Bailing and Jailing the Fast and Frugal Way

MANDEEP K. DHAMI1* and PETER AYTON2
1Max Planck Institute for Human Development, Berlin, Germany, and City University, London, UK
2City University, London, UK

ABSTRACT

Legal decisions such as the decision to bail upon adjourning a case have major consequences for both defendants and society. In the English system, magistrates, most of whom are lay people, are afforded considerable discretion and must work under constraints such as time pressure. Judgment analysis of the bail decision making policies of 81 magistrates from 44 courts throughout England and Wales revealed intra- and inter-magistrate inconsistency in bail decisions, discrepancies between stated and elicited cue use, and high levels of post-decisional confidence. Furthermore, magistrates’ policies were better described and predicted by a fast and frugal model characterized by noncompensatory cue use, than by either of two compensatory integration models. The fast and frugal model portrays a picture of bail decision making that conflicts with the ideal practice as defined by the due process model of justice. We discuss the implications of these findings for judgment and decision making research and criminal justice policy. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS judgment analysis; fast and frugal; legal decision making; bail

Decisions made in the criminal justice system are regularly criticized by organizations supporting victims, groups representing defendants and professional agencies such as the prosecution service. Criticism is directed both at existing legal rules and procedures and the individuals who are formally trained to apply them. The motivation for such scrutiny lies in the belief that crime and order are socially constructed and so can be reconstructed, and the realization that legal decisions have consequences for both the public purse and the lives of the public.

Past psychological research on legal decision making includes studies investigating particular decisions made by judges (e.g. Ebbesen and Konecni, 1975; Sensibaugh and Allgeier, 1996) and studies of jury decision making (e.g. Klevorick et al., 1984; Pennington and Hastie, 1981). Although both types of research provide scope for challenging and changing existing legal rules and procedures, only the former can also critique the performance of individuals who are formally trained and entrusted to apply them on a regular basis.

Many published studies of legal decisions made by judges have tended to focus on decisions made in the American criminal justice system. For example, Ebbesen and Konecni (1975) examined the information

*Correspondence to: Mandep K. Dhani, Department of Psychology, City University, Northampton Square, London EC1V 0HB, UK. E-mail: M.K.Dhani@city.ac.uk

Copyright © 2001 John Wiley & Sons, Ltd.
used by San Diego County judges when setting bail. To our knowledge, judgment and decision making researchers have not studied legal decisions made by judges in the English system. It may not be safe to generalize the findings of research from one jurisdiction to others because although the decisions may be common to different systems, the legal rules and procedures, and the characteristics of the individuals trained to apply them are often peculiar to a particular jurisdiction. For example, the bail decision (whether defendants should be allowed to ‘go free’ until the next hearing of the case or whether they should be held in custody) differs in the American and English systems.

In the present study we empirically investigate bail decisions made by individuals who are appointed and trained to make these decisions in the English system. The bail decision has major consequences for defendants, their families, the criminal justice system and the general public (see Cavadino and Gibson, 1993). In the past, research has been conducted by criminologists and they have highlighted concerns with, for instance, the nature of the information used (e.g. Huckleby, 1996; Simon and Weatheritt, 1974). We aim to determine whether these concerns are still justified and in doing so, we explore the potential impact that judgment and decision making research can have on legal policy and practice. We hope that this will prompt further studies of legal decisions made in the English system. Before highlighting criminologists’ concerns and describing the present study, we introduce the individuals who make bail decisions and describe characteristics of the decision making task. We then review the method of judgment analysis that is commonly employed to investigate judgment and decision making in applied domains.

Who are the decision makers?
In the English system, around 98% of all criminal cases are dealt with from start to finish by magistrates; the remaining cases are sent by magistrates to crown court for jury trial or sentence (Lord Chancellor’s Department, 1999). The large majority of magistrates (99.70%) are lay magistrates who are not required to have any formal legal training. They are given induction and basic training, followed by refresher training every three years after that. Lay magistrates are members of the local community who are appointed to perform judicial duties on a part-time, unpaid basis. The rationale behind the lay magistracy is that local justice be served by local people. Although lay magistrates usually act as a bench of two or three, they are allowed to make bail decisions alone. A very small minority (0.30%) of the magistracy are ‘stipendiary’ magistrates who are required to be legally qualified and have typically practiced law for a number of years. They are trained for at least two years and they perform judicial duties on a full-time, paid basis. Stipendiaries can and usually do act alone.

One of the most frequent tasks performed by magistrates is to decide whether to make a punitive or non-punitive bail decision whenever a case is adjourned for subsequent trial, sentence or appeal. Magistrates must decide whether to bail (release) a defendant unconditionally or react punitively by bailing with conditions such as a curfew or remanding a defendant in custody (imprisonment). The purpose of the bail decision is to ensure that the defendant returns to court for the next hearing of the case. In 1998, magistrates made decisions on around two million defendants (Home Office, 1999). Many of the cases would have been adjourned at least once during some point (Whittaker, et al., 1997), thus requiring magistrates to make a bail decision.

What are the characteristics of the bail decision making task and do they matter?
The law on bail
The magistrates’ bail decision making task is governed and guided by the law on bail, namely the Bail Act 1976 (reproduced in Cavadino and Gibson, 1993) with its subsequent revisions. The law stipulates the principles upon which a defendant may be granted or refused bail and the conditions that may be attached when granting bail. The law provides a general right to bail, except in some circumstances. (Note that bail may still be granted at the court’s discretion.) For example, bail may be denied if there are ‘substantial grounds’ for
believing that a defendant may abscond, or offend, or interfere with witnesses/obstruct justice. The law also states that in order to assess the risk of these events occurring, magistrates should ‘have regard to’ information concerning the nature and seriousness of the offence, the likely sentence if convicted, the defendant’s bail record, if any, the strength of the prosecution case, the defendant’s character, antecedents (previous convictions), associations and community ties, and any other factors that ‘appear to be relevant’. It is evident that the law is vague and ill-defined. It is silent on exactly what information should be used and how that information should be weighted and integrated.

Availability of information
Numerous studies have documented the lack of information available to magistrates when they make bail decisions (e.g. Dhami and Ayton, 1998, poster presented at the 1998 meeting of the Society for Judgment and Decision Making, Dallas; Hucklesby, 1996; King, 1971; Morgan and Henderson, 1998). Background information such as the defendant’s name, age and address, details of the current offence and any previous bail decisions is provided on a courtsheet. Additional information may be provided by the prosecution and defense on these matters and on other matters such as the defendant’s previous convictions, bail history, community ties and personal circumstances. However, as Hucklesby (1996) observed, additional information was provided in only one third of her sample of 1524 cases from three courts in Wales. The provision of information is particularly important because the law on bail states that insufficient information is a ground for a remand in custody. This allows the court time to gather further information on which to make an informed decision.

In 1988, Bail Information Schemes were introduced in England and Wales to collect, verify and provide information about a defendant’s community ties. Studies evaluating these schemes have found that some defendants are diverted from custody when magistrates are provided with such information (e.g. Lloyd, 1992; Stone, 1988). Thus, it seems likely that providing more information would lead to fewer punitive decisions. Research investigating the effects of missing information in other judgment domains supports the assumption that missing information leads to negative decisions. For example, in a study on personnel decision making, Jagacinski (1994) found that, on average, missing information which was considered of average importance to the decision led participants to rate a candidate less favorably than comparable candidates with complete information. Johnson and Levin (1985) found that in consumer decision making, less favorable evaluations were made of products as the amount of missing information increased.

Quality of information
When information is available, magistrates do not know how useful different information is in predicting whether a defendant if bailed unconditionally will abscond, offend or interfere with witnesses. The relative predictive validity of different pieces of information has not been objectively measured in any comprehensive way, and so is not available to magistrates when they make a bail decision. However, even if data were gathered, it would at best provide only a partial measure of the objective predictive validity of the information. This is because it would never be known how many defendants who were imprisoned would have offended if they had not been imprisoned. It is also impossible to measure exactly how many defendants who were released on unconditional bail actually offended, as many crimes are not detected or reported. Magistrates may learn the usefulness of different pieces of information from informal sources such as the media’s coverage of offending by defendants on bail. However, these sources are prone to bias and error.

Outcome feedback
An informal way of establishing the relative predictive validity of different pieces of information is for magistrates themselves to gather outcome feedback after having made a bail decision on a case. For example,
they may realize that particular defendants whom they bailed unconditionally returned to court charged with committing another offence. In the English system, there is no formal outcome feedback. Magistrates do not know if they made an appropriate decision and this may restrict their ability to learn whether they are using the right information in the right way. Learning from experience of the task is difficult (Brehmer, 1980), and is also prone to bias and peculiar to an individual.

Court procedure
There are no statutory rules of procedure for bail hearings in the magistrates’ court (Lydiate, 1987). Information may be presented in any order. However, it is not uncommon for a court to adopt a particular procedure and for specific information such as the offence the defendant is currently charged with, to generally precede other information such as the defendant’s previous convictions.

Evidence from the field of judgment and decision making suggests that the order in which information is presented affects the judgments made. Moreover, the effect is moderated by the amount of information that is provided, the complexity of information and whether the response is made after each piece of information is presented or not (Hogarth and Einhorn, 1992). Studies in other judgment domains have found evidence for a primacy effect (Adelman et al., 1996) and a recency effect (Highhouse and Gallo, 1997). In their study on legal decision making, Kerstholt and Jackson (1998) found that a recency effect occurred when participants were required to respond after all the information had been provided and background information was also available.

Time pressure
Magistrates are often faced with a heavy caseload and despite the lack of time limits this may lead to an implicit feeling of time pressure. Evidence suggests that although over the past three decades the average duration of a bail hearing has increased slightly, magistrates nevertheless make bail decisions rapidly (e.g. King, 1971; Dhami and Ayton 1998; Doherty and East, 1985). Dhami and Ayton (1998) observed that in their sample of 342 bail decisions made in two London courts, the duration of the bail hearings ranged from 50 seconds to 62 minutes, with an average of six minutes. They began timing cases from the moment the defendant stepped into the courtroom to take into account the accessing and processing of non-verbal information.

Research has revealed that time pressure affects human judgment in a number of ways. Davis and Davis (1996) found that under conditions of time pressure individuals become more inconsistent. Individuals also tend to switch to simple judgment strategies. For example, decision making strategies tended to be noncompensatory, that is, once a decision has been made further information cannot alter it (Ford et al., 1989). It has also been demonstrated that individuals do not use all of the relevant information (Rothstein, 1986; Wright, 1974). Edland (1979) found that time pressure resulted in greater selectivity of information and greater use of noncompensatory strategies. Recently, Rieskamp and Hoffrage (1999) reported that under conditions of time pressure, individuals’ judgments were better described by a simple, noncompensatory ‘fast and frugal’ strategy characterized by little information search and where decisions were based on only one piece of information.

Work pattern
Lay magistrates work on a sporadic basis because each sits in court for a minimum of 26 half days a year (i.e. a morning or afternoon every one or two weeks). This situation differs from the full-time duties of stipendiary magistrates. It is reasonable to assume that the difference in work pattern may lead to differences in the consistency with which decisions are made between the two types of magistrate. Moreover, there may be differences in the consistency of decisions made by more and less experienced magistrates. Although we
have found no studies directly comparing the effects of massed practice and distributed practice on performance in judgment tasks to support this assumption, studies of multiple cue probability learning typically provide participants with blocked trials (e.g. Todd and Hammond, 1965). These studies demonstrate that under conditions of massed practice participants can develop a relatively consistent judgment policy, in that it can be adequately modeled (i.e. predicted). Thus, the implication is that both lay magistrates and less experienced magistrates may show more inconsistency in their decisions.

**Past research on magistrates’ bail decision making**

When studying magistrates’ bail decision making, criminologists have evaluated magistrates’ performance against an ideal practice as defined by the due process model, which is synonymous with our notion of justice (Packer, 1968). The due process model aims to reduce crime whilst minimizing the number of innocent people who are wrongly convicted, by setting procedural constraints. There is a presumption of innocence. Magistrates are viewed as impartial adjudicators between the prosecution who represents the state and the defense who represents the individual, and both parties are considered equal. All the court participants must adhere to formal rules of procedure. Magistrates, for example, must carefully examine all the evidence, and treat individuals fairly and without bias. This is not necessarily related to discovering the truth (or making the ‘correct’ decision). For example, a successful appeal against a decision to convict does not establish the defendant’s innocence; it merely states that the correct procedures were not adhered to. It is assumed that if the function of the magistrates’ court is to serve justice, then due process must be observed (King, 1981) and due process ideals are often recommended as ways to regulate pre-trial decisions such as the bail decision (Galligan, 1987).

Criminologists have adopted sociological methods such as observations of bail hearings in the courtroom, document analyses of court registers and criminal statistics, and questionnaires and interview surveys of magistrates to investigate magistrates’ bail decision making (e.g. Hucklesby, 1996). The large majority of studies have tended to focus on bail decisions made in the adult magistrates’ court where defendants are aged 18 years or over.

Criminologists have investigated the information used by magistrates when making their bail decisions and differences in bail rates between courts. Although evidence has shown that magistrates respond to factors explicitly proposed by the law on bail such as the nature and seriousness of the offence (e.g. Doherty and East, 1985; Jones, 1985; Hucklesby, 1996; King, 1971; Morgan and Henderson, 1998; Simon and Weatheritt, 1974), it has also been revealed that magistrates are sensitive to other factors. Studies have found that magistrates’ bail decisions are associated with defendant-related factors such as race (see Fitzgerald, 1993), gender (e.g. Eaton, 1987; Hucklesby, 1996; Simon and Weatheritt, 1974) and age (Jones, 1985; Simon and Weatheritt, 1974). According to these studies, defendants of ethnic origin, male defendants, defendants not fulfilling traditional gender roles and younger female defendants are all more likely to be treated punitively than their counterparts. Second, there is evidence to suggest that magistrates’ bail decisions are also related to the earlier police bail decision (which is made after arrest before appearance at court) (e.g. East and Doherty, 1984; Jones, 1985; King, 1971; Morgan, 1994) and the prosecution request regarding bail (e.g. Doherty and East, 1985; Morgan and Henderson, 1998; Zander, 1979). For example, Morgan (1994) reported that in 99% of the cases in her sample where the police had decided to grant bail after charge, magistrates also granted bail. Hucklesby (1996) found that magistrates agreed with the prosecution request 95% of the time. Finally, there are differences in bail rates among courts that cannot be fully explained by the differences in the cases presented, thus highlighting disagreement between magistrates (e.g. Jones, 1985; Hucklesby, 1996; King, 1971). The picture portrayed by criminologists differs from the ideal practice as defined by the due process model.

Although the studies conducted by criminologists have high external validity because they are based on real cases, they suffer from a number of limitations, including the fact that they do not control for the
inter-correlations that may exist between variables either at the design or the analysis stage of research. (In fact, they do not always report the size of cue inter-correlations). This means that the effect of one variable such as race cannot be discerned independently of the effect of another variable such as offence. Huckleby (1996), for example, made observations, analyses of court registers and questionnaire and interview surveys of magistrates. She concluded that the prosecution request was ‘very influential’ in magistrates’ bail decisions, but she did not statistically disentangle the relationship between the prosecution request and other factors such as offence. Some studies do not control for the information available to magistrates and so they can at best only speculate as to the information attended to when magistrates made a bail decision. For example, Jones (1985) conducted an analysis of the official statistics on bail decisions and concluded that the police bail decision was the ‘most significant factor’ in magistrates’ bail decisions. He failed, however, to point out that this information may not have been available to magistrates when they made their decision. Finally, much of the criminological research on magistrates’ bail decision making reviewed above was conducted before some of the revisions were made to the Bail Act 1976. It is likely that magistrates’ bail decision making has been altered as a result of such changes; rendering much of the criminological research ‘out of date’.

**Researching human judgment and decision making**

**Judgment analysis**

The method of judgment analysis, sometimes also referred to as policy capturing, has been widely used to study judgment and decision making in a variety of applied domains such as education, medicine and accounting (see Heald, 1991; Wigton, 1996; Waller, 1988, respectively). It has also been used to investigate legal decisions made by judges in the American criminal justice system (e.g. Sensibaugh and Allgeier, 1996). Studies have typically investigated individuals’ consistency in making decisions, agreement between different individuals’ decisions on the same cases, their use of information, and their insight into their own policies.

Judgment analysis involves asking individuals to make decisions on a set of cases that may be either real or hypothetical and which comprise a combination of cues (information). Each individual’s judgment policy is then inferred from his or her behavior. This method attempts to avoid the pitfalls of direct report methods such as interviews and questionnaires that are susceptible to social desirability response bias (Arnold and Feldman, 1981) and inaccuracy in recall from memory and introspection (Wilson and Stone, 1985). Traditionally, researchers using judgment analysis have employed integration models such as multiple linear regression to analyze the judgment data obtained (see Brehmer and Joyce, 1988). An individual’s decisions are regressed on the cues presented. Judgment policies are described in terms of the number and identity of information used (i.e. cues with significant weights), and how it is weighted (e.g. standardized beta weights) and integrated (e.g. additive, compensatory rule) (Cooksey, 1996; Hammond et al., 1975).

The ability of a model to describe and predict individuals’ judgment data is limited by their consistency in making decisions because it is assumed that an inconsistent individual will be difficult to describe and predict. Researchers therefore have often presented individuals with a simple test–retest situation. In addition to modeling judgment policies, researchers have examined individuals’ insight into their own policies by comparing the model that describes an individual’s implicit policy with his or her explicit statement of policy, as elicited by a direct report method. It has been argued that direct methods provide an unreliable method for demonstrating self-insight because of the difficulties in articulating policies (e.g. Reilly and Doherty, 1992). Therefore, in the present study, we consider a policy captured via such direct methods as a statement of an individual’s explicit policy, that which he or she is consciously willing and able to express to others, rather than as an expression of insight. This is useful because legal decision makers are often asked to publicly account for their decisions. Finally, often one of the goals of research is to improve existing judgment policies by providing cognitive feedback and/or decision aids (Hammond et al., 1975), and so one way of deter-
mining whether individuals will be amenable to such intervention is to measure their degree of post-
decisional confidence. High confidence in a policy may imply an unwillingness to change it (Zakay, 1997).

Although there are some exceptions (see Brehmer and Brehmer, 1988), studies employing judgment analysis have generally yielded consistent findings, irrespective of the number and type of decision makers sampled and the nature and content of the judgment tasks studied (Brehmer and Brehmer, 1988; Libby and Lewis, 1982; Slovic and Lichtenstein, 1971). The main findings are that although over a set of judgments individuals are inconsistent, their judgment policies can be adequately represented by an integration model such as multiple linear regression, that contains only a handful of statistically significant cues. Furthermore, although there are inter-individual differences in policies, subgroups of individuals with similar policies sometimes emerge. There are also discrepancies between the cues used as indicated by an individual’s model and his or her own explicit statement of cue use (which researchers often interpret as a lack of self-insight). In addition, evidence from the field of judgment and decision making in general has shown that where an outcome criterion is known, individuals (including experts) appear highly confident in relation to their degree of accuracy (see Zakay, 1997).

**Criticisms of the regression model**

Although judgment analysis has proved fruitful in investigating human judgment and decision making in applied domains, researchers have questioned the use of regression models as descriptions of human judgment behavior (e.g. Dhani and Harries, in press; Einhorn, 1970, 1971; Gigerenzer and Goldstein, 1996; Gigerenzer et al., 1999). Regression models are structural models that provide a static description of judgment behavior where the same information is used in the same way when deciding on each case. Although cue weights may be noncompensatory and nonlinear terms may be included, it is generally assumed that judgments are the product of a linear, compensatory integration of multiple cues that are weighted optimally (Brehmer and Brehmer, 1988). However, the human mind is characterized by limited cognitive processing capacity (e.g. Kahneman, 1973; Miller, 1956) and evidence demonstrates that judgment strategies are chosen in relation to the structure and demands of the task (e.g. Billings and Marcus, 1983; Einhorn, 1971; Einhorn and Hogarth, 1981; Hammond, 1996a; Hammond et al., 1987; Mertz and Doherty, 1974; Payne, 1982; Payne et al., 1993). We have already presented studies that show, for example, how individuals switch to simpler noncompensatory strategies under conditions of time pressure. Thus, one can question the psychological plausibility of regression models.

**Are there alternative models?**

Some researchers have proposed static, structural, noncompensatory, nonlinear models. For instance, Einhorn (1970, 1971) proposed the conjunctive model and the disjunctive model. In the former model, all cue values must pass a specific threshold or cutoff point before a judgment is made at a specific level, thus cue sums are not important. In the disjunctive model, the cue values of only one cue need to pass a cutoff point before a judgment is made at a specific level. Brannick and Brannick (1989) proposed a version of the scatter model. In a choice task, this model, in addition to taking into account the weighted sum of the cue values, also takes into account the dispersion of cue values for each alternative. Therefore the pattern of cue values is considered to be more influential in making a judgment than solely the weighted sum of the values.

Generally, these researchers have found that for at least some, if not all, of the individuals in their samples, noncompensatory models provided a better fit to judgment data than compensatory models (Brannick and Brannick, 1989; Einhorn, 1970, 1971; Ganzach, 1995; Park, 1978). Einhorn (1970), for example, found that predictions from the disjunctive model correlated more highly with two of his participants’ judgments than predictions from a multiple linear regression model or the conjunctive model. The conjunctive model better
predicted the judgments of one participant and the three models failed to predict the judgments of the fourth participant. Cooksey (1996), however, points out that these models are not as parsimonious as the multiple linear regression model and so do not cross-validate well. These alternative mathematical models are also considered difficult for participants to interpret in studies aiming to improve judgment through cognitive feedback of policies (Dhami and Harries, in press). As such, these models have attracted relatively little attention from researchers using judgment analysis.

Process-tracing techniques have proved a popular alternative to regression analysis (see Juslin and Montgomery, 1999). Where regression analysis involves inferring information use from judgment behavior, process tracing involves either elicitation of verbal descriptions of information use, or identification of the information to which people attend by study of eye movement, information selection or reaction time data. The resulting process models provide a step-by-step description of information use in terms of production rules that are easy to understand. However, because the formation of these models is cumbersome, researchers focus on either modeling only one participant (Einhorn, et al., 1979) or modeling only particular tasks (Billings and Marcus, 1983). The models may also be criticized for involving artificial methods to assess what information is searched for, and for being vulnerable to the limitations of think aloud methods (e.g. socially desirable responses and difficulty in retrieving information from memory). These concerns raise doubts about the reliability and validity of the description of judgment behavior provided by process-tracing techniques (Payne et al., 1978).

Recently, Gigerenzer and Goldstein (1996) proposed the use of process models, which they call fast and frugal, and which, like regression models, are formed on the basis of the structural relationships between the cues and judgment. Thus, they are not reliant on verbal protocols or artificial methods.

**Fast and frugal models**

Fast and frugal models are simple process models that do not search through all available information, do not integrate all relevant information and base their decision on only one cue (Gigerenzer et al., 1999). Many of these models are also noncompensatory. These models consist of principles for information search, stop and decision making. For example, information search may be random or ordered and is limited by simple stopping rules such as stopping once the first cue which discriminates between two alternatives in a choice task is found, and so decisions are based upon one cue alone.

Studies have compared the fit of fast and frugal models with compensatory integration models (including regression models) on both simulated judgment data (Gigerenzer and Goldstein, 1996; Gigerenzer et al., 1998, 1999) and human judgment data (Dhami and Harries, in press; Rieskamp and Hoffrage, 1999). The former studies have evaluated the accuracy of the models in describing and predicting an outcome from statistical data (thus simulating judgments), whereas the latter studies have evaluated the models’ ability to describe human judgments made on a set of cases. The evidence reveals that fast and frugal models are as equally good as compensatory integration models at correctly describing both simulated and human judgment data (Gigerenzer and Goldstein, 1996; Gigerenzer et al., 1998, 1999; Dhami and Harries, in press; Rieskamp and Hoffrage, 1999). Moreover, it has been found that fast and frugal models are better at correctly predicting simulated judgments on a new set of cases (Gigerenzer et al., 1998, 1999).

Rieskamp and Hoffrage’s (1999) study involved inexperienced participants working under conditions of time pressure. In contrast, Dhami and Harries’ (in press) study involved participants who were trained and experienced at making professional decisions. They compared the abilities of a fast and frugal model and a logistic regression model’s ability to describe doctors’ prescription decisions on hypothetical cases. They did not, however, compare the models’ ability to predict doctors’ decisions made on a new set of cases. In the present study, we explore both the descriptive and predictive validity of a noncompensatory fast and frugal model and two compensatory integration models, by modeling experienced magistrates’ decisions to grant bail on hypothetical cases.
The present study
The rationale for the present study was to discover if criminologists’ past criticisms of magistrates’ bail decisions in the English system are justified and to describe the extent to which current practice differs from ideal practice as defined by the due process model of justice. We used the method of judgment analysis to study magistrates’ bail decisions in the adult magistrates’ court. The aims were to:

1. Measure each magistrate’s test–retest consistency in making bail decisions
2. Measure agreement between magistrates’ bail decisions made on the same set of cases
3. Measure each magistrate’s post-decisional confidence
4. Compare the relative fit of different models, namely two compensatory integration models and a non-compensatory fast and frugal model in order to select the most valid descriptor and predictor of magistrates’ bail decisions and
5. Compare the policies described by the best-fit model with magistrates’ own explicit accounts of their bail decision making policies.

It is hypothesized that magistrates’ bail decision making policies will be adequately described by a fast and frugal model, as magistrates may be using simple strategies because in the courtroom they work under constraints such as limited time and information, and as humans they have limited cognitive computational capacity. It is also hypothesized that the findings of the extent of magistrates’ consistency, agreement, post-decisional confidence and concordance between their implicit and explicit policies are likely to be compatible with the relatively stable body of findings reported by previous research using judgment analysis.

METHOD

Participants
Two hundred and seventy booklets were distributed by mail to practicing magistrates sitting in a random sample of 51 courts in England and Wales. The sampling frame was obtained from Shaw’s 1996/97 Directory of Courts in the United Kingdom (Morris, 1996). Eighty-one magistrates from 44 courts, fully completed and returned the materials within the time limit. The 30% response rate is high for a postal survey involving participants from a ‘closed group’.

The sample of magistrates who participated are considered to be representative of the magistracy in the English system. Of those who chose to reveal their demographic characteristics, 70 were lay magistrates and nine were stipendiary magistrates. Magistrates’ experience on the bench ranged from six months to 35 years, with a mean of 13 years. Forty-six magistrates sat in courts located in metropolitan areas and 32 sat in courts located in provincial areas.

Bail decision making task
The bail decision making task consisted of a set of 41 hypothetical cases on which the participants made a bail decision and then indicated their degree of confidence in their decision. In the present study, nine cues were manipulated (i.e. treated as the predictors in the models) and ten cues were held constant. The cues were identified on the basis of: (1) a review of the law on bail, (2) an analysis of a bail decision making training package for lay magistrates (Miles and Thomson, 1992), (3) semi-structured individual interviews with six lay magistrates, one stipendiary magistrate and two clerks to the justices (court managers), (4) observations of 35 bail hearings in both lay and stipendiary magistrates’ courtrooms and (5) a review of the literature on bail. The nine cues, their values and the distribution of their values are shown in Exhibit 1. Although some of these cues may not always be available to magistrates in the courtroom as mentioned earlier, their inclusion in the present study will provide us with evidence of their use when they are
### Exhibit 1. Cues, their values and the distribution of their values

<table>
<thead>
<tr>
<th>Cue</th>
<th>Valueb</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Male</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>(2) Female</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) White</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(2) Asian</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(3) Black</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) 18–20</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>(2) 21+</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Police bail decision (polbail)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Unconditional bail</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(2) Surety</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(3) Remand in custody</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Prosecution request (prosreq)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Don’t oppose bail</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(2) Conditional bail</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(3) Oppose bail</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Offence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Summary</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(2) Triable either-way</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(3) Indictable</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Previous convictions and bail record (pcbr)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) None–none</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>(2) None–good</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>(3) Yes, dissimilar–good</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>(4) Yes, similar–good</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>(5) Yes, dissimilar–poor</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>(6) Yes, similar–poor</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Strength of prosecution case (proscase)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Weak</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>(2) Strong</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Strength of community ties (commts)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Strong</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>(2) Weak</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

*aBased upon the interviews with seven lay and stipendary magistrates, the age cue was made categorical rather than continuous. In the English criminal justice system, defendants aged under 21 who are remanded in custody are sent to young offenders’ prisons, while those aged over 21 are sent to adult prisons. Summary offences are mostly minor offences which are tried only in the magistrates’ court, indictable offences are mostly serious offences which are tried only in the crown court in front of a judge and jury and triable either-way offences may be tried in either court. The specific offences chosen to represent each of these categories of offences were selected on the basis of their commonality as indicated by the 1995 recorded crime figures in England and Wales, which were the most recent at the time of design (Home Office, 1996). A poor bail record may be defined as that where the defendant has absconded, offended or interfered with witnesses whilst on bail in the past. The prosecution case against the defendant may be strong where details of witnesses and/or forensic evidence are presented at the time of the bail hearing. Finally, a defendant may be considered to have weak community ties if he or she is unemployed, does not have a fixed address or any family ties.

*bPolytomous cues were dichotomized before analysis, so the nonitalicized and italicized values indicate the two values for each cue. For each cue the nonitalicized value was coded as 0 and the italicized value was coded as 1 for analysis.

Available. The constant cues, such as the duration of the adjournment requested, were used to provide background information to the cases, while the values of the nine cues were combined to form the cases themselves.

Ultimately, we wanted to examine the effects of one cue on the decisions independently of the effects of another cue. In theory this can be achieved at either the design stage using an orthogonal design or the analysis stage using partial correlations, for example. Correlation-based analysis such as hierarchical logistic regression requires a large case to cue ratio (Tabachnick and Fidell, 1996), which would mean that in the present study at least 122 cases would be needed to study the nine cues. A pilot study on 20 law students revealed that participants would be willing to complete a maximum number of about 40 cases. Magistrates are also unlikely to volunteer for a time-consuming task. We opted to eliminate inter-correlations at the design stage. In another study (in preparation) we collected data on the inter-correlations between the nine
cues studied here from observations of 342 real cases appearing in two courts. In one court there were only four significant correlations between the nine cues and in the other court only two significant correlations were found. Therefore, we feel that, at least in the present study, little has been lost in removing the cue intercorrelations.

In fact, orthogonal designs are commonly used in judgment analysis research (Cooksey, 1996). A complete combination of the cue values would have yielded an unmanageable number of cases (i.e. $2 \times 3 \times 2 \times 3 \times 3 \times 6 \times 2 \times 2 = 7776$). Therefore, we decided to use a fractional factorial design. We used the orthogonal design option in SPSS version 7.5 for windows to elicit the smallest subset of possible cases while simultaneously retaining the orthogonality of the cues. This yielded a set of 27 cases (the modeling set) and created a main effects design. In similar circumstances other studies have taken advantage of techniques that reduce the number of cases selected without comprising the orthogonality of the cues (e.g. Brown and Allgeier, 1996). In addition, the program was used to select a further seven cases (the holdout set) that would be used to validate the models fitted on the modeling set. Finally, seven cases were randomly selected from the modeling set and duplicated (although the names of the defendants were changed) to measure test–retest consistency. Altogether that makes 41 cases—about the maximum we had established respondents would be prepared to complete.

The cue values were equally distributed among the set of 27 cases for all cues except for the dichotomous cues gender, age, prosecution case and community ties. Here, the unequal distributions reflected the real-world distributions. For example, there were more males than females in the set of 27 cases.

The cues were placed in the order as shown in Exhibit 1 and were in the same order in each case for ease of reading. Despite the lack of procedural rules concerning the presentation of information, our courtroom observations suggested that there was a general order for most of the cues that we were presenting. For example, as cases are not always heard in the order they are presented on the courtsheet, magistrates generally see the defendant’s gender and race upon his or her entrance into the courtroom when a case is called before they learn of the offence he or she is charged with which is on the courtsheet. Although the cue inter-correlations may not be realistic (Brunswik, 1956), the construct and face validity of the task were assessed by two magistrates. They compared the hypothetical cases to those presented in court and concluded that they were plausible. See the Appendix for an example of a hypothetical case used in the study.

The cases were presented in a random order to eliminate any order effects due to unfamiliarity with the task and shifts in attention. The holdout cases were randomly intermixed with the cases in the modeling set. The duplicate cases were placed at the end to restrict the probability of two identical cases being presented one after another.

Participants were instructed to respond to the hypothetical cases by first choosing a bail decision. The bail decision options were unconditional bail, conditional bail (with the conditions specified) or remand in custody. (The results of the conditions imposed will not be reported here.) Participants were then asked to indicate how certain they were that they had made the appropriate decision based upon the information provided on an 11-point scale. Zero on the scale represented ‘absolutely uncertain’ and 10 represented ‘absolutely certain’. These post-decisional confidence ratings were requested for each decision individually rather than for the set of decisions overall because cases differed in terms of the information provided (i.e. on the cue values).

Ranking task
In order to capture magistrates’ explicit bail decision making policies a direct ranking task was used. Cook and Stewart (1975) found no difference in the results of direct tasks. Participants were asked to rank order the nine cues according to the relative importance they attached to them when making bail decisions on the hypothetical cases presented in the bail decision making task. A rank order of one indicated the most important cue.
Procedure
The bail decision making task was followed by the ranking task in a booklet format. The booklet also contained instructions that listed the cues and described the two tasks. Participants were instructed to complete the tasks individually, to complete the bail decision making task first, not to spend too much time on each case and not to return to cases which had already been completed. Participants were also asked to specify what further information, if any, they would have liked in order to make their bail decisions in the bail decision making task. Participants’ demographic details, namely type of magistrate, location of court and number of years’ experience on the bench were also requested. We decided to measure magistrates’ overall experience on the bench, rather than the extent of their experience in making bail decisions in particular, because we felt that their sporadic work pattern may prevent accurate recall.

A handful of booklets were sent to each court, addressed to the court manager who was informed of the study and asked to distribute them to magistrates sitting in the courthouse. Magistrates’ names and addresses are not made public and so we were advised by the Magistrates’ Association to gain access to magistrates in individual courts via their managers (Bracey, personal communication, 1997). A covering letter was included for magistrates, which introduced the study, guaranteed respondent’s anonymity and asked for volunteers to participate in the study. The three weeks deadline for returning the completed booklets directly to the researchers was highlighted and a stamped, self-addressed envelope was provided.

ANALYSIS AND RESULTS

The magistrates made from 2 to 23 unconditional bail decisions (mean = 12.07), 2 to 25 conditional bail decisions (mean = 12.09) and 0 to 9 remand in custody decisions (mean = 2.84), on the cases in the modeling set.

Intra-magistrate consistency
Each magistrate’s consistency in making bail decisions was measured by computing a Cohen’s Kappa value which corrects for chance. The decisions made on the set of seven duplicate cases were compared with those made on their original counterparts in the modeling set. The Kappa value ranges from 0 (indicating that agreement or consistency is no better than chance) to 1 (indicating perfect agreement or consistency). Fleiss (1981) suggests that a value of 0.40 to 0.60 is ‘fair’, 0.60 to 0.75 is ‘good’ and a value above 0.75 is ‘excellent’. For the whole sample, Kappa ranged from 0 to 1 (mean = 0.69). The consistency of lay and stipendiary magistrates was also compared. For the 70 lay magistrates only, Kappa ranged from 0 to 1 (mean = 0.69). For the 9 stipendiary magistrates, Kappa ranged from 0.09 to 1 (mean = 0.74).

In order to assess the relationship between magistrates’ consistency and their experience on the bench, a Pearson’s correlation was computed between these two variables, across the sample. A nonsignificant correlation of 0.11 was found (one-tailed p > 0.05, n = 79).1

Agreement between magistrates
In order to measure the extent of agreement (as defined by whether different magistrates make the same decision on the same case), the percentage of magistrates who disagreed with the modal response on each

1N refers to the number of correlations computed and n to the size of the sample on which a correlation was computed.
case was calculated. Researchers using judgment analysis traditionally measure agreement in terms of the size of the correlation between two individuals’ judgments made over a set of cases (Cooksey, 1996). The measure used here is less cumbersome and is sensitive to the extent of disagreement shown on a particular case. Others have similarly used a measure of disagreement based on consensus (e.g. Strauss et al., 1995). There was disagreement between magistrates as to the decision to be made on each of the 27 cases in the modeling set. This figure ranged from 4.90% to 50% of magistrates (mean = 31.40%). Magistrates disagreed with the modal response on from 4 to 25 cases (mean = 14.95). Lay magistrates disagreed from the modal response on from 4 to 25 cases (mean = 14.96) and stipendiary magistrates disagreed on from 7 to 22 cases (mean = 14.89).

In order to examine the relationship between the number of cases on which magistrates disagreed from the modal response and their experience on the bench, a Pearson’s correlation was computed between these two variables. A significant correlation of −0.31 was found (two-tailed \( p < 0.05, n = 79 \)). A Spearman’s rank correlation was computed to examine the relationship between the extent of disagreement and the modal bail decision made over the cases in the modeling set. A nonsignificant correlation of 0.31 was found (two-tailed \( p > 0.05, n = 27 \)). There was no significant correlation between the number of cases magistrates disagreed with the modal response and their consistency in decisions (\( r = −0.08, \) two-tailed \( p > 0.05, n = 81 \)).

**Magistrates’ post-decisional confidence**
For each magistrate, a Spearman’s rank correlation was computed to examine the relationship between the bail decisions made on the cases in the modeling set and the post-decisional confidence ratings provided for these decisions. For the whole sample, the correlations ranged from −0.72 to 0.57 (mean = −0.25). The correlation was significant for 29 magistrates (two-tailed \( p < 0.05, n = 27, N = 75 \)). For the lay magistrates only, the correlations ranged from −0.63 to 0.57 (mean = −0.23. Twenty-five correlations were significant) and for the stipendiary magistrates the correlations ranged from −0.72 to −0.23 (mean = −0.44. Four correlations were significant).

Mean post-decisional confidence ratings in bail decisions made over the cases in the modeling set were then calculated for each magistrate. For the whole sample these ranged from 6.22 to 10 (mean = 8.31). For the lay magistrates only, mean post-decisional confidence ratings ranged from 6.22 to 10 (mean = 8.27). The stipendiary magistrates’ mean post-decisional confidence ranged from 7.22 to 10 (mean = 8.87). In order to examine the relationship between magistrates’ mean post-decisional confidence and their consistency in decisions, a Pearson’s correlation was computed between these two variables, across the sample. No correlation was found (\( r = 0.00, \) two-tailed \( p > 0.05, n = 81 \)). There was also no significant correlation between magistrates’ mean post-decisional confidence and their experience on the bench (\( r = 0.10, \) two-tailed \( p > 0.05, n = 79 \)).

**Magistrates’ bail decision making policies**
The polytomous cues were dichotomized for ease of analysis and for each cue all non-italicized values were coded as 0 and italicized values were coded as 1 (see notes to Exhibit 1). The dichotomization was based on semi-structured individual interviews with six lay magistrates, one stipendiary magistrate and two clerks to the justices, and so should reflect how magistrates may simplify the information presented in the courtroom. The cue inter-correlations remained zero. The three bail decision options were also simplified into a binary decision, where unconditional bail represented a non-punitive decision and conditional bail or remand in custody represented a punitive decision. Analysis of the frequency of the bail decisions made by each magistrate on the modeling set revealed that magistrates made a relatively equal number of punitive and non-punitive bail decisions.

The bail decision making policy of each magistrate was modeled on the set of 27 cases. We aimed to compare a noncompensatory fast and frugal model called the Matching Heuristic (developed by Dhami
and Ayton, 1998 for binary categorization tasks) with a compensatory integration model, in order to choose which of these models best describes and predicts magistrates’ bail decisions. Although we would have liked to use a regression model such as logistic regression, this was not possible because of the low case-to-cue ratio (i.e. 3:1). We therefore decided to use two of the other models used by Gigerenzer and his colleagues, namely a differentially weighted integration model called Franklin’s rule (also known as Prudential Algebra) and a unit weighted integration model called Dawes’ rule. Both these models provide characterizations of judgment behavior similar to that provided by a regression model. Moreover, although both models do not weight the cues optimally in the way the least squares method does in a regression model, for example, studies have demonstrated that both models are excellent approximations to regression models in terms of descriptive and predictive validity (Dawes and Corrigan, 1974; Cattin, 1978; Dorans and Drasgow, 1978; Einhorn and Hogarth, 1975; Gigerenzer et al., 1998; 1999; Gigerenzer and Goldstein, 1996; Schmidt, 1971; Wainer, 1976). Dhami and Harries (in press) found that there was no statistically significant difference between the fit of the Matching Heuristic and a logistic regression model on doctors’ prescription decisions. They defined fit in terms of the percentage of decisions correctly predicted by the models. We define fit similarly. In this sense, we do not consider the present study to be especially limited due to the inability to compare the Matching Heuristic with a regression model.

In the sections below we will first describe the procedure for modeling magistrates’ bail decision making policies with each of the three models (Franklin’s rule, Dawes’ rule and the Matching Heuristic). All the models were developed so that they aimed to predict a punitive decision and only predicted a nonpunitive decision by default because this is the procedure followed by the law on bail. We then present the results of the fit of the models on the modeling set and the models’ ability to predict magistrates’ decisions on the seven cases in the holdout set. The model that performed best on average across magistrates was accepted as the description of magistrates’ bail decision making policies.

Franklin’s rule

In this model each cue was weighted according to its influence on the decision. Then, for each case this model multiplied the cue values by their weights and then summed them. (Where a case is made up of binary cues, the cues can be coded 0 and 1 and so the sum is the sum of the weights alone for all cues taking a value of 1 in the case.) If the sum was equal or greater than the threshold value then a punitive decision was predicted. If not, then a nonpunitive decision was predicted.

The cue weights were calculated by choosing the value on each cue that had the greatest proportion treated punitively in the set of 27 cases. For example, if the proportion of males treated punitively was greater than the proportion of females treated punitively, then the former proportion would have been the weight for the gender cue. Cue weights could alternatively have been calculated using methods such as the likelihood ratio, phi coefficient and chi square. These methods take into account all the information (i.e. number of males treated punitively or nonpunitively and number of females treated punitively or nonpunitively). The method used in the present study also does this because column totals for each cue are the same for all cues as each magistrate made a specific number of punitive and nonpunitive decisions. The threshold value was calculated by first taking the sum of the cue weights for each of the 27 cases, then totaling these 27 sums, and then dividing the total by the number of cases (i.e. 27). This is a reasonable method for calculating the threshold value because each magistrate made roughly an equal number of punitive and nonpunitive decisions.

Gigerenzer et al. (1999) call these models Franklin’s rule and Dawes’ rule respectively, because the American statesman Benjamin Franklin described the procedure for the former model and Robyn Dawes, a judgment and decision-making researcher, showed that equal-weighted models provide good approximations to differential weighted models.

If we tried to optimize the weights through various techniques and recalculated the goodness of fit the model would overfit the data, and thus do worse at generalizing to the holdout set.
To provide a concrete example, we use magistrate-1’s bail decision making policy as described by Franklin’s rule to predict this magistrates’ decision on case three in the modeling set. In this case the defendant was male, an ethnic minority, 18–21 years old, charged with a triable-either way/indictable offence, the prosecution requested a punitive decision, he had no previous convictions and no good bail record, the prosecution case was weak, he had weak community ties and the police made a punitive bail decision. The weights attached to the cues were as follows: gender(0)(0.72) + race(1)(0.67) + age(0)(0.67) + offence(1)(0.78) + proseq(1)(0.72) + procr(0)(0.73) + procase(0)(0.78) + commsies(1)(0.67) + polbail(1)(0.67). The sum is 3.51, which is less than the 3.52 threshold value calculated for this magistrate. Thus, in this case, Franklin’s rule would incorrectly predict that magistrate-1 made a nonpunitive decision.

Dawes’ rule
In this model cues were unit weighted. For each case this model counted how many cues had the critical cue value, and if the unit sum of these was greater than or equal to the threshold value, it predicted a punitive decision. If not, it predicted a nonpunitive decision.

The critical cue value of a cue was defined as the value with the greatest proportion that had been treated punitively in the set of 27 cases. For example, if the proportion of males in the set of 27 cases treated punitively was greater than the proportion females treated punitively, then the critical cue value for the gender cue would have been male. The critical cue value was then given a weight of 1, while the other value was given a weight of 0. The threshold was determined as in Franklin’s rule.

As an example, on case three, magistrate-1’s Dawes’ rule sums to 5, which is less than the 5.89 threshold value calculated for this magistrate. Thus, Dawes’ rule would also incorrectly predict that magistrate-1 made a nonpunitive decision.

Matching Heuristic
This model searched through $K$ of the available cues in rank order of importance, looking for a critical value on each cue that indicated a punitive decision. If a critical value was found the model stopped searching and predicted a punitive decision. Otherwise, the model searched through the value of the next rank ordered cue. The model continued this procedure until $K$ cues had been searched. If by this time no critical value had been found the model predicted a nonpunitive decision. For illustrative purposes, Exhibit 2 shows the bail decision making process by a Matching Heuristic model where $K=2$.

The following three steps describe how the critical values on the cues, the rank ordering of cues and $K$ are determined for each magistrate. First, a critical value was defined as the value on a cue that was most frequently treated punitively in the set of 27 cases. For magistrate-1, for example, the critical value for the gender cue was male because this magistrate made a punitive decision on more male than female defendants. Therefore, despite no explicit inclusion of base rates, the critical value was affected by the fact that there were actually more males than females in the modeling set. If the absolute frequencies of the number of cases treated punitively were equal among the values of a cue, then the value with the lowest absolute frequency treated nonpunitively was chosen. The lowest absolute frequency in this situation would result in choosing the cue value with the highest ratio of punitive to nonpunitive decisions. Where the absolute frequencies of the number of cases treated nonpunitively were also equal, a critical cue value was chosen randomly. Second, a cue-utilization validity was calculated for each cue and was defined as the proportion of cases with the critical value that were treated punitively in the modeling set. For magistrate-1, for example, the validity of the gender cue was defined as the proportion of male defendants who were treated punitively. The validities were then used to rank order the nine cues, where the first rank was assigned to the largest validity. This rank order indicated the order in which the model searched through the cues. Cases with a tied rank order were placed in the order they were presented in the judgment task. Finally, in order to choose the maximum
number of cues for the model to search ($K$), the fit of the model with all nine possible maxima was systematically tested on the modeling set. The model with the best fit in terms of percentage of correct predictions was chosen as the model of magistrates’ bail decision making policy and where two or more models had the same fit, the more parsimonious model (i.e. the model employing the fewest cues) was chosen. Parsimony is the hallmark of the fast and frugal approach, which emphasizes simplicity. As Exhibit 3 illustrates, a model that looked at only one cue was the best fit for magistrate-1. The Matching Heuristic stated that this magistrate only used the offence cue to make a decision; where a serious offence (triable-either way or indictable) predicted a punitive decision and a summary offence predicted a nonpunitive decision. Thus, for case three, magistrate-1’s Matching Heuristic would only search for information regarding the offence and would correctly predict a punitive decision.
A Matching Heuristic which searched through one cue only (out of a possible nine cues) \((K = 1)\) proved to be the best fit for 75.30% of the sample. \(K = 2\) for 21% of magistrates and \(K = 3\) for the remaining 3.70% of magistrates.

**Describing and predicting magistrates’ bail decisions**
Each of the three models was used to make a prediction first on the set of 27 cases used to develop the models and then on the set of seven holdout cases used to validate the models. As stated earlier, we refer to the percentage of decisions correctly predicted by the models as ‘fit’. The binary nature of the decision to be described and predicted implies that any valid model should be expected to perform better than chance (i.e. predict more than 50% of decisions). Dawes’ rule provided the best fit on the modeling set for 9.90% of magistrates, Franklin’s rule proved to be the best fit for 38.30%, and the Matching Heuristic proved to be the best fit for 32.10% of the sample. For the remainder of the sample either two or all of the three models provided an equal fit.

The superiority of Franklin’s rule, however, diminishes when looking at the fit of the models on the holdout set. Here, Franklin’s rule was better than the other two models at correctly predicting the decisions of 14.80% of magistrates, while Dawes’ rule was the best for 19.80% of magistrates, and the Matching Heuristic was the best for 33.30% of magistrates. Once again, for the remainder of the sample either two or all of the three models were equally good at correctly predicting magistrates’ decisions on the holdout set.

Exhibit 4 presents the results of the average fit of the models across the sample. It can be seen that the Matching Heuristic provides a better fit on average across magistrates (mean = 73.98%) than either...
Franklin’s rule (mean = 73.57%) or Dawes’ rule (mean = 69.36%) on the modeling set. Moreover, the Matching Heuristic proves to be better at predicting magistrates’ bail decisions (mean = 65.61%) than Franklin’s rule (mean = 59.26%) and Dawes’ rule (mean = 62.96%), on the holdout set. Although Franklin’s rule does better than Dawes’ rule at describing the data in the modeling set, it does worse at predicting the decisions in the holdout set. Thus, we concluded that the Matching Heuristic model best captured magistrates’ bail decision making policies.

Cues usage
The Matching Heuristic was used to elicit magistrates’ cue use when making bail decisions. Cue use is defined broadly as the number of cues searched (including the cue on which the decision is based) and so this number may vary from case to case where \( K > 1 \). The mean number of cues used over the cases in the modeling set was calculated for each magistrate. Across magistrates the mean number of cues used ranged from 1 to 1.67 (mean = 1.10).

When comparing the Matching Heuristic across magistrates, it was found that magistrates differed in terms of the cues they used to make their bail decisions. Previous convictions and bail record was used by 23.46% of magistrates, followed by offence, which was used by 22.22%. Prosecution request was used by 19.75%, gender was used by 17.28%, police bail decision was used by 16.05% and strength of community ties was used by 12.35% of magistrates. Finally, race, age and strength of prosecution case were each used by 6.17% of the sample, respectively.

Analysis of the critical value on the cues used revealed that for seven of the nine cues, those magistrates using the cues, used them in the same direction. Here, with the exception of the strength of prosecution case cue, the legal cues were used in the direction expected, and the ‘extra-legal’ cues were used in the direction reported by previous research. Four of the five magistrates who used the race cue made a punitive decision when the race of the defendant was described as white. Seven of the 10 magistrates who used the strength of community ties cue made a punitive decision when the defendant was described as having weak ties.
Requests for further information
A total of 144 requests were made by magistrates for more information in response to the question of what further information they would have liked in the set of hypothetical cases which would help them to make their bail decisions. Fifty-seven percent of the requests were for further information regarding the nature of the current offence such as time of the offence. Nineteen percent of requests were for more information regarding the current bail position such as availability of bail hostel places, 9% were for more information regarding the defendant’s community ties and 7% were for more information on defendant related factors such as the defendant’s ‘lifestyle’. The remaining requests were for more information regarding the defendant’s bail record and for court related factors such as the clerk’s advice.

Comparison between Matching Heuristic and magistrates’ explicit statements of policy
Although according to the Matching Heuristic none of the magistrates used all the cues, we were interested in assessing magistrates’ explicit consideration of cue importance for each of the nine cues. Therefore, the rank ordering of the nine cues as described by the Matching Heuristic were compared with the rank ordering of cues explicitly provided by magistrates in the ranking task. Kendall’s tau-b correlations were computed between each magistrates’ implicit and explicit rank order of cues. The correlations ranged from −0.39 to 0.67 (mean = 0.09). There were only two significant correlations (2-tailed p < 0.05, n = 9, N = 81). For the lay magistrates only, these correlations ranged from −0.39 to 0.56 (mean = 0.08), and for the stipendiary magistrates these correlations ranged from −0.22 to 0.67 (mean = 0.13).

In order to examine the relation between the degree of concordance between magistrates’ implicit and explicit policies and their experience on the bench, a Pearson’s correlation was computed between concordance and experience. A nonsignificant correlation of 0.10 was found (two-tailed p > 0.05, n = 78). There was also no significant correlation between concordance and consistency in decisions, across the sample (r = 0.06, two-tailed p > 0.05, n = 80).

For an indication of the cues that were ranked differently in magistrates’ implicit and explicit policies, we computed the mean rank order of each cue in both types of policies, across magistrates. Exhibit 5 illustrates the concordance between magistrates’ implicit and explicit rank ordering of cues summarized across all magistrates. Note that the rank order of importance was reversed for ease of illustration, so a rank order of nine represents the most important cue.

Exhibit 5. Comparison between magistrates’ implicit and explicit bail decision-making policies
DISCUSSION

Bailing and jailing the fast and frugal way

We found that the majority of magistrates showed some inconsistency in their bail decisions. Magistrates also disagreed as to the decisions to be made on the same cases. Nevertheless, all magistrates were highly confident that they had made the appropriate decisions. Magistrates’ bail decision making policies were better described and predicted by a fast and frugal model called the Matching Heuristic characterized by non-compensatory processing of information than by either of two compensatory integration models. According to the Matching Heuristic, magistrates based their bail decisions on one cue and they differed as to the information they used. Some based their decisions on defendant and crime control-related cues. Finally, magistrates’ explicit statements of cue use did not match that proposed by the models of their decision making policies.

These findings are based on individual magistrates’ decisions and the use of hypothetical cases. Magistrates’ bail decision making training packages comprise hypothetical cases similar to those used in the present study. The external validity of hypothetical cases has been questioned (Ebbesen and Konecni, 1980). However, after reviewing 30 years of policy-capturing research, Brehmer and Brehmer (1988) concluded ‘thus far, then, it seems that the paper format as such does not lead to any important distortions in the policies obtained’ (p. 89). Cooksey (1996) agreed with this conclusion after considering the internal and external validity of studies using hypothetical cases. Both these reviews included research using paper cases that, like those in the present study, involved hypothetical cases with orthogonal cues. We do not consider the present study to be especially limited by the fact that we investigated decisions made by magistrates on hypothetical cases consisting of an orthogonal cue set. Indeed, previous research suggests that, though people may claim to use configural cues, incorporating such terms into models of judgment policies adds little, if anything, to their descriptive validity (see Brehmer and Brehmer, 1988). In our interviews with magistrates during the design phase of the present study, they did not report anything to suggest configural cue use.

By examining the influence of each variable upon magistrates’ bail decisions independently of the effects of other variables, we have confirmed criminologists’ claims that magistrates are influenced by defendant and crime control-related cues. However, only a small minority of magistrates used the race cue and they mostly used it in the opposite direction to that claimed by criminologists. One possible explanation is that magistrates were sensitive to our research aims and so consciously avoided using this cue. A letter received from the chairman of a bench of magistrates during the design stage of the study stated ‘it worries me that you are bringing race into it. What does the color of a defendant’s skin have to do with ... the bail decision ... ? I do hope you are not trying to prove that whites are given bail more frequently.’ Although magistrates were provided with information regarding a defendant’s community ties, only a handful of magistrates used this cue, and when they did use it, the majority used it in the direction expected. This implies that Bail Information Schemes providing this type of information to magistrates may not be as effective as has been indicated in the past by observational studies for example. Future studies employing an experimental approach may yield more reliable and valid findings on this issue. The use of cues such as gender, age, race, prosecution request and police bail decision, together with the requests for further information, reveal how magistrates interpret the catch-all clause contained in the law on bail.

Perhaps not surprisingly, factors which are considered to be both socially and professionally undesirable for making a bail decision such as defendant and crime control-related cues were ranked lower in magistrates’ explicit statements of policies than according to the models describing their behavior. Statements

---

4 At first sight, the findings of our study may seem discrepant with those of Morgan (1994) and Huckleby (1996), for example. These criminologists’ studies involve magistrates from a relatively small sample of courts in England and Wales, and their results are consistent with the results we report for a small sample of magistrates from our much larger sample of courts.
made by magistrates and court managers during the data collection phase of the study supports our findings. For example, a clerk to the justices stated that ‘the physical appearance and the presence of the defendant . . . do make a difference’. In the present study we distinguished between what magistrates publicly state they do and what they actually do. Konecni and Ebbesen (1984) suggest that these should also be distinguished from what legal decision makers privately think they do, which refers to the concept of self-insight. Future research could investigate magistrates’ insight into their policies by using policy recognition methods like those advocated by Reilly and Doherty (1992).

It was also confirmed that different magistrates disagree as to the decisions to be made on the same cases. The extent of disagreement may be partly explained by the differences in the cues used and magistrates’ individual inconsistencies. The present study extended our understanding of magistrates’ bail decision-making to include a measure of their individual consistency in their decisions and their post-decisional confidence. The high level of post-decisional confidence demonstrated by magistrates in the present study is arguably inappropriate where magistrates were in disagreement with one another and where they were inconsistent. Despite the absence of an outcome criterion, we can assert that the extent of inconsistency and disagreement found indicates that magistrates are likely to make errors, for example releasing a defendant who would abscond. In the same way that reliability is a pre-requisite for the validity of a psychometric test, consistency (whether it be within an individual or between individuals as defined by the measure of disagreement in the present study) is a pre-requisite for accuracy with respect to an individual’s decisions. The extent of inconsistency and disagreement between magistrates in the present study is surprising considering that they were performing a structured judgment task and were presented with the same information, but unsurprising in the light of other research. It is likely that there will be greater inconsistency and disagreement in the courtroom, where there are no rules of procedure, a lack of information, time pressure and an erratic work pattern.

Although Hucklesby (1997) reported that stipendiary magistrates are more likely than lay magistrates to disagree with the prosecution request and to remand in custody, our findings confirm evidence (Doherty and East, 1985; King, 1971) that suggests there are no significant differences between the two types of magistrates. Observed differences may be due to the fact that stipendiary magistrates deal with more serious and complicated cases. Intra-individual inconsistencies and characteristics of the bail decision making task may explain the lack of significant differences in performance between the two types of magistrates and between more and less experienced magistrates.

The present study differed from other studies employing judgment analysis because we tested the use of alternative models as descriptions of judgment behavior. We found magistrates’ bail decision making policies could be better represented by a fast and frugal model. Others have shown that people often use heuristics (Tversky and Kahneman, 1974). Although we studied lay magistrates as individuals rather than as a bench, empirical evidence has revealed that people use less cognitive effort, as indicated, for example, by the number of cues they use, when under conditions of shared responsibility (Weldon and Gargano, 1985). Thus, it is likely that the bench may also be fast and frugal.

The fast and frugal model does not search through all available information, does not weight information in an optimal way, does not integrate all relevant information and bases its decision on only one cue. In this sense, although the fast and frugal model is descriptively valid, it lacks prescriptive utility, because magistrates behaving in a fast and frugal manner are not observing due process requirements as they are currently defined. The law on bail is silent regarding exactly what information magistrates should use and how they should weight and integrate it. The due process model would require that magistrates do not use被告 and crime control-related cues and they carefully search through all the available information and appropriately weight and then integrate the relevant information. This way of making a decision characterizes how compensatory models such as regression models portray the judgment process. However, we have already mentioned that human cognitive limitations and certain task characteristics may prevent individuals from using such judgment strategies.
Implications for research in judgment and decision making

Gigerenzer and his colleagues have evaluated the prescriptive utility of fast and frugal models solely in terms of their accuracy (this is often global accuracy). In many applied domains, such as the legal domain, however, accuracy is not the primary concern (Hammond, 1996b). Thus, as we have argued, although fast and frugal models are descriptively and predictively valid, they may lack prescriptive utility.

In studying simulated judgments (or accuracy in predicting an outcome), the efficacy of fast and frugal models has been attributed to their ability to exploit the redundancy stemming from cue inter-correlations in the environment (Gigerenzer et al., 1999). In the present study, we examined a fast and frugal model’s ability to describe and predict human judgments made on a set of hypothetical cases comprising orthogonal cues. We found evidence of noncompensatory, fast and frugal judgment strategies.

Previous research has shown that people use noncompensatory strategies under conditions such as time pressure. The findings of the present study are not limited to such conditions. Furthermore, although previous judgment analysis research has shown that people use few cues, we have found no study that has reported participants using only one cue, as we have found in the present study.

The cognitively simple strategy embodied in the Matching Heuristic is more psychologically plausible than that characterized by the two static compensatory models. First, people are flexible in their use of information. The Matching Heuristic, like other fast and frugal models, indicates that the same cue is not used to make a decision on every case, as it is in static models. Second, as we mentioned earlier, past research suggests that people may switch to noncompensatory strategies under conditions such as time pressure. Thus, magistrates, and other professionals, may well be using simple strategies because they often work under such conditions. Third, although Gigerenzer and Goldstein (1996) assumed an optimal learning strategy is involved in the development of their fast and frugal model, the Matching Heuristic does not. It is more compatible than earlier fast and frugal models and the two static models presented here with evidence that people are selective and use frequencies when they learn relations between cues and an outcome. The Matching Heuristic uses frequencies when determining the critical cue value. It is claimed that this is a natural form of processing (Cosmides and Tooby, 1996; Gigerenzer and Hoffrage, 1995). The way in which the Matching Heuristic defines the critical cue values and the cue-utilization validities is compatible with research showing how humans learn about and judge causation and covariation from direct experience (Nisbett and Ross, 1980). Evidence suggests that when humans learn about the relations between cues and an outcome they look at only a subsection of the available information. For example, in the modeling set, the number of male and female cases who were treated punitively and non-punitively can be put into a 2 × 2 table, and when determining the critical value, the model looks at only the two punitive cells. When determining the validity, the model only looks at the two cells of the value chosen as the critical value. Finally, the Matching Heuristic embodies the principle of matching characteristics of individual cases with a prototype. This is consistent with evidence from the field of categorization (see Estes, 1994). From the above it is clear that if researchers wish to develop psychologically plausible decision mechanisms, they should attempt to integrate evidence from the broader context of psychology.

Despite the criticism that regression models are not psychologically plausible descriptions of judgment behavior, the vast majority of judgment analysis research uses regression models. Researchers continue to paint a complex picture of human judgment and they seem to have overlooked an indication that individuals may be using simple strategies from the regression models themselves. This is the finding that usually only a handful of cues (three on average according to Brehmer (1994)) are statistically significant (see Brehmer and Brehmer, 1988; Libby and Lewis, 1982; Slovic and Lichtenstein, 1971). A number of explanations for the seemingly ‘ritualistic’ use of regression models have been offered, including that they provide an adequate fit to data (Brehmer and Brehmer, 1988) and the tools to apply them are readily available (Stewart, 1988). Although alternative models have been proposed, we have found no instance of researchers using judgment analysis to investigate legal decision making or judgment in any other domain for that matter who have compared the relative descriptive validity of different types of models. This recently led Doherty and Brehmer
(1997) to conclude that ‘the conception of human judgment emerging from regression studies of human judgment can be refuted only by evidence that shows that judgment is something other than a matter of combining pieces of information that are weighted according to their importance. So far, such evidence has failed to materialize’ (p. 547). We feel the present study has provided some evidence to this effect.

Moreover, as we have demonstrated, judgment analysis is not synonymous with the use of regression models. Researchers interested in legal decision making and judgment in other domains should reconsider their use of these models. Perhaps we are finally witnessing a turning of the tide. For instance, recently, Kenneth Hammond (1996b), a prominent judgment analyst who during his career has consistently used regression models to investigate professionals’ judgment policies, confessed that a ‘sin of commission on my part was to overemphasize the role of the multiple regression (MR) technique as a model for organizing information from multiple fallible indicators into a judgment’ (p. 244). Finally, in order to avoid a similar situation regarding the future use of fast and frugal models, research should compare the relative descriptive and predictive validity of these models and other noncompensatory models.

**Implications for the criminal justice system**

As in previous research on legal decision making, the present study has shown that the current practice of magistrates’ bail decision making is far from ideal practice. Our finding contradicts how magistrates and their managers believe the bail decision is made. For example, a lay magistrate wrote to us stating that ‘the situation . . . depends on an enormous weight of balancing information, together with our experience and training.’ The chairman of the council stated ‘we are trained to question, and to assess carefully the evidence we are given.’ In Konecni and Ebbesen’s (1984) terms, these quotes highlight the ‘mythology of legal decision making.’ In the past, criminologists criticizing magistrates’ bail decision making have tended to focus solely upon fallibilities in magistrates to the exclusion of an analysis of the bail decision making task that magistrates are expected to perform and the conditions under which they perform this task. Performance is contingent on both human cognitive limitations and task constraints. Future research should examine the effects of task conditions such as time pressure, order of information presentation and work pattern upon magistrates’ bail decisions. Although we did not study the effects of characteristics of the bail decision-making task upon magistrates’ performance, previous research in other domains suggests that such characteristics will have predictable effects.

Therefore, interventions should focus on both the characteristics of the bail decision making task and on magistrates themselves. As an example, we list a few of the possible interventions for change. The Bail Act 1976 could be better specified. The catch-all category could be unpacked into its component parts and the Act could incorporate a weighting scheme for the information that is informed by at least some objective measures of the predictive validities of the factors. Thus, research needs to be directed at measuring cue validities. (If statistical data could be collected and a decision rule could be formulated, it is conceivable that magistrates could be replaced by an automated system.) Magistrates’ cognitive processing limitations could be overcome by providing them with a cognitive aid that helps them implement the law. Cognitive aids have been recommended in other judgment domains such as psychiatry (Erdman, 1988), and are not a new concept in the legal domain (Larsen et al., 1997). Statutory rules of procedure for bail hearings could be introduced and time pressure could be reduced. Magistrates’ consistency and disagreement could be improved by providing them with cognitive feedback of their own and other magistrates’ bail decision making policies, respectively. Cognitive feedback has been shown to be effective in other domains such as medicine for example (Kirwan et al., 1983). Finally, although the accuracy of magistrates’ bail decision making was not directly measured in the present study, it is suggested that formal, objective outcome feedback could help magistrates to develop an appropriate policy. These suggestions may be relevant to other legal decisions and other jurisdictions where decision makers are afforded considerable discretion and work under similar task conditions.
Any attempt to improve performance needs to take into consideration the idea that decision makers may be reluctant to change. In the present study, magistrates’ resistance to change may be indicated by their high degree of post-decisional confidence. Nevertheless, an awareness of the magnitude and significance of legal decisions such as magistrates’ bail decisions may highlight the urgency with which practice needs to be improved.

Recognizing the importance of legal decisions

The magistrates’ bail decision has significant consequences. For example, evidence suggests that defendants who are remanded in custody in England and Wales are more likely to lose their homes, jobs and suffer loss of reputation and deterioration in family ties (see Cavadino and Gibson, 1993; King, 1971). They are also more likely to plead guilty, be given a custodial sentence and less likely to be acquitted than their bailed counterparts (e.g. Davies, 1971; Jones, 1985). This can be partly explained by the implicit presumption of guilt inherent in a remand in custody, the difficulties in preparing a defence while in custody and the long duration spent in custody awaiting trial. Remand prisoners place a financial burden on the prison service (White, 1999) and remand prisons are notoriously overcrowded and lack regime activities (Morgan and Jones, 1992). It is not surprising that this circumstance, coupled with the uncertainty of knowing what the future holds, makes it more likely for remand prisoners to kill themselves whilst in prison. In 1999, 64% (58) of the self-inflicted deaths amongst the English prison population were by remand prisoners (figures provided by Suicide Awareness Support Unit, HM Prison Service, 1999). It has been argued that attaching conditions to bail such as reporting to the police station, residing at a specific address or keeping curfew do not prevent offending, interference with witnesses or absconding (e.g. Block, 1990), but they do curtail a defendant’s liberty. On the other hand, if too many of the ‘wrong’ defendants were released, the general public would be the victims of offending on bail. A recent study by Brown (1998) found that 15% of those released on bail (unconditionally or conditionally) in his sample committed offences and 9% failed to attend court at least once.

It is clear that individuals who are trained and entrusted to apply legal rules and procedures must perform a tremendous balancing act. They must protect the public while simultaneously respecting an individual defendant’s right to liberty. The lack of due process observed in magistrates’ bail decision making raises doubts about their performance on more complicated and less frequently made decisions such as the decision to convict. Fortunately, our conceptions of crime and order are socially constructed and so the concerns expressed by different groups in society regarding the appropriateness of some of the legal decisions made have often led to specific changes in legal rules and procedures. Unfortunately, judgment and decision making researchers, despite being equipped with the necessary tools, have tended to neglect legal decision making, especially in the English criminal justice system. We propose that they, like criminologists, should take the opportunity to challenge and change the system.

APPENDIX

Below is a copy of a hypothetical case used in the present study:

The defendant Ravi Kumar is male, Asian, and 23 years old. He has been charged with inflicting grievous bodily harm. The police bailed him with sureties after he was charged. The prosecution has not objected to unconditional bail. Ravi has one previous conviction for a similar offence, and has offended while on bail in the past. The prosecution case against him at present does not seem particularly strong. He has a fixed address.

(please tick one of the options)
Based upon the information provided, what is your decision?
unconditional bail [ ]
conditional bail [ ] please specify condition
remand in custody [ ]

(please circle one of the options)
On a scale of 0 to 10 how certain are you that you have made the appropriate decision?
absolutely uncertain 0 1 2 3 4 5 6 7 8 9 10 absolutely certain

ACKNOWLEDGEMENTS

The first author was awarded the 1999 De Finetti Prize by the European Association for Decision Making, for the research presented in this paper. We would like to thank Torsten Mohrbach for programming Franklin’s rule, Dawes’ rule and the Matching Heuristic.

REFERENCES


Authors biographies:

**Mandeep K. Dhani** is a PhD student at City University, London. She was a pre-doctoral fellow at the Max Planck Center for Human Development where she completed part of this work. Her research interests include legal decision making, Brunswikian psychology and prison psychology.

**Peter Ayton** is a Reader in Psychology at City University, London. His research interests include subjective probability, risk, and human decision making.

Authors’ address:

**Mandeep K. Dhani** and **Peter Ayton**, Department of Psychology, City University, Northampton Square, London EC1V 0HB, UK.