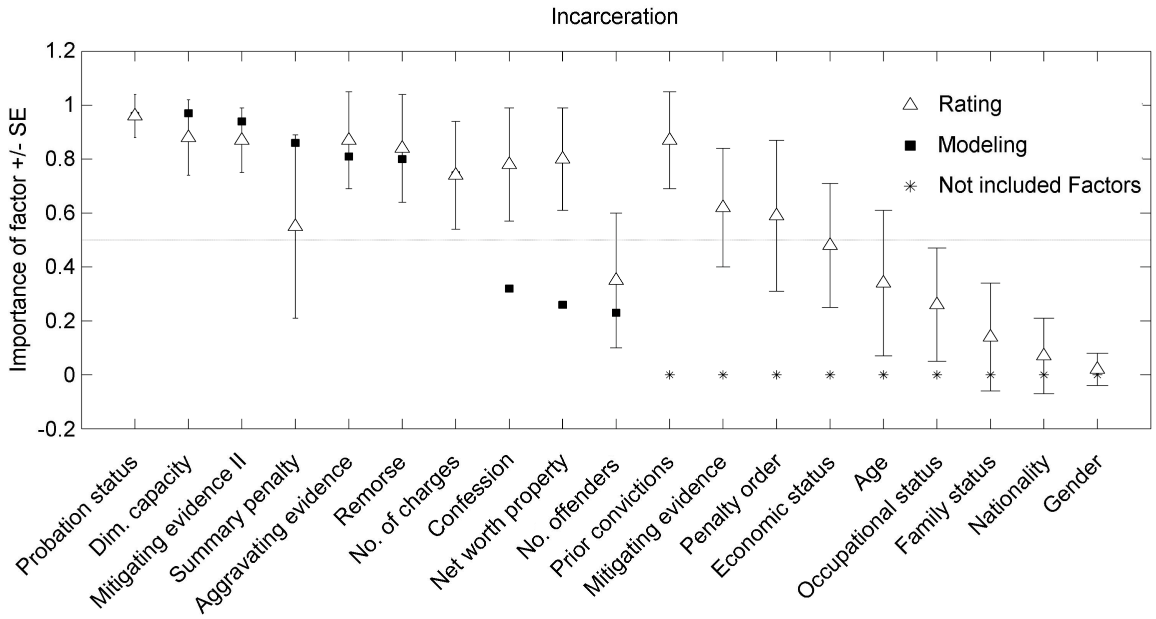

Incarceration
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
posterior model probability
Linear Regression
Mapping Model
Unit Linear Weight Model
1.12
1.14
1.16
1.18
1.2
1.22
1.24
x 10^4
1000 best models

Figure 3


Fines
Importance of factor +/- SE
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
Rating
Modeling
Not included Factors
Net worth property
Penalty order
No. of charges
Aggravating evidence
Prior convictions
Summary penalty
Mitigating evidence II
Confession
Nationality
Remorse
Age
Mitigating evidence
Economic status
No. offenders
Occupational status
Family status
Gender


Incarceration
Importance of factor +/- SE
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
Rating
Modeling
Not included Factors
Probation status
Dim. capacity
Mitigating evidence II
Summary penalty
Aggravating evidence
Remorse
No. of charges
Confession
Net worth property
No. offenders
Prior convictions
Mitigating evidence
Penalty order
Economic status
Age
Occupational status
Family status
Nationality
Gender