

# VERDICTS AND DELIBERATION CONTENT ANALYSIS OF JUDGES AND JURIES IN THE SPANISH LEGAL CONTEXT

Ramón Arce\*, Francisco Tortosa\*\* and Elisa Alfaro\*\*\*

\*University of Santiago de Compostela, \*\*University of Valencia, \*\*\*University Jaume I, Castellón

---

*Verdicts and deliberation content analysis of judges and juries in the Spanish legal context. A comparative approach. In view of the fact that jury performance has been subject to considerable criticism in various countries, the aim of the present paper is to compare the performance of judges and juries. A total of 15 nine-member juries and 15 three-member panels of judges deliberated and reached a verdict on real cases. The results of a comparative analysis reveal that: a) neither judges nor juries perform in a normative way; b) judges perform better than juries; c) judges and juries have different tendencies, juries towards the demands of the defence and judges towards the demands of the prosecution; and d) juries' deliberations are not more exhaustive than those of judges, as we might expect in order to compensate for their lack of legal knowledge and experience, and tend to be redundant in content. Finally, we discuss the practical implications, especially in relation to mechanisms for mitigating the non-normative decision-making of judges and juries.*

*Sobre la base de que son numerosas y consistentes inter-contextos jurídicos las críticas a la eficacia de los Tribunales de Jurados, nos hemos planteado un estudio comparativo de la ejecución de Tribunales de Jueces y Jurados en el ordenamiento jurídico español (L.O. 5/1995). Tomando como material estimular tres casos reales, formamos 15 Tribunales de Jurados y 15 Tribunales de Jueces, que emitieron el objeto del veredicto tras la discusión grupal. Los resultados mostraron que tanto Jueces como Jurados no ejecutan de un modo normativo; que la ejecución de los Jurados es de peor calidad; que Jueces y Jurados ejecutan tareas distintas, más orientados a la defensa los Jurados y más a la acusación los Jueces; y que las deliberaciones de los Jurados no son más profundas para compensar su falta de conocimientos en la materia. Finalmente, discutimos las implicaciones para contrarrestar estos déficit en la actuación de Jueces y Jurados.*

Juries have often been accused of incapacity, ineptitude and lack of background or training for exercising the function entrusted to them as members of the judicial apparatus. In brief, juries have been labelled as incompetent. Nevertheless, systematic research on the subject does not appear to lead to the same conclusions. In one study in this field, Kalven and Zeisel (1966), basing themselves on judges judgements, categorized cases already judged by juries as easy, difficult and very difficult. If juries were incompetent, the divergences on the verdict between judges and juries should occur in the difficult and, above all, very difficult cases. However, the data suggest that the discrepancies are evenly distributed across the three conditions, so that it cannot be concluded that the jury fails to understand the evidence presented. In an attempt to replicate the work of Kalven

and Zeisel, Baldwin and McConville (1979) found that judges sometimes understood that the jury had applied "what was fair" rather than "what was legal". These results were confirmed by Myers (1979), who observed, after studying 201 penal cases, that juries rarely deviated from what was legal, and that when they did so it was not out of incompetence, but rather in pursuance of their perceptions of what was "fair and right". In a complementary explanatory line, MacCoun and Kerr (1988) showed that judges and juries differ in that there is a tendency of judges towards guilty verdicts in cases of reasonable doubt, whilst juries in such cases tend towards not guilty verdicts. Where there are results that may back up accusations of incompetence is in the area of the understanding of instructions (Elwork, Sales & Alfini, 1977) and in civil cases, especially complex ones (e.g., Chin & Peterson, 1985).

As regards judicial decisions made by judges, the literature has focused on the study of disparity (for an exhaustive review, see Kapardis, 1997). There are numerous studies indicating great disparity in judicial decisions, in relation to both the decision to imprison (e.g.,

---

The original Spanish version of this paper has been previously published in *Psicothema*, 2003, Vol. 15. No 1, 127-135

.....  
Correspondence concerning this article should be addressed to Ramón Arce, Facultad de Psicología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain.

E-mail: psarce@usc.es

Partridge & Eldridge, 1974; Diamond & Zeisel, 1975) and length of sentence (e.g., Sutton, 1978). Attempts to explain this disparity point to the transcendence of both legal and extra-legal variables. As regards variables of a legal nature, and after a review of 140 studies, Kapardis (1985) found the following legal factors as the most relevant in explaining the variability of sentences: recentness of previous conviction; criminal history; type of charge; previous interaction with the judicial system; sentence recommended by the officer supervising the conditional discharge; and provocation by the victim of the crime. Recently, one of us (Arce et al., 2001) found that the evaluation of each item of evidence in the reliability (e.g., credibility of the different testimonies) and validity (value of each piece of proof for the decision) dimensions, a legal factor, explains the majority of the variance of the disparity. Among the numerous extra-legal variables studied for their influence on the sentence, one of the most notable, in view of its relevance and consistency, is constituted by the role of the deciding agent, and, in relation to this agent, the penal orientation of a rehabilitative or utilitarian nature that gives rise to disparate decisions (e.g., Sobral & Prieto, 1994). Finally, a further line of research in the extra-legal area, carried out in Spain, has shown through the study of past sentences that they are largely determined by the use of heuristics, that is, through systematic judgement biases (Garrido & Herrero, 1995; Arce et al., 1996; Fariña, Novo & Arce, 2002).

In sum, the scientific literature has given sufficient indication that legal decisions are subject to biases and, by extension, to the commission of errors; indeed, among the goals of research has been that of seeking solutions to these shortcomings (e.g., Heuer & Penrod, 1994). However, the great majority of these results cannot be extrapolated to the Spanish judicial system. As far as the work of the jury is concerned, it is regulated by a statute (*Ley Orgánica 5/1995 del Tribunal de Jurado*) that makes provision for action within the penal framework, restricted to certain offences and aimed at reaching a verdict. The concept of verdict in this context goes beyond the judgement of guilty or not guilty, being based as it is on a questionnaire drawn up by the presiding magistrate based on the allegations of the two parties. Likewise, it differs in its composition and the rules that define it: nine members who decide by qualified majority on each item of evidence. This phenomenology

of the jury has some implications that differentiate it in relation to decisions reached in other modalities (Fariña, Arce & Vila, 1999).

Within this context, we set out to make an experimental study, highly faithful to real situations, comparing the performances of judges and juries in the Spanish judicial system. The chief objective was to compare the content of group discussions and decisions reached, which constitutes the central axis of decision-making that is frequently ignored in scientific research.

## METHOD

### *Participants*

The sample was made up of two well differentiated groups. On the one hand, a total of 135 jurors. All of them were adults, with full civil and political rights, with their certificate of secondary education, and with no physical or mental incapacities – the basic requirements for jury service. Mean age was 25.53 ( $S_x = 7.73$ ), with a range of 18 to 50 years. By gender, and taking into account the participants' voluntary responses, 18.6% were males and 81.4% females. As regards marital status, 87.3% reported being single, while 16.3% were married. The other part of the sample was made up of judges with a minimum of one year's experience (the sociodemographic data of this part of the sample were not provided, at their request, to avoid their being identified).

### *Procedure*

The jurors were previously instructed about the duties of a jury and what their task would be. Next, both judges and jurors were shown a video-recording of one of the real cases described in the Materials section. In all, 15 juries were formed, with nine members each, and 15 panels of judges, with three members each. When the groups had been formed, the task began. It involved four steps:

- a) The juries and judges provided individually a series of sociodemographic data (e.g., name, age, gender, marital status, educational level, place of residence). The sample of judges expressly requested exclusion of this data in order to avoid identification.
- b) They watched the video-recording of a real case (see Materials section).
- c) They then proceeded to consider their decision. These deliberations, which were recorded on video, had no time limit, and the juries were told that that

they must (in accordance with the rules in the Spanish system) try to reach a unanimous decision. It was not necessary to give instructions to the panels of judges, given their knowledge, nor to apportion roles, since they were actually performing as judges.

d) Finally, we recorded the group verdict as it had been defined in the original case.

### **Material**

Juries and judges were presented with one of three real cases recorded on video, which are summarized below.

In Case I, against J.M.A.<sup>1</sup>, the public prosecutor's summary was "...the accused J.M.A., adult and with paranoid disorder that affects the psychobiological bases of his imputability, on day D around time H, when his neighbour E.L.C. left her house to walk to her car, appeared suddenly, carrying an iron bar 64 cm long and 12 mm in diameter, and with the intention of killing her, struck her on the head from behind. As a result, Mrs. E.L.C. fell to the ground, where J.M.A. continued to hit her on the head and body, until another neighbour intervened. Mrs. E.L.C. suffered craneo-encephalic trauma with bruising and an open wound measuring 15 cm on the scalp, which prevented her from carrying out her normal duties for 22 days. Therefore, the public prosecution considers the offence as one of *asesinato frustrado*. The accused's offence is partially mitigated by mental derangement. The sentence recommended is of 10 years' imprisonment plus costs".

For the private prosecution, the facts constitute an offence of *asesinato frustrado*. There is the extenuating circumstance of mental derangement. The recommendation is a sentence of twelve years' imprisonment plus costs.

The summary by the defence counsel was as follows "... in disagreement with the summary of the public prosecutor and with the facts as described by the private prosecution, since they omit to mention the state of excitation and continued aggression to which my client was subjected by the person assaulted; also to be taken into account is his unbalanced mental state over several years; and this without pretending to justify the rest of the events and the specific assault committed. In a person with a disturbed mental state and with his volition

totally distorted, we cannot delimit his voluntariness, his intention or lack of it for causing injuries as serious as those caused, given his lack of awareness of what he was doing. Consequently, we consider the events as constituting an offence of *asesinato frustrado*, with the mitigating circumstance of mental derangement".

In Case II, against J.C.O., the public prosecutor's summary was "... the accused, J.C.O., adult and with a paranoid disorder that notably affected his intellectual and volitional faculties, on day D went to a field belonging to J.A.A. in search of him, moved by a desire to kill him because of disagreements they had had over some land. The accused pursued J.A.A. with an axe, but was unable to fulfil his intention due to the intervention of E.V. Three months later, the accused, moved by the same desire and carrying a wooden pole to which he had fixed an axe, went to the same place, once again in pursuit of J.A.A., and saying the words "I'm going to kill you". He was unable to carry out his threat thanks once more to the intervention of E.V., who was there at the request of J.A.A., given the latter's fear that the accused might try to kill him. E.V., in his car, drove up to J.A.A. so that he could jump into the car and thus escape. I therefore consider the facts to constitute two offences of *homicidio en grado de tentativa*. There is partial mitigation of mental derangement. I recommend a sentence of two years' imprisonment for each one of the offences".

For the private prosecution, "...the facts constitute two offences of *homicidio en grado de tentativa*. There is partial mitigation due to mental derangement. The recommended sentence is of two years' imprisonment for each one of the offences".

Finally, the defence summarized the case as follows, "...in disagreement with the summary of the public prosecution and of the private prosecution, considering that the account of the facts as reported by the prosecution is at odds with the true events.... The events as they actually occurred do not constitute any type of offence, or at most one of threatening behaviour. In addition, there is the mitigating circumstance of mental derangement, and indeed we may well consider the mental derangement as an extenuating circumstance".

In Case III, against F.M.C., the public prosecutor's summary was as follows: "... the accused F.M.C., adult,

<sup>1</sup> Names of people and places were changed so that actual cases could not be identified. As regards the different summaries, we present a literal transcription, including the legal jargon employed. The facts are only related in accordance with the descriptions by the public prosecutor and in the other official interventions.

was with S.E.L in the house in which they lived together, when the accused became annoyed by the conversation. Suddenly and unexpectedly, F.M.C. took a knife 9 cm long and 2 cm wide from his pocket, and with the intention of killing him, jumped on S.E.L. and stabbed him twice. One of the stabs was in the left hemithorax, and caused a wound 2 cm long, 9 cm below and 1 cm. to the right of the left nipple; the other wound was of the same length, 2 cm below and 4 to the left of the same nipple. The stab causing this latter wound was within the precordial area, and though it did not affect the heart, it and the former stab perforated the thoracic wall, producing pleural and small pneumothoracic bleeding. Given the unexpected nature of the attack, S.E.L. was unable to defend himself. His annoyance having subsided, the accused left the scene. I thus consider the facts as constituting *asesinato frustrado*, with the mitigating circumstance of mental derangement. The recommended sentence is of 18 years' imprisonment".

For the private prosecution, "... the accused F.M.C., at his home, which is also that of S.E.L, after a conversation that made him angry, suddenly took a knife out of his pocket and stabbed S.E.L. twice, producing the wounds described in the report. Thus, the facts constitute *asesinato frustrado*. We consider that there is an mitigating circumstance of mental derangement. The recommended sentence is of 18 years' imprisonment".

For the defence, "... the accused F.M.C. has suffered from mental disorders for several years, having been in psychiatric clinics such as B. On day D, after a heated discussion with his uncle S.E.L., he became so angry that he jumped on him and stabbed him twice, producing the injuries described in the report. This act was committed without any intention of killing S.E.L., being simply the result of the loss of temper after the argument, and due to the mental disorders mentioned previously. Consequently, we consider the events as constituting an offence of actual bodily harm. There is an extenuating circumstance of mental derangement. The recommended sentence is of 18 months' imprisonment".

#### ***Analysis of content of the protocols (deliberations)***

The deliberations were subjected to an analysis of content, taking as the unit of analysis each individual verbal intervention.

These verbal messages were categorized on the basis of speaker and content. There were six coding categories of

the messages: identifying the speaker (category 1); defining the content of the deliberation, or the argumentation about facts (category 2); defining the content of the legal argumentation (category 3); mention of the reliability of the testimonies (category 4); mention of the validity of the different evidence (category 5); and the valence of each intervention, that is, the classification as positive, negative or neutral for the accused (category 6). These categories were taken, for the study of the content, from Hastie, Penrod and Pennington (1986) and Fariña, Arce and Vila (1999); and for the study of the valence of the interventions, from Tanford and Penrod (1990). Through the assessment of the reliability of the different testimonies and of the validity of the evidence, we measured the dimensions "reliability" and "validity", in the "Integration of Information Models" proposed as a valid reference for explanation of the formation of individual legal judgements (e.g., Ostrom, Werner & Saks, 1978), by both judges (Arce et al., 2001) and juries (Arce, Fariña & Real, 2000); however, these have never been studied from a group discussion perspective.

To establish the content categories, on the basis of the systems previously described, we proceeded with a method of successive approaches after the study of the protocols. For the final fixing of the categories we followed the norms drawn up by Anguera (1990). Thus, we created a categorial system that was mutually exclusive, reliable and valid, in the context Weick (1985) refers to as that of systems of methodical categories. The resulting categories were as follows:

- *Category 1.* Identification of the subject who intervenes. For deliberations of the panels of judges, 1, 2 and 3, and for those of the juries, from 1 to 9.
- *Category 2.* Argumentation about the facts:
  01. No reference made to the facts of the case.
  02. Facts that occurred.
  03. Assumption of facts.
  04. Explanation/Justification of the cause.
  05. Personal history of the accused.
  06. Personal anecdotes.
  07. Intervention of the public prosecutor.
  08. Intervention of the defence counsel.
  09. Intervention of private prosecution.
  10. Intervention of forensic personnel/technical experts.
  11. Intervention of the accused.

12. Intervention of the victim.
13. Intervention of the defence witnesses.
14. Intervention of the prosecution witnesses.

- *Category 3. Legal argumentation:*

00. No reference made to law.
01. Legal detail.
02. Extenuating circumstances.
03. Aggravating circumstances.
04. Mitigation.
05. Recommendation for sentence.
06. Time of imprisonment.
07. Verdict intention.
08. Definition of the offence.
09. Formulation of conjectures.
10. Weight of evidence and *In dubio pro reo*.
11. Criminal responsibility.
12. Other cases.
13. Criminal danger.

- *Category 4. Assessment of the reliability of evidence.*

01. Credibility of the accused.
02. Credibility of the victim.
03. Credibility of forensic personnel/technical experts.
04. Credibility of defence witnesses.
05. Credibility of prosecution witnesses.

- *Category 5. Assessment of value of the evidence (all evidence was recoded to give universality for the three cases used):*

01. No reference made to the evidence of the case.
02. Arms.
03. Attributions to mental state of the accused.
04. Importance of the accused's testimony.
05. Importance of the victim's testimony.

06. Importance of technical experts' testimony.
07. Importance of the psychologist's testimony.
08. Importance of defence witnesses' testimony.
09. Importance of prosecution witnesses' testimony.
10. Importance of technical experts' reports.

- *Category 6. Valence of the intervention (direction of the intervention for the accused):*

01. Positive: pro-accused.
02. Negative: contra-accused.
03. Neutral.

### **Reliability**

The deliberations were coded by a thoroughly trained coder. Furthermore, a second coder assessed as reliable in other codings (Jólluskin, 2000) evaluated part of the material so as to check the reliability of the main coder. With these two coders, we computed two modes of reliability: one intra-coder and the other inter-coder, for 10% of the deliberations of each sample group (juries and panels of judges). With a view to obtaining intra-coder consistency, we left a period of one month between the original coding and the re-coding. For the calculation we used Cohen's Kappa, which has a correcting index for hits occurring at random.

The values obtained show that both inter- and intra-coder consistency of the different measures are good (see Table 1).

Moreover, as regards establishing the reliability beyond the instruments used, it is important to stress that these have shown themselves to be reliable, effective and valid in other studies, as well as consistent with other methods (e.g., Arce et al., 1996; Arce et al., 2001, Jólluskin, 2000). Thus, considering this inter-subject, inter-study and inter-method consistency, it can be stated that the values obtained are reliable (Wicker, 1975).

### **Data analysis**

Analyses of the deliberations were made through the assignment of categories to each intervention of the jurors/judges. Thus, each category is converted into a variable. A key question concerns fixing to these variables a condition of discrete variable (i.e., frequencies) or continuous variable. Traditionally, they have been considered as continuous variables (i.e., Hastie, Penrod & Pennington, 1986), presumably because of the greater

**Table 1**  
**Reliability of the coding. Kappa values**

Category	Intra-Coder	Inter-Coder
Subject	1.00*	1.00*
Facts	0.95*	0.78*
Law	0.90*	0.65*
Valence	1.00*	0.72*
Evidence (reliability and validity)	0.80*	0.65*

Note: \*p<.001. Kappa values are the average for the categories making up that category. Moreover, the values for each of the categories in isolation were significant.

power of the dataanalysis. For our specific purposes, it is also advisable to consider these variables as continuous. Consequently, we have taken a precautionary measure that guarantees their status as continuous variables: their transformation into continuous variables through the square root method. Thus, the variances are homogenized, stabilizing at approximately  $\sigma^2=1$  if the mean of the original observations is  $>.8$ , as in our case (Dixon & Massey, 1983, p. 373). Nevertheless, the means presented in the tables correspond to the raw scores, in order to provide an immediate idea of the numerical impact of the variables. As regards multivariate tests, we opted for the Pillai-Bartlett Trace, since it is more robust in the context of the heterogeneity of the variance matrices (Olson, 1976).

Furthermore, we computed a cognitive construct that indicates the rigidity in the content of the discussion: redundancy (e.g., Arce et al., 1999; Fariña, Arce & Vila, 1999). The significance of this construct resides in the fact that a redundant debate in terms of content has less value than a more balanced one, given the greater depth and fairness of the deliberation. This is obtained via the following formula: redundancy =  $\Sigma [f_o - f_e]$ .

With regard to studying of consistency of the judges' and juries' verdicts, we made a comparison of the highest percentages, with a given value, by means of the transformation into "z-scores". As given value with which to compare the empirical value we took .80 (that is, 80%), a point Tversky (1977) considers as the one after which it can be said that the judgements are concordant.

## RESULTS AND CONCLUSIONS

### *Study of the group verdict*

In Case I, the panels of judges all reached the unanimous decision of "*asesinato frustrado* with partially mitigating circumstances". That is, concordance on the decision was 100% and there is correspondence with the demands of the public prosecutor. For their part, the juries reached this same unanimous decision, "*asesinato frustrado* with partially mitigating circumstances", in two deliberations (40% of the juries),  $Z(n=45)=6.67$ ;  $p<.001$ , while in another two cases the decision was "guilty of *asesinato frustrado* but mentally ill" (40% of the juries, defence's request),  $Z(n=45)=6.67$ ;  $p<.001$ ; finally, one jury reached the most punitive decision, "murder" (20%, private prosecution's request). Consequently, the juries' deliberation, far from leading the decision towards a homoge-

nization of criteria – as occurs with the traditional system of decision in the framework of an adversarial procedure demanding a verdict of guilty or not guilty (Fariña, Arce & Vila, 1999) –, led it towards a disparity of criteria. In sum, the deliberation led the juries to a performance that was internally inconsistent.

In the case against J.C.O. (Case II), the panels of judges reached the decision of "threatening behaviour", the defence's recommendation, in three deliberations (60%),  $Z(n=15)=18.18$ ;  $p<.001$ ; in one, the decision reached was "*homicidio en grado de tentativa parcial*", the public prosecutor's request, (20%); and in another, it was "*homicidio en grado de tentativa completa*", the recommendation of the private prosecution, (20%). The dispersion is even greater in the case of the juries. Thus, two reached a decision of "threatening behaviour" (40%),  $Z(n=45)=6.67$ ;  $p<.001$ , another opted for "*homicidio en grado de tentativa parcial*" (20%), another one for "*homicidio en grado de tentativa completa*" (20%), and one was undecided (20%). The situation is, therefore, one of inconsistent decisions inter-group, among both judges and juries.

In Case III, against F.M.C., three of the panels of judges reached a verdict of "*asesinato frustrado* with mental derangement" (60%, public prosecutor's request)  $Z(n=15)=18.18$ ;  $p<.001$ ; one decided on a sentence of "actual bodily harm with mental derangement" (20%, defence's recommendation); and another reached a decision of "*asesinato frustrado*" (20%, request of the private prosecution). The juries, on the other hand, reached a verdict of "*asesinato frustrado* with mental derangement" in two cases (40%, public prosecutor's request)  $Z(n=45)=6.67$ ;  $p<.001$ ; another two decided on "actual bodily harm with mental derangement" (40%, defence's request); and one gave a verdict of "not guilty", (20%, defence's request). Once again, we find indications of decisional inconsistency, applicable, in this case, to both judges and juries.

Considering the three cases together, we find that the decisions of the panels of judges are in line with those proposed by the public prosecutor, 60%,  $\chi^2(2,n=45)=15.6$ ;  $p<.001$ , while those of the juries tend more towards the recommendations of the defence, 50%,  $\chi^2(2,n=126)=24.43$ ;  $p<.001$ .

### *Length of the discussion and argumentational ability*

There are two forms of understanding the length of a deliberation, referring either to time employed or to number

of interventions. We opt for the latter interpretation, as it more faithfully reflects the argumentational activity carried out. The results show that the juries' ( $M=186.133$ ) deliberations included more interventions than those of the panels of judges ( $M=71.2$ ),  $F(1,28)=10.066$ ;  $SM=99072.533$ ;  $p<.001$ ;  $T.E.=.264$ . However, this does not mean that there is greater argumentational ability, since in the context of judicial decisions, the larger the group, the longer the discussion (Fariña, Arce & Vila, 1999). Indeed, we found the interventions of the jurors ( $M=143.67$ ) to be more unrelated to the facts of the case than those of the judges ( $M=58$ ),  $F(1,28)=9.034$ ;  $p<.01$ ;  $SM=55040.833$ . Likewise, as regards the relation of interventions to legal aspects of the case, the juries ( $M=133$ ) present more interventions unrelated to legal ques-

tions than the panels of judges ( $M=42.47$ ),  $F(1,28)=10.571$ ;  $p<.01$ ;  $SM=61472.133$ ;  $E.S.=.274$ . Finally, the results in relation to connections between interventions and evidence show that the jurors ( $M=164$ ) make more interventions unrelated to the assessment of the reliability of evidence than the judges ( $M=60.33$ ),  $F(1,28)=11.4121$ ,  $p<.01$ ;  $SM=80600.833$ ;  $E.S.=.29$ . In sum, while the jurors make more interventions in the course of the deliberation, these are emptier of content than those of the judges.

All of this leads us to ask ourselves whether the fact that the deliberations of the juries are longer truly implies that they are more productive. It may well be that they focus excessively on some categories of discussion to the detriment of others, that is, they are redundant. The results in this direction (see Table 2) suggest that the deliberations of the juries are more redundant in relation to both the facts of the case and the legal aspects or references to evidence; in sum, they do not examine the content equally, but rather focus on certain content to the detriment of other. Consequently, in this particular case, greater length is not synonymous with greater quality in terms of even-handed study of the legal factors (e.g., facts, evidence and legal relationships). In brief, the juries appear to focus the discussion too narrowly, ignoring important content.

### *Argumentation about the facts*

The results in relation to the facts of the case indicate that the discussions of judges and juries are comparable in this aspect,  $F_{\text{multivariate}}(13,16)=1.497$ ;  $ns$ ;  $E.S.=.546$ . Nevertheless, examination of the univariate effects (see Table 3) shows that there are significant differences in the variables "reference to the facts that occurred" and "assumptions of facts". In either case, the juries show higher values than the judges. The implications are different for each category of analysis. Thus, while the juries enter into a discussion on the legal, factual evidence via the mention of "the facts that occurred", they also open the way to more discussion of the extra-legal evidence through the "assumptions of facts". Thus, they stray considerably more than the panels of judges from the object of the deliberation through excessive concern with extra-legal evidence.

### *Legal argumentation*

In the study of the legal argumentation, we found once

**Table 2**  
ANOVAS for the different measures of redundancy

Dependent variable	$M_{\text{judges}}$	$M_{\text{juries}}$	SM	F	P
Redundancy in facts	96.61	234.87	143374.652	9.832	.004
Redundancy in the law	103.42	250.26	161726.639	6.837	.014
Redundancy in evidence	109.01	284.68	231435.563	10.484	.003

Note: D.F. (1,28)

**Table 3**  
Tests of inter-subjects effects in facts of the case

Dependent variable	$M_{\text{judges}}$	$M_{\text{juries}}$	MC	F	p	eta <sup>2</sup>
Facts occurred	12.067	31.733	2900.833	9.496	.005	.253
Assumption of facts	.6	7.067	313.633	8.489	.007	.233
Justification of the cause	.1467	1.333	.133	.028	.868	.001
Personal history of the accused	.333	1	3.333	1.687	.205	.057
Personal anecdotes	.267	2.333	32.033	3.613	.068	.114
Intervention of the public prosecutor	.933	2	8.533	.411	.527	.014
Intervention of the defence counsel	.8	.267	2.133	1.792	.191	.060
Intervention of private prosecution	.333	.0667	.533	1.455	.238	.049
Intervention of forensic personnel/technical experts	4.4	1.867	48.133	2.921	.098	.094
Intervention of the accused	1.467	1.933	1.633	.355	.556	.013
Intervention of the victim	2.2	1.467	4.033	.360	.554	.013
Intervention of the defence witnesses	.333	.933	2.700	.698	.410	.024
Intervention of the prosecution witnesses	.467	.000	1.633	2.579	.120	.084

Note: D.F. (1,28)

more that judges and juries are comparable, as regards the legal resources associated with the interventions,  $F_{\text{multivariate}}(13,16) = 1.362$ ; ns; E.S. = .525.

However, the univariate effects, which can be seen in Table 4, show the jurors' interventions to be more oriented to "the intention of the verdict", to the "formulation of legal conjecture", and to "analysis of the penal consequences". In sum, juries show greater orientation to the verdict; to making assumptions or deductions of a legal nature that require specific legal knowledge they do not have; and to drawing legal inferences about the type of sentence that go beyond their brief according to Spain's judicial regulations (e.g., *Ley Orgánica 5/1995 de Tribunal del Jurado*). Thus, the juries' deliberations show an orientation toward the decision that is in contraposition (given the nature of discussion dynamics) to the "orientation toward the evidence" that would allow them to reach decisions of higher quality (Hastie, Penrod & Pennington, 1986).

#### Discussion on the reliability of the testimonies

The results show that the dimension "reliability of the evidence" also has a similar presence in deliberations of the panels of judges and the juries,  $F_{\text{multivariate}}(4,25) =$

Table 4 Tests of inter-subjects effects in references to legal aspects						
Dependent variable	M <sub>judges</sub>	M <sub>juries</sub>	SM	F	p	eta <sup>2</sup>
Legal detail	1.733	1.267	3.333	.623	.437	.022
Extenuating circumstances	1	3	30.000	2.877	.101	.093
Aggravating circumstances	1.333	1	.833	.165	.688	.006
Mitigation	1.067	3	28.033	2.876	.101	.093
Recommendation for sentence	.333	.267	3.333E-02	.076	.785	.003
Penal consequences	2.533	10.333	456.300	6.575	.016	.190
Verdict intention	4.333	9.733	218.700	4.535	.042	.139
Definition of the offence	1.333	1	.833	.165	.688	.006
Formulation of conjectures	.0667	1.933	26.133	5.308	.029	.159
In dubio pro reo	.0667	.467	1.200	3.877	.059	.122
Criminal responsibility	1	2.2	10.800	.730	.400	.025
Other cases/jurisprudence	1.667	.667	7.500	1.864	.183	.062
Criminal danger	.267	1.2	6.533	2.564	.121	.084
Note: D.F. (1,28)						

.429; ns; E.S. = .137. Likewise, the univariate effects indicate that none of the variables measuring reliability is involved to a different extent in the deliberations of judges and juries (see Table 5).

Table 5 Tests of the inter-subjects effects in the reliability/credibility dimension						
Dependent variable	M <sub>judges</sub>	M <sub>juries</sub>	SM	F	p	eta <sup>2</sup>
Credibility of the accused	1.33	.4	.533	.700	.410	.024
Credibility of the victim	.000	.733	4.033	3.233	.083	.104
Credibility of technical experts	.000	.267	.533	1.672	.207	.056
Credibility of defence witnesses	.000	.467	1.633	1.000	.326	.034
Credibility of prosecution witnesses	.000	.0667	3.333E-02	1.000	.326	.034
Note: D.F. (1,28)						

Table 6 Tests of the inter-subjects effects in the discussion on the value of the evidence						
Dependent variable	M <sub>judges</sub>	M <sub>juries</sub>	SM	F	p	eta <sup>2</sup>
Arms	.2	2.6	43.200	5.449	.027	.163
Mental state of the accused	2.933	11.533	554.700	4.023	.055	.126
Accused's testimony, validity	1.067	2	6.533	1.513	.229	.051
Victim's testimony, validity	2.267	1.2	8.533	.901	.351	.031
Technical experts' testimony, validity	2.533	2.067	1.633	.106	.747	.004
Psychologist's testimony, validity	.267	.2	3.333E-02	.054	.818	.002
Defence witnesses' testimony, validity	.0667	.2	.133	.400	.532	.014
Prosecution witnesses' testimony, validity	.0667	.000	3.333E-02	1.000	.326	.034
Technical experts' reports, validity	.467	.0667	1.200	1.260	.271	.043
Note: D.F. (1,28)						

Table 7 Valence of judges' interventions in deliberation			
Category	N observed	N expected	Residual
Pro-accused	316	359.3	-43.3
Against accused	556	359.3	196.7
Neutral	206	359.3	-153.3



### ***Discussion on the validity of the evidence***

We also found no significant differences between judges and juries with regard to the content of the discussions on the value of the different evidence presented at the trial,  $F_{\text{multivariate}}(9,20) = 1.669$ ; ns; E.S. = .568. That is, judges and juries perform similarly in weighing up the evidence for the decision that emerges from the group discussion. According to the data on univariate effects (see Table 6), the juries make more references to “arms described in the account of the facts” as evidence. In short, the results, as regards the relationship between the interventions in group discussion and the validity of the evidence, point to minor differences, since just one item of evidence differentiates them. This greater presence of references to arms may be due to the stress-inducing role of arms for lay subjects, given that, in line with the findings of eyewitness studies, arms cause the situation to produce anxiety, which in turn leads to poorer performance (e.g., Johnson & Scott, 1976; Tooley et al., 1987; Maas & Köhnken, 1989).

### ***Study of the valence of the intervention***

Taken in isolation, the panels of judges (contingencies in Table 7) present a greater tendency to go against the accused in their interventions,  $\chi^2(2, n = 1078) = 178.293$ ;  $p < .001$ , and less tendency to be favourable or neutral to him. In other words, the deliberations are oriented against the defendant.

For their part, the jurors (see contingencies in Table 7) show a tendency in their interventions greater than the expected one both in favour of and against the accused, to the detriment of neutral interventions: that is, they “orient the decision”,  $\chi^2(2, n = 2769) = 331.033$ ;  $p < .001$ .

In terms of strategy, the juries follow one of “integration of the information”, rather than one of exclusion, as would be reflected if the information were pushed, either in favour of or against the accused. In contrast, the judges use a strategy of “exclusion” of information favourable to the accused; that is, there is an orientation towards guilt.

The implications and significance of the results presented here are limited. Thus, they cannot be generalized to other legal contexts because the way Spanish juries work is not directly comparable to that of others for which information is available. Moreover, and circumscribing ourselves to the Spanish case, the results should be treated with some caution, since, while it is true that the simu-

lation in the present study is a highly faithful one, it is also the case that these types of results are frequently different from those obtained in a real context or one based on archive cases (e.g., Konecni & Ebbensen, 1992). It is therefore necessary to proceed with further research in this line in order to delimit more precisely the significance of the implications emerging from the present study. Bearing in mind this need for caution, we can formulate the following conclusions based on the results presented here. First, judicial decisions, both of panels of judges and of juries, fall short in relation to a central objective, that of inter-rater consistency. This conclusion applies especially to the case of juries. Thus, the deliberations fail to fulfil the purpose assigned: control of biases through the homogenization of judgements (Kaplan & Miller, 1978). Furthermore, it should be borne in mind that enshrined in our judicial system is the principal of free assessment of the evidence, which excludes any type of subjection to any criterion other than that of *the panel of judges or jury*. Second, in relation to judgement tendencies, the panels of judges show, in accordance with the results of other studies using Spanish archive sentences, an effect of decisional anchoring of the sentences in the recommendations of the public prosecutor (Garrido & Herrero, 1995; Arce, Fariña & Novo, 1996; Fitzmaurice & Pease, 1986; Fariña, Novo & Arce, 2002). In contrast, the juries tend to lean towards the recommendations of the defence counsel. In sum, while the panels of judges are oriented towards “the demands of the prosecution”, the juries are oriented towards “the demands of the defence”. This constitutes one more demonstration (e.g., MacCoun & Kerr, 1988) of the tendency for bias in lay subjects towards the thesis of the defence and in professionals towards that of the prosecution. Third, the deliberations of juries do not involve a more profound discussion of the content by way of compensation for their lack of experience, knowledge and training by comparison with judges. Thus, though juries’ deliberations are longer, they are not more “content-oriented”. Surprisingly, juries’ deliberations are emptier of content. Fourth, nor do juries make a deeper analysis of the reliability of testimonies or the value of the evidence provided. This result is also surprising, since we would have expected more thorough discussion in compensation for their lesser ability. Fifth, judges’ deliberations do not reflect (as we would expect in view of their training) more grasp and use of legal notions. Even so, the data may be misleading, since judges’ legal knowledge may be impli-

cit (e.g., references to jurisprudence in themselves suggest an understanding of all the legal content involved in the case); moreover, juries' references are not strictly "legal", but based rather on "legal conjecture", associations with "intention of the verdict" and "assumptions about the penal consequences", none of which constitute their brief according to the law. Sixth, in line with group decision theory, the valence of the intervention shows that judges' deliberations are more "oriented against the accused", while those of jurors appear to follow a process of "integration of the favourable and unfavourable versions for the accused". This result is indicative of the fact that differences between the decisions of judges and juries not only relate to the verdict (e.g., Tanford & Penrod, 1990; Arce, 1995), but rather go beyond it, affecting assessment of the evidence, to the detriment of the defendant in the case of judges. Furthermore, and in line with the findings of Tanford and Penrod (1990), the interventions of both judges and jurors are directly linked to the verdict through the favourable or unfavourable valence towards the accused. That is, they are "task-oriented". Seventh, as regards cognitive activity, the deliberations of lay subjects show greater activity, but this is not as effective as it might be given that it is more redundant – that is, they do not discuss the facts, evidence and legal arguments so homogeneously as the judges, but rather place excessive emphasis on certain categories to the detriment of others. And finally, eighth, considering all of the above, we can conclude that juries and judges perform different tasks.

In sum, we have found considerable indication of inadequate performance in judges and juries, though more in the latter case, and where measures such as training about the deficits presented would help to improve the situation. Nevertheless, there is a need for ad hoc studies with a view to finding the most appropriate solutions. In the meantime, we propose the training of decision-makers in relation to the sources of bias so that they can confront the metacognitive deficit which has been identified as a source of anomalies in informal reasoning (Perkins, 1989). Thus, our belief is that a first step for improving the non-normative performance of these samples is to point out the problem to the actors involved so that they can self-correct. As regards professional practice, in view of the fact that the involvement of a jury is optional, it would be more favourable to the interests of the defence, while panels of judges would be more favourable for the prosecution. Likewise, strategies of

persuasion are not the same for lay subjects and professionals in that they focus on different items of evidence.

## REFERENCES

- Anguera, M<sup>a</sup> T. (1990). Metodología observacional. In J. Arnau, M<sup>a</sup>. T. Anguera, & J. Gómez (Eds.), *Metodología de la investigación en las ciencias del comportamiento* (pp. 125-236). Murcia: Servicio de Publicaciones de la Universidad de Murcia.
- Arce, R. (1995). Evaluating evidence on jury decision making. In R. Bull & D. Carson (Eds.), *Handbook of psychology in legal context* (pp. 565-580). London: John Wiley and Sons.
- Arce, R., Fariña, F., Novo, M. & Real, S. (1996). Cognition and judicial decision making. *International Journal of Psychology*, 31(3-4), 93.
- Arce, R., Fariña, F., Novo, M. & Seijo, D. (1999). In search of causes of hung juries. *Expert Evidence*, 6, 243-260.
- Arce, R., Fariña, F., Novo, M. & Seijo, D. (2001). Judges' decision-making from within. In R. Roesch, R. R. Corrado, & R. J. Dempster (Eds.), *Psychology in the courts: International advances in knowledge* (pp.195-206). New York: Routledge.
- Arce, R., Fariña, F. & Real, S. (2000). The assessment of information integration theory and confirmatory bias hypothesis on judicial proceedings. A case of rape and murder. In A. Czerederecka, T. Jaskiewicz-Obydzinska, & J. Wójcikiewicz (Eds.), *Forensic psychology and law* (pp. 296-303). Krakow: Institute of Forensic Research Publishers.
- Baldwin, J. & McConville, M. (1979). *Jury trials*. London: Oxford University Press.
- Chin, A. & Peterson, M.A. (1985). *Deep pockets, empty pockets: Who wins in Cook Country jury trials?* Santa Monica, CA.: The Rand Corporation.
- Diamond, S.S. & Zeisel, H. (1975). Sentencing council: A study of sentence disparity and its reduction. *University of Chicago Law Review*, 43, 109-149.
- Dixon, W.J. & Massey, F.J. Mr. (1983) (4<sup>a</sup> ed.). *Introduction to statistical analysis*. New York: MacGraw-Hill.
- Konecni, V.J. & Ebbensen, E.B. (1992). Methodological issues on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law*.

- International perspectives* (pp. 413-423). Berlín: Walter de Gruyter.
- Elwork, A., Sales, B.D. & Alfini, J.J. (1977). Juridic decisions: In ignorance of the law or in light of it? *Law and Human Behavior*, 1, 163-189.
- Fariña, F., Arce, R. & Vila, C. (1999). Efectos de la composición del jurado en los estilos deliberativos, análisis cognitivo y de contenido de la deliberación. *Anuario de Psicología*, 30(1), 105-122.
- Fariña, F., Novo, M. & Arce, R. (2002). Heurístico de anclaje en las decisiones judiciales. *Psicothema*, 14(1), 39-46.
- Fitzmaurice, C. & Pease, K. (1986). *The psychology of judicial sentencing*. Manchester: Manchester University Press.
- Garrido, E. & Herrero, C. (1995). Fiscales y jueces: Concordancia entre percepción y realidad. In E. Garrido & C. Herrero (Comps.), *Psicología política, jurídica y ambiental* (pp. 269-281). Salamanca: Eudema.
- Hastie, R., Penrod, S.D. & Pennington, N. (1986). *La institución del jurado en los Estados Unidos. Sus intimitades*. Madrid: Cívitas (Orig. 1983).
- Heuer, L. & Penrod, S.D. (1994). Trial complexity. A field investigation of its meaning and effects. *Law and Human Behavior*, 18(1), 29-51.
- Johnson, C. & Scott, B. (1976, September). *Eyewitness testimony and suspect identification as a function of arousal, sex of witness, and scheduling of interrogation*. Paper presented at 18th Annual Convention of the American Psychological Association, Washington, D.C.
- Jólluskin, G.S. (2000). *Influencia social mayoritaria y minoritaria, y decisiones de jurados*. Doctoral dissertation, Universidad de Santiago de Compostela.
- Kalven, H. Jr. & Zeisel, H. (1966). *The American jury*. Boston: Brown.
- Kaplan, M.F. & Miller J. H. (1978). Reducing the effects of juror bias. *Journal of Personality and Social Psychology*, 36, 1.443-1.455.
- Kapardis, A. (1985). *Sentencing by English magistrates as a human process*. Nicosia, Cyprus: Asselia Press.
- Kapardis, A. (1997). *Psychology and law: A critical introduction*. Cambridge: Cambridge University Press.
- Maas, A. & Köhnken, G. (1989). Eye-witness identification: Simulating the «weapon effect». *Law and Human Behavior*, 13, 397-408.
- MacCoun, R.J. & Kerr, N.L. (1988). Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology*, 54(1), 21-33.
- Myers, M.A. (1979). Rule departures and making law: Juries and their verdicts. *Law and Society Review*, 13, 781-797.
- Olson, C.L. (1976). On choosing a test statistic in manova. *Psychological Bulletin*, 83, 579-586.
- Ostrom, T.M., Werner, C. & Saks, M.J. (1978). An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology*, 36, 436-450.
- Partdridge, A. & Eldridge, W. (1974). *The second circuit sentencing study*. New York: Federal Judicial Center.
- Perkins, D.N. (1989). Reasoning as it and could be: An empirical perspective. In D. M. Topping, D. C. Crowell & V. N. Kobayaski (Eds.), *Thinking across cultures: The third international conference on thinking*. Hillsdale, NJ: LEA.
- Sobral, J. & Prieto, A. (1994). *Psicología y ley. Un examen de las decisiones judiciales*. Madrid: Eudema.
- Sutton, L. (1978). *Variations in federal criminal sentences: A statistical assessment at the national level*. Washington, D.C.: National Criminal Justice Information and Statistics Service.
- Tanford, S. & Penrod, S. (1990). Las deliberaciones del Jurado: Contenido de la discusión y procesos de influencia en la toma de decisiones de un Jurado. *Estudios de Psicología*, 42, 75-97 (Orig. 1986).
- Tooley, V., Brigham, J.C., Maas, A. & Bothwell, R.K. (1987). Facial recognition: Weapon effect and attentional focus. *Journal of Applied Social Psychology*, 17, 845-859.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Weick, K.E. (1985). Systematic observational methods. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (vol. 1) (pp.567-634). Hillsdale, N.J.: LEA.
- Wicker, A.W. (1975). An application of a multiple-trait-multimethod logic to the reliability of observational records. *Personality and Social Psychology Bulletin*, 4, 575-579.