

Time Series Analysis of a Murder Trial

GERHARD KETTE and HERMANN BRANDSTÄTTER

University of Linz, Austria

ABSTRACT

As observers of an entire trial by jury on attempted murder thirty-seven law students continuously rated their impression of the defendant's guilt any time they experienced a just noticeable change in guilt perception. With the transcripts of the trial recorded on tape a detailed content analysis was performed as to extralegal vs. legal evidence, abstract vs. concrete presentation of information, factual vs. emotional emphasis, compassion for defendant and attribution of responsibility (external vs. internal). Each aspect of the information units was rated by eight different subjects (four male, four female). The units of information were presented to half of them in the reverse order. It was found that legal relevance, attribution of responsibility, and abstract vs. concrete information are the best predictors of the time series of guilt perception.

Key words: Legal psychology, jury decision-making, time series analysis, impression formation, information integration.

INTRODUCTION

A model of jury decision-making ought to explain how individual jurors reach their pre-deliberation voting stances, and how those individual preferences are combined into a group decision resulting in a guilty/not guilty verdict. The pilot study to be reported here deals solely with the first stage of this highly complex decision process. The analysis of the second stage, i.e. the final, interactive decision-making of the jurors, will be presented in a separate report.

According to the Information Integration Model (Anderson and Graesser, 1976; Burnstein, 1982; Anderson, 1982), the task of the juror is to listen to conflicting testimony from a variety of sources and to integrate that information into a single decision to convict or acquit the defendant. The juror presumably arrives at a subjective likelihood that the defendant is guilty, then compares it with a threshold of reasonable doubt. If the subjective likelihood of guilt is greater than the threshold of reasonable doubt, then the decision is to convict. Testimony, observations of the defendant, preexisting biases, and other sources of information (judge's instructions, counsel's arguments, etc.) lead to beliefs or inferences about the defendant's ability, motivation and opportunity to commit the crime, and about his/her character, etc. These beliefs are combined into a single likelihood that the defendant is guilty. The specific combination rule proposed in legal applications

represents the juror's decision as a weighted average of an initial opinion and information presented during the course of the trial. The trial provides both evidence and, legally speaking, irrelevant information.

The Information Integration Model shows (compared to other models of individual juror Decision-Making, like Bayesian Models, Poisson Model, sequential weighting models or nonmodels – for a review see Pennington and Hastie, 1981) a high generality and empirical adequacy and offers a practicable measurement method and a close interdependence between theory and the methodology. We will use this model as a general framework, not for further empirical testing of its assumptions, but as a background in which our hypotheses and method are embedded.

Here the primary question is whether the information processing strategy required by the legislator really determines the process of guilt perception, that is, whether only legally relevant facts contribute to the judgements of guilt, and not legally irrelevant aspects like emotional tuning and concreteness of information.

The interface of emotions with the processing of legally relevant information forms the focus of this study with thirty-seven law students attending a murder trial and acting as mock jurors.

How informational influence in group decision-making is modified by social emotions has been studied intensively in our laboratory for some years (for reviews see Brandstätter, 1985, 1987). Depending on the topic (facts or values), on the role of the subjects (participant or observer), on the contingency of emotional utterances of the discussion partner (contingent on the subject's argument or not contingent on it), and on personality characteristics of the subject (exchange oriented vs. reinforcement oriented), the positive or negative social emotions communicated had a different effect.

It is not far-fetched to assume that social emotions also influence the process by which jurors observing a trial in the courtroom form their impression of the defendant's guilt. For whatever reason they may like or dislike his/her appearance and/or behaviour in the courtroom and, as a result, may be inclined towards a lenient or a harsh judgement. Social emotions come into play in yet another way: when the incriminated interaction between the defendant and his victim is presented before the court, or when, for example, the prosecutor, the defending counsel, or some expert witness may draw a vivid picture of moving or abominable events in one's life events in order to impress the jurors for the better or worse.

As to the function of affect in information processing, we agree with Tomkins (1980) who assumes that affect amplifies both the perceived prominence of the source of the affect and the acts following it. According to Tomkins (1980, p. 147), affect 'combines urgency and generality. It lends its power to memory, to perception, to thought, and to action no less than to the drives.' Thus, listening to an emotionally-tuned account of the defendant's misbehaviour will amplify the perceived guilt and amplify as well the impulse to judge him as guilty. Moreover, whether the accounts of the defendant's behaviour are abstract or concrete will make a difference in their impact on the jurors. Concrete information is expected to arouse more attention, to be understood better, remembered better and, therefore, to be more influential than abstract information (cf. Eagly and Chaiken, 1984, p. 277f.).

Phenomenological understanding of the trial process

Everybody has some intuitive understanding of what goes on at a murder trial and how the jurors form their private impressions of the defendant's guilt. Although such an intuitive understanding is not enough, it would be unwise to rush into constructing and testing a formal time series model of the process under consideration without first trying a kind of phenomenological analysis, unrestricted by theoretical assumptions and methodological constraints.

We have distinguished between three different, although interconnected, processes. The first is the process of interaction between the participants in the trial (judge, prosecutor, counsel for the defence, witnesses, defendant, jurors, audience). The second is the process of the formation of impressions of the defendant's guilt by observers of the trial (law students in the role of mock jurors). The third is the process of content analysis of the trial transcripts performed by a different sample of students.

The real trial interaction follows procedural rules: reading the indictment, questioning the defendant, examining the witnesses, the hearing of expert witnesses, and the prosecutor's and defence counsel's addresses to the jury. This means that the incriminating or exonerating evidence is presented variously: for a certain time the arguments are more or less supportive of the defendant's case, and then information is similarly presented stressing the defendant's guilt.

Coherence is an essential characteristic of discourse production and understanding (cf. Foss, 1988). However, besides the flow of arguments suggesting guilt or innocence and notwithstanding the planned interventions of the judge who in the Austrian (and continental European) jurisdiction strongly influences the sequence of questions and answers, unpredictable events interfere heavily with the planned process structure. Thus, we can expect the sequence of arguments for or against the defendant's guilt to be a mixture of systematic and chance components.

The second process, the impression formation in the minds of the (mock) jurors is probably less erratic, although at all stages dependent on the (real) process of interaction. We may expect coherence not only from the structure of the trial itself; the cognitive structure of the jurors, their endeavour to integrate all the divergent information coming up during the trial into a comprehensive evaluation of the entire evidence, makes the series of guilt judgements more predictable; i.e. what the jurors mark on their scales early in the process largely determines their later marks.

Each new piece of information, in itself without great significance, usually adds less and less to the existing evidence. We may be right in assuming that in most cases the jurors do not usually restructure their prior knowledge completely when confronted with new evidence, but do in some way continuously sum up or average the units of information as Anderson's (1982) information integration theory suggests.

In reviewing recent research on social cognition and social perception, Higgins and Burgh (1987) also stress the point that stored judgements heavily influence the subsequent processing of information about a person. What is not consistent with prior beliefs is often ignored or transformed, and, in reconstructing the information from memory, people rely on their prior judgements, more so in delayed than in immediate recall. On the other hand, there is some evidence that people will make more of an effort to be accurate if describing and predicting events correctly is

particularly important to them. Serving as a juror in a trial is obviously such a situation in which most people will deem accuracy to be particularly important.

The jurors do not influence the first level process of interaction between participants except within their role as observers. Thus, it seems evident that the trial influences the jurors' impressions, but that these impressions do not influence the trial. This would simplify the analysis of the different time series. However, there is a person who may form impressions parallel and similar to those of the jurors and who is very much in a position to influence the trial process: the judge. Therefore, we have to think of, if not to implement in the model, the reciprocal influence of the processes at the first and second level.

The third process, i.e. the sequence of the raters' judgements of legal relevance, emotionality, etc., in the units of information presented and recorded during the trial and transcribed afterwards, is a mapping of the 'real' process as regards several important aspects. What was really going on in the trial appears in the content ratings in highly abstract terms and in interpretations of persons who did not attend the trial.

Half of the raters did not even follow the natural sequence of the information, but started with the last unit and ended with the first. This was meant to counterbalance the effects (autocorrelations) of expectations formed earlier on in the rating process.

Since the task of the raters was not to tell whether the defendant was guilty or not guilty, but whether the unit of information was legally relevant or not, how strongly it argued for or against the defendant, and whether its presentation was emotional or unemotional, concrete or abstract etc., there was no need to integrate the single pieces of information. Although we cannot exclude the fact that the raters soon formed some general impressions of the defendant's guilt and were consequently influenced by those impressions in their single content ratings, we assume that the two time series (one generated by the jurors, the other by the raters) are different in structure because the first one implies integration, whereas the second does not. Therefore, differences, i.e. step by step changes in the first series, may correspond to the original (not differenced) values of the second series.

Choosing the right ARIMA model

For various reasons, the ordinary least square regression analysis is not suitable for testing the hypotheses we are dealing with. Methods of multiple time series analysis seem more appropriate, of which ARIMA (autoregressive integrative moving average) models seem to be the most convenient (McCleary and Hay, 1980; Möbus and Nagl, 1983).

Before we can relate the raters' time series of the jurors' time series in order to explain the latter on the basis of the former, we have to decide which of the ARIMA models to choose for each of them.

Although it seems to be common practice with ARIMA models not to start with hypotheses and theoretical justification but to look which model fits the data best, we prefer to think first about what kind of process we expect for theoretical reasons.

Since jurors must continuously integrate past evidence in order to give repeated overall judgements of the defendant's guilt, we first have to test the most simple integrative model: ARIMA (0, 1, 0), the so-called random walk model.

$$Y_t = Y_{t-1} + a_t$$

where a_t is a normally distributed 'random shock' with zero mean.

The random walk model is compatible with the (not quite realistic) assumption that the jurors start at the indifference point of the scale and that the sequence of incriminating and exonerating evidence is randomly distributed around the indifference point.

Testing the model essentially means differencing the series ($Y_t - Y_{t-1}$) and looking at the autocorrelation functions (ACFs) which should not be significantly different from zero if the model remains valid.

If differencing ($Y_t - Y_{t-1}$) gives a mean of differences which is significantly different from zero, the original series implies a trend θ_0 in addition to the random walk components

$$Y_t = Y_{t-1} + \theta_0 + a_t$$

(cf. McCleary and Hay, 1980, p. 43).

This would be in line with the more realistic assumption that negative evidence is more frequent (i.e. on the average the guilt content of the arguments is negative), but nevertheless randomly fluctuating around the mean guilt load of arguments. In this case, a trend in the jurors' time series will be expected if the initial guilt judgement is different from the mean of the guilt content of the units of the trial evidence.

However, assuming random distribution of guilt content (around the average guilt load) is not very realistic either. As we have mentioned before, because of procedural rules and the jurors' 'natural' preferences for coherence and proper order in queries and answers, and for pro and con arguments, we will probably have longer runs of positive or negative arguments than would be expected by chance. This means that the ACFs of the differenced series may be significantly different from zero (towards the positive).

Another reason for a positive autocorrelation of the differenced time series of the jurors could be that noticing a change induces the expectation of further evidence corroborating this change. Such an expectation could make congruent evidence more influential than the incongruent one (cf. Lilly, 1984).

If the random walk model has to be rejected, i.e. if the ACFs are different from zero, we have basically two reasonable options: assuming that the difference series is generated by a first order autoregressive model ARIMA (1,1,0) or by a first order moving average model ARIMA (0,1,1).

Assuming that we have differenced the jurors' series, we have to choose between

$$\begin{aligned} \text{ARIMA (1,0,0): } Y_t &= \phi Y_{t-1} + a_t \text{ and} \\ \text{ARIMA (0,0,1): } Y_t &= a_t - \theta a_{t-1} \end{aligned}$$

(McCleary and Hay, 1980, p. 57 and p. 63).

What could be the psychological meaning of these models in our case? In seeking convincing psychological interpretations of the models, we should remember that the autocorrelations of the jurors' time series can be caused by the structure of the 'real' process (the specific sequence of evidence for or against the defendant) as

well as by the cognitive structures of the jurors. Let us imagine that during the trial each unit of information changes the guilty – not guilty balance of ‘objective’ evidence, depending on whether the argument is put on the left (guilty) or the right (not guilty) side of these ‘scales’. When a later argument is a repetition of a preceding argument, it will contribute less to a change of the balance. An argument which contradicts the preceding argument is usually less redundant; it should therefore have a stronger effect. This is represented rather well by a first order MA-model; i.e. ARIMA (0,0,1) with a positive weight ($0 < \theta < 1.0$) as both logical reasoning and simulation trials with variations of θ -coefficients have made clear.

ARIMA (1,0,0) can be interpreted more easily with respect to the information processing of the jurors. A positive ϕ -coefficient would be compatible with the assumption that each piece of information that has caused a change predisposes the subject to expect further evidence justifying his/her move, thus making him/her more responsive to information congruent with the past move.

Since there are reasonable arguments supporting ARIMA (1,0,0) and others which speak for ARIMA(0,0,1), we will have to test both models in order to find out which is the most appropriate.

METHOD

Observations in the courtroom

37 students (all of them law students in their second term at Linz University; 17 were female, 20 male) rated the ‘guilt’ of the defendant during a whole trial. In order to make the situation as realistic as possible, our mock jurors entered the courtroom at the same time as the actual jurors and had exactly the same information, except for being sworn in. They were carefully instructed to put themselves in the position of the real jurors and to form a valid judgement of the defendant’s guilt/innocence.

The defendant was accused of the attempted murder of his live-in partner. The court had to make the decision between (1) attempted murder (which is, according to Austrian Criminal Law, equivalent to murder in the first degree), (2) grievous bodily injury or (3) innocence.

The trial took one day, during which the presentation of evidence lasted from 8 to 12 a.m. After the lunch-break there followed the summing up by the prosecutor and the defence counsel, the instruction of the jurors by the judge, deliberation of the jury and the proclamation of the verdict by the judge.

Some days before the trial, our mock jurors had already been instructed in the use of the ‘guilt-scale’ (10 cm graphic rating scale from ‘guilty, very sure’ to ‘not guilty, very sure’; see Figure 1) and trained in the handling of the scale with fictitious tape-recorded passages of a trial. The subjects were to record *any change*,



Figure 1. 10 cm graphic rating scale with two extreme anchors

even the slightest one, of perceived guilt *any time* such a change occurred. In addition, the exact time, the speaker and a key-word were to be recorded, allowing for the necessary time marks in the process analysis.

The four hour period of evidence presentation in the court room was split into 80 3-minute-intervals. Since the time intervals of guilt ratings varied within and between subjects depending on the distribution of points in time when a juror perceived a change in his/her guilt impression, a subject’s last mark on the scale was kept till a new mark was made on the scale signalling a change in guilt perception. This means that we have a guilt impression score from each subject for each time interval. For every 3-minute-interval the guilt perceptions were averaged across the 37 mock jurors.

Content variables of transcripts

For content analysis of the trial transcripts, the first 50 consecutive information units (3-minute-intervals) were selected. This was also the period where the most dramatic changes occurred. All units had to be judged by eight raters (four male, four female) according to five different aspects (legal relevance, concreteness, emotionality, stimulation of compassion with the defendant and attribution of responsibility). Each group of raters was assigned to one aspect only in order to avoid transfer effects (halo) from one aspect to the other.

By ‘legally relevant information’ we mean all data which the law allows to be presented in a trial as evidence for or against the defendant. The raters received a clear description of what is defined by law as admissible or inadmissible evidence. Since most of the raters were law students it could be assumed that they had a precise understanding of this scale. ‘Concreteness of information’ is suggested as a possibly useful aspect by the research on cognitive processes to which among others Eagly and Chaiken (1984) refer.

‘Emotionality’ refers to the degree to which the information arouses emotions. The fourth scale deals with the dimension ‘compassion with defendant’. Certainly this scale is not semantically independent of emotionality, but it defines a facet of emotionality which, in another study, has been found to influence the jurors’ decision-making (Austin *et al.*, 1976).

‘Attribution of responsibility’ for the incriminated behaviour was included (see Fincham and Jaspers, 1980 and Haisch, 1980), acknowledging that perceived personal responsibility is not identical with, but, nevertheless, an important component of legal responsibility. The instruction for the raters was ‘This dimension describes the “intentionality” and therefore also “responsibility” for the incriminated behaviour. Internal causes exist if the motive for the criminal action lies within the person itself (e.g. motivation for revenge, possession of certain goods, etc.). External causes are causes in the environment (e.g. to be forced to commit a crime, influence of drugs, etc.).’

Finally, we have a global scale representing ‘comprehensive judgement of guilt load’ given with each single piece of information. This is the only directional variable (for or against the defendant). Should we have missed something with the more specific categories, we could check whether the residuals (variance not explained by the more specific variables) were related to the general impression of guilt load.

Each of the preceding (unidirectional) variables (legal relevance, concreteness of information, emotionality, attribution of responsibility) can be given direction by multiplying it with the sign (-1, +1) of the directional variable 'comprehensive judgement of guilt load'. The values of -1 and +1 are given for scores below or above the midpoint of the scale (extending from definitely for to definitely against the defendant).

RESULTS

Phenomenological interpretation of the time series of guilt judgements

Looking at Figure 2 it is clear at a glance that the subjective perception of guilt undergoes considerable fluctuations, as new evidence is added during the trial.

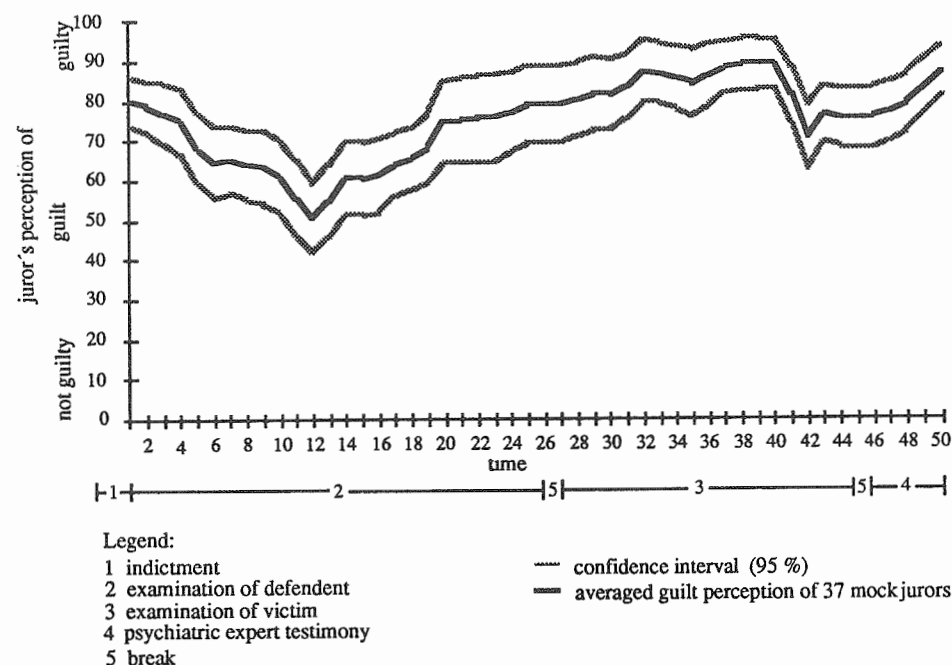


Figure 2. Averaged guilt perception of 37 mock jurors with 95% confidence intervals

Figure 2 shows the averaged trace of guilt perception by our 37 mock jurors with the 95% confidence interval during the first 50 time units (3-minute-intervals) of evidence presentation. During these 150 minutes of evidence presentation the defendant as well as the victim were examined and the expert testimony of a psychiatrist was read.

Let us begin with a purely descriptive interpretation of the curve. The initial value of guilt perception is relatively high. The first rating was made by most of our jurors after the indictment was read out, that is approximately 10 minutes after the opening of the trial. After this relatively negative rating, the defendant gains points. During this phase the judge explored the curriculum vitae of the defendant, his home, his professional background, career and family status. This biographical outline failed to show features which deviate much from the average; it did not

show a typical criminal career. Of particular note was a cumulation of difficult life events (early loss of parents, frequent change of employment, serious road accident followed by a partial invalidity which led to unemployment and divorce).

Until about 45 minutes later there is a marked upturn in the curve, showing the defendant as being increasingly felt to be guilty. For the next hour, the defendant continues to lose ground. At this time the defendant was examined about the antecedents of the attack against the victim and the attack itself. The defendant had to report how he became acquainted with the victim, about the circumstances which led to the attack and how the attack was conducted. The beginning of the examination of the victim, who also had to report the antecedents of the criminal act and the act itself, falls within this period of the trial.

Another characteristic turn occurs after about two hours when things seem to go better for the defendant for a short time. During this time the victim made statements about the mental state (question of diminished responsibility) of the defendant of the time of the attack. The victim testified that the perpetrator was under the influence of alcohol and drugs. However, his position deteriorates again towards the closing of the session. This was caused by the expert testimony of the psychiatrist. He stated that the defendant was not in a state of diminished responsibility due to drugs or a low level of intelligence.

In this case observers started off with a rather high subjective probability of 'guilty'. This corresponds with interview studies of Kairys *et al.* (1975) and Saks *et al.* (1975) who found that between 50% and 60% of potential jurors think the defendant is guilty, even before they have any information about the case.

Obviously, the defendant's right to be considered innocent until proven guilty is largely hypothetical. A juror without any knowledge of the law will come to the conclusion that 'there must be something to it' if a person is subjected to the whole process of police questioning, legal inquiries, formal charges and a full-scale trial.

In all phases of the trial, the 17 female raters turned out to be more severe than their 20 male counterparts (Figure 3). This is no surprise, since the victim was a woman. Obviously, differences in emotional involvement and identification with the victim or with the defendant make a difference in guilt perception. This raises the question of sex distribution in juries. In general, women - particularly young women - are more on the defendant's side than male jurors (Efran, 1974; Sealy and Cornish, 1973). But there are also many studies on the influence of the sex of the juror on decision-making which show inconsistent or nonsignificant effects. However, there is no research on how the type of offence influences the attitude of male vs. female jurors. In our case, a jury with a female majority would have been less favourable for the defendant; as this is known to lawyers and judges, they are aware of the legal possibilities in controlling the composition of juries, if at least indirectly.

The curve traced by the male observers is on the whole parallel to that traced by the female observers, which underlines the high reliability of the ratings ($r_{mf} = 0.88$).¹

¹ Nevertheless the difference between male and female guilt ratings is highly correlated with the content variables 'compassion with the defendant' ($r = 0.66, p < 0.001$), 'attribution of responsibility' ($r = -0.43, p < 0.01$) and 'concreteness of information' ($r = 0.43, p < 0.01$). This indicates that sex differences in guilt perception are partly explainable by those variables. Male jurors, compared to female jurors, are more lenient with the defendant the more the content of the trial process suggests compassion with the defendant, lack of responsibility and concreteness of information.

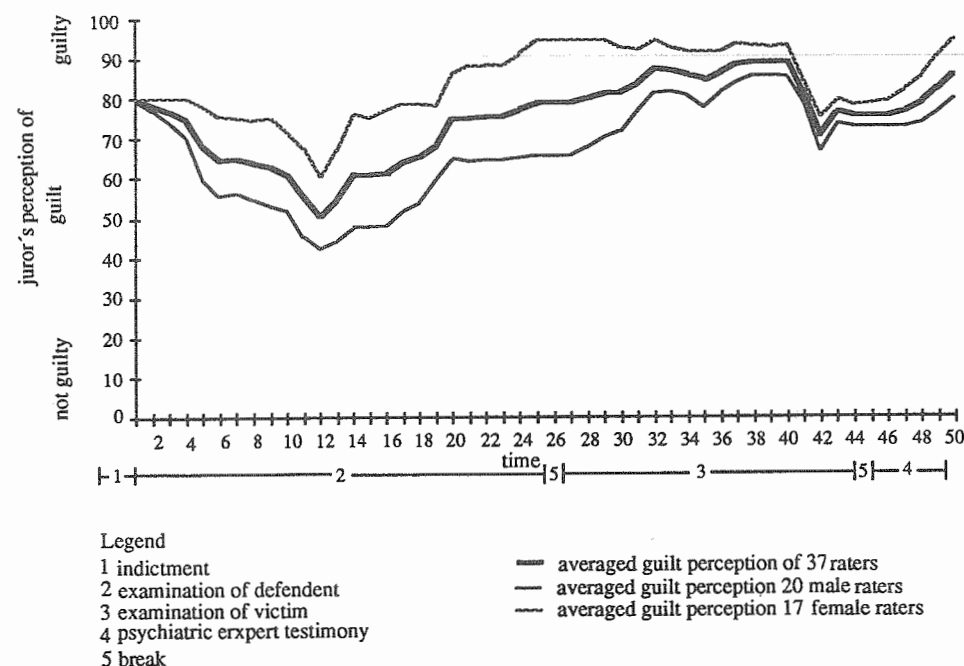


Figure 3. Averaged guilt perception of male and female raters

Content analysis of trial transcripts

We have six content scales: (1) legal relevance (irrelevant-relevant), (2) concreteness of information (abstract-concrete), (3) emotionality of the content (unemotional-emotional), (4) compassion for the defendant (low-high), (5) attribution of responsibility for incriminated behaviour (external-internal causes) and (6) global guilt load (low-high).

As an indication of the reliability of ratings we took the average intercorrelation which is known as the intraclass correlation (Guilford, 1954, p. 395). Table 1 shows the intraclass correlations (r_{cc} and r_{kk}) for each scale.

Table 1. Intraclass correlation of ratings (8 raters, $n = 49$)

| Scale | r_{11} | r_{88} |
|-------------------------------|----------|----------|
| Legal relevance | 0.54 | 0.90 |
| Concreteness | 0.22 | 0.69 |
| Emotionality | 0.41 | 0.85 |
| Attribution of Responsibility | 0.32 | 0.79 |
| Compassion with defendant | 0.51 | 0.89 |
| Overall guilt load | 0.24 | 0.71 |

Note: r_{11} is the reliability of the judgements of one rater, r_{88} represents the reliability of the averaged judgements of eight raters.

These variables are potential predictors of the step by step changes in guilt impression of the jurors. The correlations between the five variables are shown in Table 2. The reliability of all scales is acceptable, so we were in the position to use all of them. Nevertheless we abandoned 'compassion for the defendant' because of its low predictive power for the perception of guilt ratings.

Table 2. Correlation between the dependent variable (GJC) and the five (unidirectional) independent variables

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|-------|-------|-------|------|-------|-------|
| 1. Guilt Judgement Change | — | 0.47 | 0.36 | 0.30 | -0.36 | -0.14 |
| 2. unirect. Legal Relevance | 0.56 | — | 0.16 | 0.32 | -0.18 | -0.22 |
| 3. unirect. Concreteness | 0.42 | 0.18 | — | 0.47 | -0.06 | 0.38 |
| 4. unirect. Emotionality | 0.38 | 0.26 | 0.35 | — | 0.19 | 0.16 |
| 5. unirect. Attribution of Responsibility | -0.50 | -0.26 | -0.38 | 0.14 | — | -0.43 |
| 6. Compassion with defendant | -0.13 | -0.30 | 0.37 | 0.11 | -0.54 | — |

Note: The correlations between the pre-whitened form of variables are shown above diagonal.

ARIMA time series analysis of unidirectional variables

For discussion of the logic of ARIMA time series analysis the reader may refer to McCleary and Hay (1980) or Möbus and Nagel (1983). Here we present the results only with some general clarifications of the types of models we used.

The first step in time series analysis is making sure that the process is stationary, i.e. that there is no drift or trend in the data. Only the Guilt Judgement Curve (Figure 2) proved to be non-stationary. First order differencing ($z_t = Y_t - Y_{t-1}$) resulted in a time series ('guilt judgement change' GJC) with ACFs (autocorrelation functions) close to zero. Thus, contrary to our expectations, the Guilt Judgement Curve turned out to follow a random walk model quite closely, i.e. ARIMA (0,1,0) which results in an ARIMA (0,0,0) after differencing, which gives us the unidirectional GJC variable.

The situation is different with the time series of the independent variables. Although stationary, the inspection of the ACFs and PACFs (partial autocorrelation functions) showed that they represent an ARIMA (1,0,0):

$$Y_t = \phi_1 Y_{t-1} + a_t;$$

and not an ARIMA (0,0,1):

$$Y_t = a_t - \theta a_{t-1}.$$

This means that at any point in time the 'random shock' a_t (a_t can be conceived of the unpredictable content of the information unit) is added to a discounted value of the judgement score at time $t-1$, whereby $0 < \phi < 1$. This is in line with our expectation concerning the information processing of the jurors.

The pattern of between-series correlations was used to identify a transfer function relationship between GJC series and the four input series, in much the same way that the ACF was used to identify an ARIMA relationship within the time series. Transfer functions are analog to multiple regression in ordinary least square analysis. Before we can determine the cross correlation functions (CCFs) for gauging the impact of the several input time series on our output series, we have to 'pre-whiten' the predictor time series (McCleary and Hay, 1980, p. 243f.), i.e. to remove the effects of autocorrelations.

Table 3 shows the CCF from lag-2 to lag-zero for the output series with all pre-whitened input series. A correlation at lag-2 e.g. would refer to a delayed impact of the presented information by a time lag of 6 minutes.

The highest correlations are at lag-zero, which indicates an immediate influence of presented information on changes in guilt perception. There is also a delayed effect – the correlations at lag-1 are lower, but still exist. This effect disappears at lag-2.

Table 3. Cross correlation function (CCF) Dependent variable (GJC) with independent

| GJC with | -2 | lag -1 | 0 |
|-------------------------------|-------|-----------|-------|
| Legal relevance | -0.09 | 0.38 | 0.47 |
| Concreteness | 0.11 | 0.31 | 0.36 |
| Emotionality | -0.04 | 0.23 | 0.30 |
| Attribution of Responsibility | 0.01 | -0.35 | -0.36 |
| Compassion with defendant | 0.12 | -0.09 | -0.14 |

Transfer Function

The Transfer Function describes the dynamic causal relationship between two or more series. In our case a multiple transfer model of the following form seemed suitable:

$$Y_t = f(X_{1t}) + f(X_{2t}) + \dots + f(X_{nt}) + N_t$$

Y_t stands for output series and $X_{1t}, X_{2t}, \dots, X_{nt}$ for the input series. This model is a generalisation of the bivariate transfer model:

$$Y_t = f(X_t) + N_t = Y_t^* + N_t$$

whereas Y_t^* includes the part of Y_t which is predicted by the variable X_t ; N_t stands for the error of causal variables not included in the model. N_t follows an ARIMA (p, d, q) model. The input variable X is an observed one, thus the transfer function must be empirically identified in advance of the estimation of the parameters. Another way of presenting the transfer model is

$$Y_t = \frac{\Omega(B)}{\delta(B)} X_t + N_t = \nu(B) + N_t$$

This model includes the moving average operator $\Omega(B)$ and the autoregressive operator $\delta(B)$, which represents the inner structure of the input and output series.

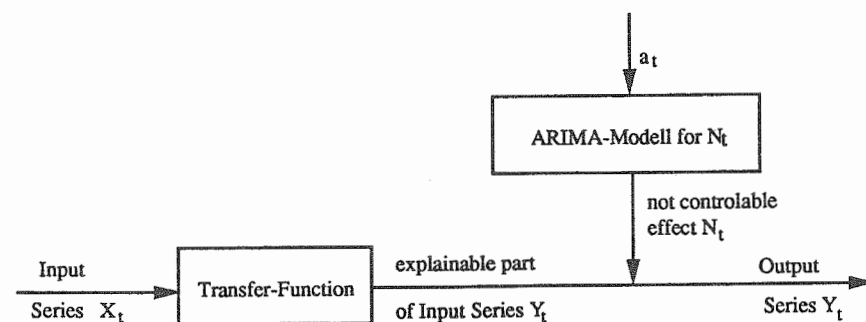


Figure 4. Process of the transfer model (dynamic form of the multiple regression)

For our purpose we used the multivariate extension of this model because we have more than one input series (see Möbus and Nagel, 1983, p. 271):

$$Y_t = \frac{\Omega_1(B)}{\delta_1(B)} X_{1t} + \frac{\Omega_2(B)}{\delta_2(B)} X_{2t} + \frac{\Omega_3(B)}{\delta_3(B)} X_{3t} + \dots + \frac{\Omega_n(B)}{\delta_n(B)} X_{nt} + N_t = \nu_1(B) X_{1t} + \nu_2(B) X_{2t} + \nu_3(B) X_{3t} + \dots + \nu_n(B) X_{nt} + N_t$$

The identification of the model takes several steps. Univariate ARIMA models must be built for all input series. All non-stationary series must be differenced appropriately to become stationary. Every X_{jt} series has to be pre-whitened to an α_{jt} series. The output series X_t is transformed the same way to the series β_{jt} . Now we can be sure that the CCF between α_{jt} and β_{jt} does not contain any autocorrelation and we can take the CCF to estimate the parameters ν_j of the transfer model.

After that, a suitable ARIMA Modell for N_t has to be identified, and the transfer model must be tested. The process of identification and estimation is finished when the model fits the data, a_t represents white noise and $\sigma^2_{a_t}$, the error variance, is a minimum (see Figure 5). In our case we had the following transfer model:

$$\begin{aligned} \text{GJC}_t &= f_1(\text{LR}_t) + f_2(\text{C}_t) + f_3(\text{E}_t) + f_4(\text{AR}_t) + N_t = \\ &= \nu_{11}\text{LR}_t + \nu_{12}\text{LR}_{t-1} + \nu_{21}\text{C}_t + \nu_{22}\text{C}_{t-1} + \nu_{31}\text{E}_t + \nu_{41}\text{AR}_t + \nu_{42}\text{AR}_{t-1} + N_t = \\ &= (0.0464 \text{ LR}_t) + (0.0342 \text{ LR}_{t-1}) + (0.0431 \text{ C}_t) + (0.0246 \text{ C}_{t-1}) + \\ &\quad + (0.0396 \text{ E}_t) + (-0.0656 \text{ AR}_t) + (-0.0634 \text{ AR}_{t-1}) + N_t. \end{aligned}$$

and for the remaining series N_t we identified an ARIMA (0,0,0) model. GJC stands for guilt judgement change, LR for legal relevance, E for emotionality of the content, C for concreteness of information and AR for attribution of responsibility for incriminated behaviour. All of the series were pre-whitened.

ARIMA time series analysis of directional variables

In the preceding paragraph we analyzed the time series of averaged unidirectional individual changes in the jurors' guilt impressions using the likewise unidirectional variables legal relevance, concreteness, emotionality, and responsibility attribution as independent variables.

Now we used the directional changes of the averaged jurors' guilt impressions as dependent time series and exonerating vs. incriminating evidence weighted by (a) its legal relevance, (b) concreteness, (c) emotionality, (d) responsibility, (e) overall guilt load, in this order as independent time series. Thus, we could see whether extra-legal characteristics of the trial information had had some impact on the juror's guilt judgement in addition to the legally relevant aspects.

Since the directional independent variables are highly correlated (following from the fact that all variables have been multiplied by the same dichotomic variable 'for vs. against the defendant') we did not follow the logic of stepwise regression analysis but entered the variables in an order which seemed most adequate, theoretically speaking. Thus, the overall guilt load has to be entered last in order to find out if the raters, in their global judgements, picked up some valid information from the transcripts which is not covered by the more precisely and narrowly-

defined aspects of legal relevance, concreteness, emotionality and responsibility attribution. We entered concreteness prior to emotionality because the former can be conceived of as a component of the latter. Responsibility attribution is entered second to last because it is conceptually close to, but less comprehensive than global guilt load.

The only valid predictor of directional changes of guilt perception is directional legal relevance which was entered first. The cross-correlation (time lag-zero) is $r = 0.68$. There is no incremental validity in the other variables.

The simple direction of the argument (pro or con) gives a cross-correlation (time lag-zero) with the dependent variable of $r = 0.60$. Thus, weighting the direction by legal relevance of the argument results in an increment of validity.

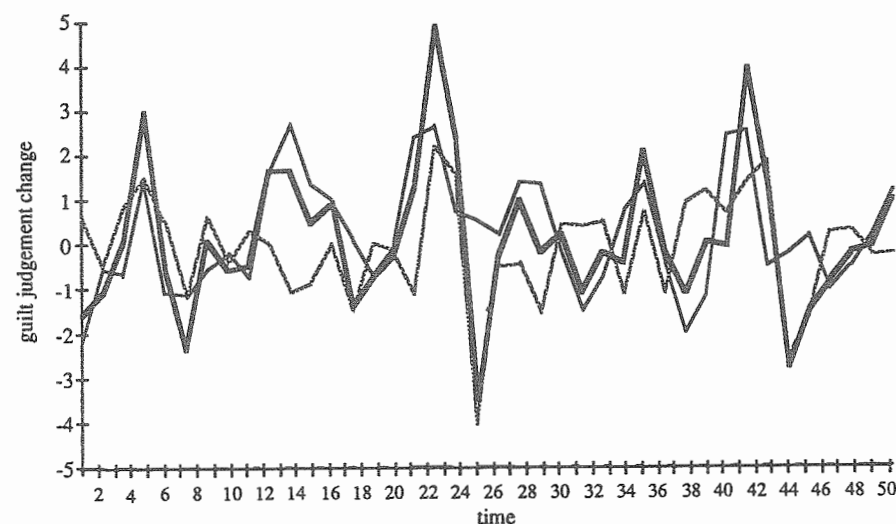


Figure 5. Standardised time series for guilt judgement change:
— observed, — estimated, --- residuals (observed - estimated)

DISCUSSION

First we have analyzed the time series of the averaged individual unidirectional changes $\sum_{i=1}^n \frac{(|Y_i - Y_{i-1}|)}{n}$. By eliminating the direction of change we can take

into account all individual changes caused by a unit of information, whether they go up or down the guilt perception scale. Even if some subjects interpret a piece of information as incriminating, others as exonerating, the sheer amount of change is supposed to depend on variables such as legal relevance, concreteness etc.

We found that, in addition to legal relevance, concreteness and responsibility attribution are significant predictors of averaged individual unidirectional changes. Unexpectedly, emotionality turned out to be irrelevant. Obviously, what emotionality has to contribute to the prediction of unidirectional change is absorbed by legal relevance, which is somewhat correlated with emotionality ($r = 0.32$; cf. Table 2).

There are two possible interpretations for such a correlation. We can assume (a) that legally relevant facts are often emotionally arousing, (b) that the judgement of legal relevance is biased by emotional arousal. Only experimental manipulation of the conceptually different but empirically related aspects of the information units could tell us which of the two interpretations is more likely.

How is it that unidirectional concreteness and unidirectional responsibility attribution have some incremental validity beyond the validity of unidirectional legal relevance, whereas directional concreteness and directional responsibility attribution have no incremental validity?

In predicting directional change, the direction of the argument alone already explains 36% of the variance. Weighting the direction of legal relevance results in an increase of explained variance up to 49% leaving little room for a further increment, particularly for variables which are highly correlated with the variable (legal relevance) entered first.

Why the other independent variables have no incremental validity becomes clear if one looks at the intercorrelations of these variables (Table 4): they are extremely high (between $r = 0.46$ and $r = 0.95$; after pre-whitening between $r = 0.24$ and $r = 0.87$).

Table 4. Correlation between the dependent variable (GJ) and the five (directional) independent variables

| | 1 | 2 | 3 | 4 | 5 | 6 |
|--|-------|-------|-------|-------|-------|-------|
| 1. Guilt Judgement | — | 0.55 | 0.24 | 0.53 | 0.59 | -0.30 |
| 2. direct. Legal Relevance (directional) | 0.52 | — | 0.51 | 0.78 | 0.79 | -0.37 |
| 3. direct. Concreteness (directional) | 0.36 | 0.75 | — | 0.73 | 0.66 | -0.18 |
| 4. direct. Emotionality (directional) | 0.49 | 0.90 | 0.74 | — | 0.87 | -0.27 |
| 5. direct. Attribution of Responsibility | 0.48 | 0.95 | 0.79 | 0.91 | — | -0.26 |
| 6. Compassion with defendant | -0.84 | -0.49 | -0.43 | -0.49 | -0.46 | — |

Note: The correlations between the pre-whitened form of variables are shown above diagonal.

Even if each of them were to be entered first one by one, they would add little to the validity of the direction of the arguments. The correlation with the global guilt load is even remarkably lower than the correlation with the argument direction ($r = 0.42$ vs. $r = 0.60$).

This could mean that the jurors were quite sensitive to the legal relevance of the information and did not uncritically follow legally unacceptable arguments for or against the defendant's guilt. Since our subjects were second-year law students already trained to some degree in differentiating between legally relevant and legally irrelevant evidence, we cannot be sure that the same would happen with real jurors.

The cross-correlations with the residuals (from predictions based on legal relevance) of the remaining variables are between $r = 0.15$ and $r = -0.04$; i.e. they are virtually zero.

With unidirectional change as dependent variable, however, we have no overriding single variable, but several variables which are not highly intercorrelated, and each of them modestly correlated with the dependent variable, thus allowing for their unique contributions to the prediction of change.

Whether we try to predict directional or unidirectional change, in both cases about 50% of the variance remain unexplained. Since most of the measures are

highly reliable we have to assume that non-verbal characteristics of the trial process have a major impact on the observers' judgements. The transcripts are void of those non-verbal utterances (facial, postural, or vocal expressions), thus depriving us of the possibility to take the most significant carriers of emotions into account. It may well be that this is the main reason why legal relevance is predictive and emotionality is not.

Final comment

To our knowledge, the study reported here is the first one where the subjects (the mock jurors) attend a real trial in the courtroom applying the judgement method of just noticeable changes. It is probably also the first time that the permission for tape recording the whole trial was given by the court and that ARIMA models were used in explaining the time series of the mock jurors' guilt impressions by time series of ratings representing relevant characteristics of the single units of information.

Considering the enormous investment of time and effort in collecting and analyzing the data, we asked ourselves several times whether we were right in leaving the laboratory and the familiar and comfortable means for producing a neat, statistically significant result in testing another tiny hypothesis about one or another condition of jurors' decision-making.

We now feel reassured that, although the pilot study did better in exploring the feasibility of a new method in a natural setting than in testing a specific theory, we are moving in the right direction.

Future studies of this kind should incorporate some non-verbal characteristics of the courtroom behaviour in the time series analysis. This could be done (a) by having several observers in the courtroom who can continuously code their impressions in theoretically meaningful dimensions or categories, (b) coding some non-verbal (vocal) aspects of tape-recorded communication. Unfortunately, in our study the low technical quality of our tape recordings did not allow such an analysis.

Moreover, some of the categories used in the content analysis need to be replaced or revised to provide a more thorough theoretical foundation of the courtroom discourse analysis. The most important and, at the same time, most difficult task will be to integrate theoretical ideas from different, although somehow related and overlapping fields of research (persons perception, information processing, attitude change, decision-making, etc.) into a realistic and verifiable model of jurors' impression-formation which precedes and pre-structures their final interactive decision-making.

ACKNOWLEDGEMENT

This study was sponsored by the Fonds zur Förderung der wissenschaftlichen Forschung Project Nr.:P5747 and Nr.:J70. We have to thank Veronika Papula for assistance in the statistical analysis and we also owe thanks to our legal friends, whose interest in our field has made the data collection possible. *Correspondence* should be addressed to Gerhard Kette, Department of Psychology, University of Linz, A-4040 Linz, Austria.

REFERENCES

- Anderson, N.H. (1982). *Method of Information Integration Theorie*, Academic Press, New York.
- Anderson, N.H. and Graesser, C.C. (1976). 'An information integration analysis of attitude change in group discussion', *Journal of Personality and Social Psychology*, **34**: 210-222.
- Austin, W., Walster, E. and Utne, M.K. (1976). 'Equity and the law: the effect of a Harmdoer's "Suffering in the Act" on liking and assigned punishment'. In: Berkowitz, L. and Walster E. (Eds) *Advances in Experimental Social Psychology*, **9**: 163-190. Academic Press, New York.
- Brandstätter, H. (1985). 'Social emotions in controversial discussions and in group decision making'. In: Lawler, E.E. (Ed.) *Advances in Group Processes*, **2**: 249-281, JAI Press, Greenwich.
- Brandstätter, H. (1987). 'Towards differential social psychology: individual differences in responding to an aggressive discussant.' In: Semin G.R. and Krahé B. (Eds) *Issues in Contemporary German Social Psychology*, 55-73, Sage, London.
- Burnstein, E. (1982). 'Persuasion as argument processing'. In: Brandstätter, H., Davis J.H. and Stocker-Kreichgauer G. (Eds) *Group Decision Making*, 103-124, Academic Press, London.
- Eagly, A.H. and Chaiken, S. (1984). 'Cognitive theories of persuasion'. In: Berkowitz L. (Ed.) *Advances in Experimental Social Psychology*, **17**: 267-359, Academic Press, New York.
- Fincham, F.D. and Jaspers, J.M. (1980). 'Attribution of responsibility: from man the scientist to man as lawyer.' In: Berkowitz L. (Ed.) *Advances in Experimental Social Psychology*, **13**: 81-138. Academic Press, New York.
- Foss, D.J. (1988). 'Experimental psycholinguistics', *Annual Review of Psychology*, **39**: 301-348.
- Guilford, J.P. (1954). *Psychometric Methods*, 2nd edn, McGraw, New York.
- Haisch, J. (1980). 'Anwendung von Attributionstheorie als normative Modell für eine rationale Strafzumessung', *Zeitschrift für experimentelle und angewandte Psychologie*, **27**: 415-427.
- Higgins, E.T. and Burgh, J.A. (1987). 'Social cognition and social perception', *Annual Review of Psychology*, **38**: 369-425.
- Kairys, D., Schulman, J. and Harring S. (Eds) (1975). *The Jury System: New Methods for Reducing Prejudice*, National Lawyers Guild, Philadelphia.
- Lilly, W. (1984). 'Die Hypothesentheorie der sozialen Wahrnehmung'. In: Frey, D. and Irle M. (Eds) *Theorien der Sozialpsychologie*. Bd I: Kognitive Theorien, 19-46, Huber, Bern.
- McCleary, R. and Hay, R.A. Jr. (1980). *Applied Time Series Analysis for the Social Sciences*. Sage, Beverly Hills.
- Möbus, C. and Nagl, W. (1983). 'Messung, Analyse und Prognose von Veränderungen'. In: Bredenkamp J. and Feger H. (Eds) *Hypothesenprüfung. Enzyklopädie der Psychologie* **B15**: 239-470, Hogrefe, Göttingen.
- Pennington, N. and Hastie, R. (1981). 'Juror decision-making models: the generalization gap', *Psychological Bulletin*, **89**: 246-287.
- Saks, M.J., Werner, C.M. and Ostrom, T.M. (1975). 'The presumption of innocence and the American juror', *Journal of Contemporary Law*, **2**: 46-54.
- Tomkins, S.S. (1980). 'Affect as amplification: some modifications in theory'. In: Plutchik R. and Kellerman H. (Eds) *Emotion: Theory, Research, and Experience*. **1** (Theories of Emotion): 141-164, Academic Press, New York.