

Validez de la Escala de Sesgo del Jurado en relación con las decisiones grupales de Jurados simulados

Leticia de la Fuente Sánchez, Juan García García y E. Inmaculada de la Fuente Solana*
Universidad de Almería y * Universidad de Granada

El objetivo fundamental de este estudio es evaluar la validez de la Escala de Sesgo del Jurado, adaptación española de la *Juror Bias Scale* (JBS, Kassin y Wrightsman, 1983), en relación con su capacidad para predecir las decisiones grupales (veredictos y preguntas relativas al objeto del veredicto) dadas por diferentes grupos de Jurados formados por sujetos clasificados previamente por la escala como «sesgados hacia la defensa» o «sesgados hacia la acusación». Se utilizan para ello simulaciones judiciales en las que se manipula la fuerza de la evidencia y se mantiene constante el tipo de caso y la forma de presentación de los casos. Para el análisis de los datos se utilizan métodos de inferencia exacta, mediante el cómputo de valores *p* exactos. Los resultados obtenidos indicaron que, bajo la condición de evidencia ambigua, la Escala de Sesgo del Jurado predice con éxito los veredictos grupales de los Jurados proacusación y prodefensa, así como las preguntas relativas al objeto del veredicto de carácter opinable. Las implicaciones de estos resultados se discuten en relación con la relevancia de los sesgos individuales en las decisiones grupales y la utilidad de este tipo de instrumentos de medida en la fase de selección de un Jurado.

Validity of the Juror Bias Scale in relation to mock jury decisions. In the present study, the validity of the Spanish adaptation of Juror Bias Scale (JBS, Kassin & Wrightsman, 1983), is tested in relation to its capability to predict the group decisions (verdicts and verdict's object questions) reached by different groups of juries composed by subjects previously classified by the JBS as «defense-biased» or «prosecution-biased». Judicial simulations are used to develop the study, in which the strength of evidence is manipulated and the type of case and the presentation format are kept constant. To analyze data exact inference methods are used. The results indicated that, under ambiguity condition, the JBS scores successfully predict the group verdicts of both defense-biased and prosecution-biased juries, as well as the verdict's object questions with «opinion character». Implications of these results are discussed in relation to the importance of individual biases in group decisions and the utility of this type of scales in jury selection.

Uno de los principales retos que se ha planteado la investigación psicológica en el ámbito de la toma de decisiones judiciales ha sido evaluar si características personales de los miembros de un Jurado tales como actitudes, valores, creencias o características de personalidad pueden llegar a ser factores con suficiente relevancia como para influir de forma importante los juicios y decisiones que las personas que actúan como jurados tienen que tomar durante un proceso judicial (en adelante, indicaremos al grupo con mayúscula inicial –Jurado–, y a los miembros en minúscula –jurado–). Globalmente hablando y, desde una perspectiva aplicada, el objetivo que se persigue es la detección de aquellos individuos que, por sus características particulares, puedan llegar a tomar decisiones que estén más influidas por sus propias actitudes y creencias que por la evaluación objetiva de la evidencia (entendida ésta como con-

junto de pruebas) que se presenta en el juicio. Así, la detección de posibles *jurados sesgados* que puedan atentar contra el principio de imparcialidad es el objetivo de los procesos de selección de jurados que se llevan a cabo antes de comenzar los juicios.

Este planteamiento ha suscitado dos cuestiones importantes: ¿Cuáles son las características de un jurado sesgado?, ¿En qué circunstancias los veredictos de los Jurados pueden estar realmente influidos por los sesgos de los miembros que los forman?. Para intentar responder a la primera cuestión, los investigadores interesados en la relación entre características individuales y decisiones de los jurados han llevado a cabo multitud de estudios en los que se han considerado múltiples factores. Se han estudiado, fundamentalmente, variables sociodemográficas (sexo, edad, estado civil, nivel socioeconómico o profesión, entre otros) y variables psicosociales o actitudinales. Incluidas en estas últimas se han estudiado actitudes generales, comunes a diferentes campos de investigación (ideología política, religiosidad, locus de control, autoritarismo, dogmatismo, conservadurismo, ideas sobre un mundo justo, etc.) variables actitudinales adaptadas al ámbito legal (autoritarismo legal, propensión hacia la culpabilidad o la inocencia) y variables específicas a los casos (actitud hacia la pena de muerte, acti-

tud hacia la violación, hacia las drogas o actitud hacia los psiquiatras) (para una revisión puede consultarse, p.e., Hastie, Penrod y Pennigton, 1983; Devine, Clayton, Dunford, Seying y Price, 2001; García, De la Fuente y De la Fuente, en prensa).

En conjunto, los resultados de este tipo de estudios indican que la relación entre características individuales y decisiones de los jurados está mediatizada por factores situacionales tales como el tipo de caso que se va a juzgar y la fuerza de la evidencia presente en los juicios y, por tanto, no puede ser entendida ni estudiada de manera universal. De hecho, la capacidad predictiva de factores individuales como actitudes, características de personalidad, estilos atribucionales, ideología, etc., parece estar inversamente relacionada con la fuerza de la evidencia. De esta manera, los resultados de diversos estudios indican que la relevancia de los sesgos individuales aumenta a medida que la fuerza la evidencia disminuye (Baumeister y Darley, 1982; Dane y Wrightsman, 1982; Kaplan, 1989; MacCoun, 1990; García, 2000; De la Fuente, 2001; Devine et al., 2001). En cuanto a la mediación del tipo de caso, existen evidencias de que determinadas características individuales pueden ser una fuente relevante de sesgo en unos casos pero no en otros. Quizá uno de los ejemplos más claros en este sentido lo podemos encontrar en relación con la influencia del género de los sujetos. Los resultados de gran parte de los estudios indican que el género puede ser una fuente de sesgos en casos relacionados con agresiones sexuales, violencia doméstica o custodia de niños, existiendo un sesgo positivo hacia las mujeres en los casos de familia y negativo en agresiones domésticas (Arce, Fariña y Fraga, 2000).

Por otra parte, de este tipo de estudios también se puede concluir que no todos los factores individuales tienen la misma capacidad de influir sobre las tendencias de veredicto (Devine et al., 2001). En este sentido, parece importante tener en cuenta qué tipo de medidas es mejor utilizar para maximizar la capacidad predictiva de variables psicológicas tales como actitudes o características de personalidad. La investigación concerniente a los sesgos de los jurados indica que medidas que evalúan directamente actitudes específicas al sistema criminal de justicia, como el *Revised Legal Attitudes Questionnaire* (RLAQ; Kravtitz, Cutler y Brock, 1993), o que evalúan actitudes relevantes a cómo se toman las decisiones acerca de la culpabilidad, como la *Juror Bias Scale* (Kassin y Wrightsman, 1983), son mejores predictores de los veredictos que otras medidas de carácter más general (Kassin y Wrightsman, 1983; Narby, Cutler y Moran, 1993; García, 2000; Moran, 2000; Kovera, Dickinson y Cutler, en prensa).

Para intentar responder a la segunda cuestión, se ha estudiado bajo qué circunstancias los sesgos individuales pueden influir sobre las decisiones grupales a las que tiene que llegar el Jurado tras el proceso de deliberación. Diversos estudios han puesto de manifiesto que cuando los jurados tienen la posibilidad de discutir el caso en grupo, los factores o fuentes de sesgo individuales son contrarrestados durante el proceso de deliberación (Izzet y Leginski, 1974; Kerwin y Shaffer, 1994). Sin embargo, otros estudios han constatado que los sesgos de los jurados pueden mantenerse, o incluso generalizarse y aumentarse después de la deliberación, provocando decisiones grupales más extremas y favoreciendo la aparición del fenómeno de 'polarización grupal' (Bray y Noble, 1978; Cowan, Thompson y Ellsworth, 1984; Kaplan, 1989; Kramer, Kerr y Carroll, 1990; MacCoun, 1990; De la Fuente, 2001). En conjunto, estos estudios sugieren que jurados sesgados pueden generar decisiones grupales sesgadas cuando el grupo de jurados no es suficientemente heterogéneo. Es decir, si los individuos que

forman el Jurado presentan homogeneidad en relación con alguna característica común que puede sesgar sus decisiones en casos concretos, es más probable que la decisión grupal esté igualmente sesgada.

Por otra parte, creemos importante considerar que la influencia de los sesgos de los jurados se puede poner de manifiesto, además de sobre el veredicto propiamente dicho, sobre aspectos como la determinación de atenuantes, la interpretación de las pruebas periciales, el criterio individual de duda razonable, etc., todos ellos componentes indirectos del veredicto final. Sin embargo, la mayoría de los estudios se limitan a la relación entre características individuales y veredicto, lo que, desde nuestro punto de vista, puede estar limitando en gran medida la detección de posibles sesgos hacia componentes indirectos del veredicto. Además, en un Jurado de hechos como es el español, donde el Jurado debe justificar su veredicto, estos aspectos cobran especial relevancia. En el procedimiento que la Ley del Jurado estipula en España, los jurados deben decidir, previamente a la emisión del veredicto, sobre una serie de cuestiones relacionadas con éste encaminadas a que los jurados justifiquen y razonen su propio veredicto, lo que se conoce como 'objeto del veredicto', y se ha podido comprobar que la interpretación que los jurados hacen de cuestiones como alevosía, imputabilidad, existencia de delito, legítima defensa, etc., está mediatizada por factores psicosociales (De la Fuente, García y Martín, 1998; García, 2000).

Estamos de acuerdo con Myers y Lecci (1998) al afirmar que los resultados obtenidos hasta el momento sugieren la necesidad de seguir explorando las condiciones bajo las cuales las características individuales pueden convertirse en sesgos importantes que afecten las decisiones de los Jurados. Igualmente, resulta de especial interés poder avanzar en qué tipo de medidas se pueden utilizar con garantías científicas para la detección de jurados sesgados. Así, en este estudio nos hemos planteado un doble objetivo. Por una parte, analizar si la condición de 'homogeneidad' en los Jurados puede provocar decisiones grupales diferenciales, acotando el tipo de caso y teniendo en cuenta la fuerza de la evidencia. Para ello, se comparará la ejecución de Jurados compuestos por personas consideradas homogéneas en cuanto a tendencias de sesgo en dos casos reales de homicidio, uno con evidencia fuerte proacusatoria y otro con evidencia ambigua.

Para la medición de la tendencia de sesgo se utiliza la Escala de Sesgo del Jurado (ESJ, De la Fuente, 2001), adaptación española de la *Juror Bias Scale* (Kassin y Wrightsman, 1983), una de las escalas más frecuentemente utilizadas en el contexto anglosajón para la detección de jurados sesgados. Así, como segundo objetivo, directamente relacionado con el instrumento de medida y complementario al anterior, nos interesa conocer si la medida de sesgo que recoge la ESJ es capaz de discriminar las decisiones a las que llegan los diferentes Jurados, formados por sujetos clasificados previamente por la escala como 'sesgados hacia la defensa' o 'sesgados hacia la acusación'. Si esto es así, podremos aportar nuevas evidencias acerca de la validez de esta escala; concretamente, 'evidencias predictivas' que se encuadrarían en lo que la APA recoge en sus últimos *Standards* en relación con la Validez como 'evidencias basadas en relaciones con otros variables' (AERA, APA, NCME, 1999).

Desde este planteamiento, nuestro interés con respecto al proceso de deliberación se ha centrado, no tanto en el proceso de deliberación en sí mismo sino, más bien, en el resultado de dicho proceso, aunque también se han recogido algunas medidas relati-

vas a la deliberación que nos pueden aportar información complementaria, como el número de votaciones o el tiempo empleado en las deliberaciones.

En el estudio que presentamos los Jurados deciden, además del veredicto, sobre otra serie de preguntas que conforman el mencionado 'objeto del veredicto'. A la hora de redactar las preguntas nos guiamos fundamentalmente por el formato en el que se presenta el 'objeto del veredicto' en los juicios reales. Por esta razón, todas las preguntas comienzan con la expresión '¿Considera probado que...?'. Pero también se redactaron para poder poner a prueba una hipótesis que manejábamos respecto a la influencia del tipo de pregunta sobre las respuestas de los jurados. Esta hipótesis parte de las propuestas de Laughlin y Ellis (1986) acerca de la relación entre el tipo de cuestión sobre la que se está decidiendo, más opinable o más racional, y el tipo de influencia, normativa o informativa, que será más probable que aparezca en las dinámicas grupales. En esta línea, planteamos que el tipo de cuestión que se proponga puede favorecer la aparición y generalización de sesgos en los jurados, ya que en las preguntas 'opinables', donde los hechos a juzgar son lo suficientemente ambiguos como para poder basarse en ellos con seguridad, será más probable que los sujetos basen sus decisiones en otros factores, ajenos a las pruebas presentadas en el juicio.

En función de los objetivos propuestos, las hipótesis de las que partimos son las siguientes: a) Existirán diferencias en los veredictos grupales en función de la tendencia de sesgo de los Jurados en el caso de evidencia ambigua, no en el caso de evidencia fuerte proacusación, donde prevemos que todos los grupos emitirán un veredicto de culpabilidad. En el caso de evidencia ambigua, esperamos que los Jurados formados por sujetos clasificados por la ESJ como proacusación emitirán más veredictos de culpabilidad que los clasificados como prodefensa. b) Existirán igualmente diferencias en función de la tendencia de sesgo en las preguntas objeto del veredicto de carácter opinable. En concreto, esperamos encontrar que los Jurados homogéneos proacusación tenderán a elegir la opción que menos favorezca a los acusados, mientras que los prodefensa elegirán la opción más favorable. c) Por lo que respecta al tiempo empleado en las deliberaciones, esperamos encontrar diferencias entre los diferentes grupos de Jurados, de manera que serán los Jurados formados por sujetos sesgados hacia la acusación los que presentarán deliberaciones más cortas. Por último, en cuanto al número de votaciones necesarias para llegar a una mayoría cualificada, tanto en el veredicto como en el objeto del veredicto, no estamos en disposición de plantear hipótesis de carácter confirmatorio debido a la falta de datos de la que disponemos al respecto. Esta variable se introduce con carácter exploratorio.

Método

Participantes

En este estudio participaron 153 sujetos, estudiantes universitarios de diferentes titulaciones que cumplieran las condiciones exigidas para ser jurado en España. De éstos, 65 eran hombres y 88 mujeres, cuyas edades variaban entre los 18 y los 39 años, con una media de 23.03 y una desviación típica de 2.21. Estos participantes proceden de una muestra más amplia, compuesta por 321 estudiantes. Los 153 sujetos de este estudio fueron seleccionados de aquellos que expresaron voluntariamente, en una fase previa, su interés por participar en el presente estudio. La forma concreta de selección se describe en el apartado de procedimiento.

Material

Para la medición de la tendencia de sesgo se utilizó la Escala de Sesgo del Jurado, adaptación española de la *Juror Bias Scale*. Esta escala está fundamentada teóricamente en los modelos de decisión individual en jurados. En la mayoría de estos modelos se asume implícita o explícitamente que para emitir un veredicto es necesario realizar previamente las siguientes «operaciones» (Pennington y Hastie, 1981): (1) evaluar la probabilidad de que el acusado cometiera el delito (probabilidad de comisión); y (2) establecer un criterio individual de duda razonable, que la probabilidad de comisión estimada debe exceder para votar por la condena. Así, los juicios de culpabilidad aparecerían cuando la probabilidad de comisión estimada por un jurado excediera su criterio individual de duda razonable, mientras que el veredicto de inocencia se daría cuando la probabilidad de comisión estimada no llega a exceder dicho criterio de duda razonable. Con base en este análisis teórico, Kassir y Wrightsman (1983) elaboran la *Juror Bias Scale*, cuyo objetivo es detectar la tendencia generalizada de los sujetos a condenar o absolver a los individuos implicados en causas penales. Se asume, por tanto, que un jurado individual va a diferir del resto a lo largo de dos dimensiones: 'probabilidad de comisión generalizada' (PC), conceptualizada como la creencia de que las personas que son llevadas a juicio probablemente cometieron el delito del que se les acusa; y 'criterio individual de duda razonable generalizado' (DR), conceptualizado como la cantidad de evidencia en contra que una persona está dispuesta a asumir para condenar. De este modo, la combinación que cada persona hace de ambas dimensiones permitiría posicionarlos en un continuo en cuyos extremos estarían, por una parte, los sujetos con tendencias generalizadas hacia la condena (sesgados hacia la acusación) y, por otra, los sujetos con tendencias generalizadas hacia la absolución (sesgados hacia la defensa).

La ESJ está compuesta por 22 ítems con formato de respuesta tipo Likert de 5 puntos; diecisiete de los cuales son afirmaciones diseñadas para reflejar las dimensiones 'probabilidad de comisión generalizada' y 'criterio de duda razonable generalizado', y los cinco restantes son *filler*, relacionados con cuestiones legales generales e incluidos con el fin de desviar el propósito real de la escala. De los 17 ítems que componen las dimensiones PC y DR, doce de ellos están escritos de manera que una respuesta de acuerdo indicara sesgo hacia la acusación (p.e.: «De cada 100 personas llevadas a juicio, por lo menos 75 son culpables del crimen que se les acusa»), mientras que los cinco restantes están redactados de forma que una respuesta de acuerdo indica sesgo hacia la defensa (p.e., «Si la mayor parte de la evidencia, pero no toda, sugiere que el acusado cometió el delito, el Jurado debería votar NO culpable»).

Por otra parte, para realizar las simulaciones judiciales se utilizaron dos grabaciones de vídeo, siendo ambas elaboraciones correspondientes a la vista oral de dos juicios reales calificados de homicidio, celebrados en dos audiencias provinciales españolas. Los vídeos que se presentaron a los sujetos fueron editados a partir de los vídeos originales. Se eligió el tipo penal de homicidio por dos razones, porque es el delito más frecuentemente juzgado mediante tribunales de Jurados en España y porque ha sido el caso más utilizado en estudios previos con la *Juror Bias Scale*. Los juicios se eligieron debido a la diferencia existente en cuanto a la fuerza de la evidencia en contra de los acusados, diferencia que viene avalada, tanto por el criterio legal de los jueces que los juz-

garon en su momento, como por estudios anteriores (p.e. García, 2000), en los que dichos casos fueron evaluados por más de 500 personas de la población general que variaban en cuanto a edad, sexo y nivel educativo, entre otras variables. En el primer caso, denominado 'Caso 1: de evidencia fuerte proacusación', los acusados admitieron haber provocado los hechos que causaron la muerte de la víctima, entregándose voluntariamente a la policía. En el segundo, denominado 'Caso 2: de evidencia ambigua', el acusado, que se declara inocente, es arrestado como principal sospechoso, pero las circunstancias que rodean la muerte de la que se le acusa son lo suficientemente ambiguas como para permitir diversas interpretaciones.

Procedimiento

Los sujetos que participaron en este estudio fueron seleccionados de una muestra más amplia entre aquellos que expresaron voluntariamente, en una fase previa, su interés por participar en los juicios simulados. En dicha fase, realizada un mes antes de las simulaciones, se administró a la muestra global un cuadernillo que contenía la ESJ junto a otros cuestionarios. A partir de la puntuación que obtuvieron los sujetos en la ESJ se formaron dos grupos: sesgados hacia la acusación (proacusación) si su puntuación estaba por encima de la mediana de la escala; y sesgados hacia la defensa (prodefensa), si estaba por debajo (Med=52). Utilizamos para la clasificación de los sujetos el mismo procedimiento que siguieron los autores de la escala original (Kassin y Wrightsman, 1983). En dicho estudio, la mediana de la escala fue 51.50. Posteriormente, se fue llamando a los voluntarios para confirmar su participación en el presente estudio, quedando compuesta la muestra definitiva por 153 sujetos, de los cuales 81 quedaron clasificados como sujetos proacusación y 72 como prodefensa. Tanto los sujetos proacusación como los sujetos prodefensa fueron, entonces, asignados a grupos de 9 personas para formar los Jurados; quedando, de esta manera, 9 grupos de Jurados homogéneos proacusación y 8 grupos de Jurados homogéneos prodefensa. La asignación de los sujetos a los grupos se realizó intentado mantener una cierta variabilidad en cuanto a las titulaciones y al sexo de los sujetos, de manera que no se obtuvieran grupos de jurados en los que todos fueran hombres o todas mujeres, y que no fueran de la misma titulación. En la fase de simulaciones judiciales se realizaron 17 sesiones, una para cada grupo de jurados. En cada sesión los participantes eran instruidos acerca de la tarea que iban a desarrollar, junto a una explicación acerca del significado del concepto 'duda razonable', circunstancias agravantes y atenuantes, la regla de decisión que debían adoptar y la forma de votación. La regla de decisión que se adoptó fue la establecida por la Ley del Jurado en España en la que, tanto la votación de los hechos (estipulados en las preguntas relativas al objeto del veredicto), como la votación del veredicto, se rigen por las llamadas «mayorías cualificadas» (culpable y hechos desfavorables al acusado: un mínimo de 7 de los 9 votos; inocente y hechos favorables al acusado: un mínimo de 5 de los 9 votos). La secuencia de presentación de los casos fue la siguiente: la mitad de los grupos vieron en primer lugar el caso 1 y en segundo lugar, el caso 2; presentando la secuencia contraria a la otra mitad de grupos. No se limitó el tiempo de deliberación para ningún grupo. Cada grupo elegía a una persona que debía ejercer como portavoz del Jurado y encargarse de recoger en un cuadernillo elaborado al respecto las cuestiones que posteriormente se utilizarían como medidas objeto de análisis. Dichas me-

das fueron las siguientes: a) resultado de la decisión grupal en cada una de las preguntas 'objeto del veredicto', en cada caso, con un formato de respuesta dicotómico (si/no); b) número de votaciones necesarias para llegar a una mayoría cualificada, en cada una de las preguntas y en cada caso; c) veredicto grupal, en cada caso, con formato de respuesta dicotómico (culpable/inocente); d) número de votaciones necesarias para llegar a una mayoría cualificada en el veredicto, en cada caso; e) en el supuesto de que el Jurado emitiera un veredicto de culpabilidad en cualquiera de los dos casos, la consideración del delito: homicidio voluntario/homicidio involuntario; f) tiempo empleado en el proceso de deliberación, en cada caso.

Análisis de datos

Una manera de realizar inferencias estadísticas válidas en presencia de muestras pequeñas es mediante el cómputo de valores p exactos, basados en la distribución permutacional del test estadístico (Metha y Patel, 1999). Esta aproximación fue inicialmente propuesta por Fisher (1925) para tablas de contingencia 2×2 y modelos de probabilidad de clasificación fija. Desarrollos posteriores han ampliado y justificado el uso de estos procedimientos a otros modelos de probabilidad que son habitualmente utilizados para analizar tablas de contingencia, como el modelo de homogeneidad, así como a tablas de contingencia multidimensionales (Metha y Patel, 1999). Aunque los métodos de inferencia exacta a los que hacemos referencia han sido desarrollados inicialmente para tablas de contingencia con datos categóricos, pueden ser extendidos a muestras pequeñas con datos continuos. Por lo tanto, estos métodos proporcionan una aproximación unificada para manejar datos no paramétricos, tanto para el caso de variables categóricas como para el caso más tradicional de variables continuas (Metha y Patel, 1999). En los análisis se ha utilizado el procedimiento NPAR1WAY del paquete estadístico SAS.

Resultados

Comenzaremos exponiendo los resultados correspondientes a los veredictos y a las preguntas relativas al objeto del veredicto. Para la comparación de grupos en estas variables se utilizó la prueba exacta de Fisher. En la tabla 1 se presentan los resultados correspondientes a los veredictos en los dos casos.

Como se puede observar en la tabla 1, es en el caso de evidencia ambigua (caso 2) donde se producen diferencias estadísticamente significativas en el veredicto entre los Jurados homogéneos prodefensa y los proacusación, obteniéndose un tamaño del efecto moderadamente alto (0.53). En concreto, el 89% de los Jurados homogéneos proacusación emitieron un veredicto de culpabilidad, frente a un 25% de los Jurados homogéneos prodefensa. Por lo que respecta a la consideración del delito, todos los Jurados que consi-

	Est. Fisher	p exacta ^a	ϕ ^b
Caso 1	9.44	1.000	–
Caso 2	7.137	0.015	0.53

^aBilateral; ^bCoefficiente phi

deraron al acusado culpable, escogieron posteriormente la opción de homicidio involuntario. En el caso 1, de evidencia fuerte proacusación, excepto uno de los Jurados prodefensa que emitió un veredicto de inocencia, todos los demás emitieron un veredicto de culpabilidad. En este caso, sin embargo, no todos los grupos escogieron la opción de homicidio involuntario, ya que 3 de los 9 grupos proacusación decidieron considerarlo como voluntario, aunque esta diferencia con el grupo de Jurados prodefensa no resultó estadísticamente significativa.

Con respecto al objeto del veredicto, las diferencias entre los Jurados se encontraron de nuevo en el caso 2, de evidencia ambigua. Concretamente, en las preguntas 3 ('¿Considera probado que en la muerte de M.G. influyeron los golpes que le propinó el acusado?') y 7 ('¿Considera probado que el acusado podía prever la muerte de M.G.?'), ambas difíciles de responder con base en la evidencia presentada en el juicio. Así, en la pregunta 3, el 100% de los grupos homogéneos proacusación responde afirmativamente, frente a un 14% de los grupos prodefensa, obteniéndose un tamaño del efecto muy elevado (0.9). En la pregunta 7, responde afirmativamente un 77,8% de los grupos proacusación y un 25% de los grupos homogéneos prodefensa, obteniéndose en este caso un tamaño del efecto moderado (0.4). Los resultados se presentan en la tabla 2.

	Est. Fisher	p exacta ^a	ϕ ^b
P3C2	13.388	0.000	0.90
P7C2	4.735	0.05	0.40

^aBilateral; ^bCoefficiente phi

Para la comparación de los grupos de Jurados en las variables 'tiempo empleado en el proceso de deliberación' y 'número de votaciones necesarias para llegar a una mayoría cualificada', tanto en el veredicto como en las preguntas objeto del veredicto, se utiliza la prueba exacta de Wilcoxon-Mann-Whitney basada en el algoritmo desarrollado por Mehta y Patel (1983). Por lo que respecta al tiempo promedio empleado en las deliberaciones, se encontraron diferencias estadísticamente significativas entre los Jurados proacusación y los prodefensa, tanto en el caso 1 (W-M-W = 96, p exacta = 0.0181) como en el caso 2 (W-M-W = 93, p exacta = 0.033), siendo los Jurados prodefensa los que presentaron, en promedio, deliberaciones más largas. El tamaño del efecto en ambas pruebas arrojó valores moderadamente altos (0.56 y 0.52, respectivamente). En cuanto al número de votaciones necesarias para llegar a una mayoría cualificada en el veredicto, sólo encontramos diferencias estadísticamente significativas entre los Jurados en el caso 1, de evidencia fuerte proacusación (W-M-W = 91, p exacta = 0.043), con un tamaño del efecto de 0.49, siendo los Jurados prodefensa los que realizan mayor número de votaciones para llegar a una mayoría cualificada en el veredicto. Por lo que respecta a los resultados obtenidos en las variables relativas al número de votaciones necesarias para llegar a una mayoría cualificada en las preguntas objeto del veredicto, no se han obtenido diferencias estadísticamente significativas entre los Jurados prodefensa y proacusación en ninguno de los dos casos.

Discusión

En este estudio nuestro interés se ha centrado en analizar mediante simulaciones judiciales si la creación de grupos homogéneos de Jurados en cuanto a tendencias opuestas de sesgo, utilizando la Escala de Sesgo del Jurado, generaría decisiones grupales diferenciales, lo cual nos estaría aportando a su vez evidencias predictivas de la validez de la escala. Dado que los casos utilizados variaban en cuanto a la fuerza de las pruebas, esperábamos encontrar decisiones grupales diferentes en el caso de evidencia ambigua, pero no en el caso de evidencia o pruebas fuertes proacusación. Los resultados obtenidos han confirmado que, en casos de homicidio, la Escala de Sesgo del Jurado es capaz de discriminar los veredictos grupales de Jurados homogéneos proacusación frente a los de Jurados homogéneos prodefensa, en el caso de evidencia ambigua, no así en el caso de evidencia fuerte proacusación, en el que todos los grupos han emitido un veredicto de culpabilidad. Podemos decir, por tanto, que las tendencias generalizadas de respuesta que mide esta escala han sido lo suficientemente relevantes como para sesgar los veredictos grupales en una situación judicial con pruebas ambiguas, lo cual viene a apoyar los resultados de otros estudios que han utilizado la *Juror Bias Scale* en el contexto anglosajón (Kassin y Wrightsman, 1983; Kassin y Garfield, 1991; Narby, Cutler y Moran, 1993; Chapdelaine y Griffin, 1997; Myers y Lecci, 1998; Lecci, Myers, Denning, Littell, Robertson y Gorbe, 2000). En general, estos estudios, aunque han variado en cuanto a la metodología utilizada, el tipo de caso o el nivel de la decisión estudiado (individual o grupal), han evidenciado que los sujetos con puntuaciones más elevadas en la escala tienden a condenar, mientras que los que presentan puntuaciones más bajas, tienden a la absolución.

Por lo que respecta a los resultados obtenidos en relación con las preguntas que conforman el objeto del veredicto, la hipótesis de la que partíamos se ha confirmado en un caso pero no en el otro. Como esperábamos, han sido las preguntas de marcado carácter opinable las que han provocado respuestas diferenciales en los grupos de Jurados, aunque esto solamente ha ocurrido en el caso de evidencia ambigua. En el caso de evidencia fuerte proacusación, aunque no se encontraron diferencias estadísticamente significativas entre los grupos, fueron las preguntas de carácter opinable las que produjeron mayores discrepancias en las respuestas. Estos resultados apoyan una vez más la hipótesis de que la fuerza de la evidencia es un factor mediador en la relación sesgo-decisiones del Jurado (Dane y Wrightsman, 1982; Kaplan, 1989; Kerr, Hymes, Anderson y Weathers, 1995; Kerr, Niedermeier y Kaplan, 1999; García, 2000; Devine et al., 2001; Wrightsman, 2001) y, por tanto, es más probable que la influencia de este tipo de sesgos juegue un papel relevante en aquellos casos en los que las pruebas existentes no son concluyentes. Estos resultados, que en principio podrían minimizar la relevancia de este tipo de sesgos, se tornan especialmente interesantes si pensamos que, tanto en España (CGPJ, 1998, 1999), como en otros países donde rige el sistema de Jurado (Devine et al., 2001), la mayor parte de los casos que son juzgados mediante un tribunal de jurados son aquellos en los cuales no existen evidencias claras a favor de un parte, ya que en caso contrario, las partes suelen buscar un acuerdo y el juicio con Jurado no llega a celebrarse.

Por otra parte, la comparación de la ejecución de los diferentes grupos de Jurados en las variables relativas al proceso de deliberación aporta otros datos de interés. Los resultados obtenidos

han indicado diferencias entre los grupos prodefensa y los grupos proacusación en el tiempo empleado en las deliberaciones en los dos casos presentados. Así, hemos encontrado que, en promedio, las deliberaciones de los grupos homogéneos prodefensa han sido significativamente más largas que las de los grupos proacusación. Sin embargo, en las variables relativas al número de votaciones necesarias para llegar a una mayoría cualificada, las diferencias entre grupos proacusación y grupos prodefensa aparecen solamente en relación con los veredictos, en el caso de evidencia fuerte proacusación, donde el promedio de votaciones necesarias para llegar a un veredicto cualificado ha sido superior en los grupos prodefensa. Estos resultados parecen estar indicando que los Jurados prodefensa discuten más que los Jurados proacusación, aunque esto no se refleja necesariamente en el número de votaciones que realizan, a excepción del veredicto en el caso 1 donde, como hemos visto, los Jurados prodefensa, en promedio, han necesitado un mayor número de votaciones para emitir sus veredictos, que finalmente han sido de culpabilidad. El que los Jurados formados por sujetos clasificados por la ESJ como proacusación hayan tenido deliberaciones más cortas confirma la hipótesis planteada anteriormente y está en relación con otros estudios en los que se ha puesto de manifiesto que Jurados formados por sujetos con tendencias hacia la culpabilidad tardan menos en deliberar (Shaffer, Plummer y Hammock, 1986; Jackson, 1991; Shaffer y Kerwin, 1992). Por otra parte, el que los Jurados homogéneos prodefensa hayan necesitado más votaciones para condenar que los Jurados proacusación podría estar dando cuenta, como algunos autores han sugerido (Boyll, 1996) del 'coste cognitivo' o 'disonancia cognitiva' que puede suponer para unos sujetos con tendencia hacia la absolución el llegar a un veredicto de culpabilidad. Sin embargo, dado la carencia de datos de la que disponemos en este sentido, no estamos en disposición de poder establecer conclusiones al respecto. Estas son hipótesis tentativas que, aunque nos parecen interesantes, deberemos de poner a prueba en estudios posteriores.

En resumen, los resultados obtenidos han indicado que Jurados homogéneos en términos de tendencias opuestas de sesgo, medidas en este caso según la ESJ, producen decisiones substancialmente diferentes cuando la evidencia es ambigua, pero no cuando ésta favorece claramente a una parte. Como algunos autores han

puesto de manifiesto (Kramer et al., 1990), la probabilidad de conseguir un Jurado homogéneo en términos de sus preferencias individuales con respecto a determinadas decisiones, por ejemplo para condenar o absolver, simplemente como consecuencia de un proceso aleatorio, es una realidad que debemos tener en cuenta si no utilizamos métodos con suficientes garantías científicas que nos permitan detectar aquellos posibles jurados que por sus características personales puedan llegar, bajo determinadas circunstancias, a tomar decisiones sesgadas. En este sentido, la información que proporciona la ESJ puede ser útil, apuntándose como un instrumento a tener en cuenta, junto a otro tipo de estrategias procedentes de la investigación en este ámbito, así como de la experiencia de jueces y abogados, en la detección de tendencias de sesgo en la fase de selección del Jurado.

Por último, comentar en relación con la metodología utilizada en este estudio, simulaciones judiciales realizadas bajo condiciones controladas, que ha sido habitualmente criticada por falta de realismo y, por tanto, por problemas con la generalización de resultados. Sin embargo, ninguna otra aproximación es capaz de obtener el mismo grado de control sobre la influencia de factores extraños, así como de manipulación de variables relevantes, particularmente en lo que a características de los individuos o de un caso específico se refiere. Además, como señala Diamond (1997), la investigación con Jurados simulados ha mejorado substancialmente en los últimos 15 años (utilización de medios audiovisuales, etc.) y los datos disponibles sugieren que los Jurados simulados no parecen operar de una manera significativamente diferente a la de los Jurados reales (Saks, 1997; Bornstein, 1999). No obstante, en investigaciones futuras, para mejorar la generalización de resultados, sería positivo poder contar con muestras más amplias y representativas, ya que aunque los estudiantes universitarios pueden ser elegidos como miembros de un Jurado al igual que otro ciudadano que cumpla con los requisitos exigidos, no representan habitualmente al típico grupo de candidatos a Jurado que se pueden encontrar en los juicios reales.

Agradecimientos

Esta investigación forma parte del proyecto 1FD97-1593-CO1 del plan nacional de I+D, financiado con fondos FEDER.

Referencias

- AERA, APA, NCME (1999). *Standards for Educational and Psychological Testing*.
- Arce, R., Fariña, F. y Fraga, A. I. (2000). Género y formación de juicios en un caso de violación. *Psicothema*, 12(4), 623-628.
- Baumeister, R. F. y Darley, J. M. (1982). Reducing the biasing effect of perpetrator attractiveness in jury simulation. *Personality and Social Psychology Bulletin*, 8 (2), 286-292.
- Bornstein, B.H. (1999). The ecological validity of jury simulations: Is the jury still out?. *Law and Human Behavior*, 23, 75-91.
- Boyll, J.R. (1996). Psychological, cognitive, personality and interpersonal factors in jury verdicts. *Law and Psychology Review*, 15, 163-184.
- Bray, J.R. y Noble, A.M. (1978). Authoritarianism and decisions of mock juries: Evidence of jury bias and group polarization. *Journal of Personality and Social Psychology*, 36 (12), 1.424-1.430.
- Chapdelanie, A. y Griffin, S.F. (1997). Beliefs of guilt and recommended sentence as a function of juror bias in the O.J. Simpson trial. *Journal of Social Issues*, 53(3), 477-485.
- Cowan, C.L., Thompson, W.C. y Ellsworth, P.C. (1984). The effects of death qualification on juror's predisposition to convict and on the quality of deliberation. *Law and Human Behavior*, 8(1-2), 53-79.
- Dane, F.C. y Wrightsman, L.S. (1982). Effects of defendants' and victims' characteristics on jurors verdicts. In: L. Keer y M. Bray (Eds.): *The Psychology of the Courtroom*. New York: Academic Press.
- De la Fuente, E.I., García, J. y Martín, I. (1998). Some individual differences in perception of the evidence and the verdict choice. *Psychology, Crime and Law*, 4, 361-373.
- De la Fuente, L. (2001). Evaluación del sesgo de los jurados en situaciones simuladas de decisión individual y grupal. Almería: Servicio de Publicaciones de la Universidad de Almería. [Edición CD].
- Devine, D. J.; Clayton, L. D.; Dunford, B. B.; Seying, R. y Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy and Law*, 7(3), 622-727.
- Diamond, S.S. (1997). Illuminations and shadows from jury simulations. *Law and Human Behavior*, 21, 561-571.

- Fisher, R.A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.
- García, J. (2000). Adaptación del Cuestionario de Actitudes Legales para la definición de perfiles psicosociales en la selección de jurados. Almería: Servicio de Publicaciones de la Universidad de Almería. [Edición CD].
- García, J., De la Fuente, L. y De la Fuente, E.I. (en prensa). Aspectos psicológicos de la toma de decisiones de los miembros de un Jurado. En G. Buela, D. Bunce y E. Jiménez (Comps). *Manual de Psicología Forense*. Madrid: Biblioteca Nueva.
- Hastie, R., Penrod, S. y Pennington, N. (1986). *La institución del Jurado en los Estados Unidos. Sus intimidaciones*. Madrid: Civitas (Orig. 1983).
- Izzett, R. y Leginski, W. (1974). Group discussion and the influence of defendant characteristic in a simulated jury setting. *Journal of Social Psychology*, 93, 217-279.
- Jackson, J. (1991). Juror decision-making and the trial process. En G. Davies, S. Lloyd-Bostock, M. McMurrin y C. Wilson (Eds.), *Psychology, Law and Criminal Justice: International Developments in Research and Practice* (pp. 327-336). Berlin: Walter de Gruyter.
- Kaplan, M.F. (1989). Studying juror cognition by studying judgments of murder mysteries. *Basic and Applied Social Psychology*, 10(3), 225-240.
- Kassin, S.M. y Garfield, D.A. (1991). Blood and guts: General and trial-specific effects of videotaped crime scenes on mock jurors. *Journal of Applied Social Psychology*, 21(18), 1.459-1.472.
- Kassin, S.M. y Wrightsman, L.S. (1983). The construction and validation of a juror bias scale. *Journal of Research in Personality*, 12, 423-442.
- Kerr, N.L., Hymes, R.W., Anderson, A.B. y Weathers, J.E. (1995). Defendant juror similarity and mock juror judgments. *Law and Human Behavior*, 19(6), 545-567.
- Kerr, N.L., Niedermeier, K.E., & Kaplan, M.F. (1999). Bias in jurors vs. bias in juries: New evidence from the SDS perspective. *Organizational Behavior and Human Decision Processes*, 80, 70-86.
- Kerwin, J. y Shaffer, D. (1994). Mock jurors versus mock juries. The role of deliberations in reactions to inadmissible testimony. *Personality and Social Psychology Bulletin*, 20(2), 153-162.
- Kovera, M.B.; Dickinson, J.J. y Cutler, B.L. (en prensa). Voir dire and jury selection: Practical issues, research findings, and directions for future research. En A.M. Goldstein (Ed.), *Comprehensive Handbook of Psychology, Volume 11: Forensic Psychology*. New York: John Wiley & Son.
- Kramer, G. P., Kerr, N. L. y Carroll, J. S. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior*, 14(5), 409-438.
- Kravitz, D.A., Cutler, B.L. y Brock, P. (1993). Reliability and validity of the original and revised legal attitudes questionnaire. *Law and Human Behavior*, 17(6), 661-677.
- Laughlin, P.R. y Ellis, A.L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, 22, 177-189.
- Lecci, L., Myers, B., Denning, C., Littell, L., Robertson, A. y Gorbe, K. (2000). *Assessment of juror bias in a Community sample*. 108th Annual Convention of the American Psychological Association. Washington, August.
- MacCoun, R. J. (1989). Experimental research on jury decision-making. *Science*, 244 (4908), 1.046-1.050.
- Metha, C.R. y Patel, N.R. (1999). Exact permutational inference for categorical and nonparametric data. En R. H. Hoyle (Ed.), *Statistical Strategies for Small Sample Research*. London: Sage.
- Moran, G. (2000). *Florida International University Criminal Juror Survey*. American Psychological Association Annual Convention, Washington, D.C.
- Myers, B. y Lecci, L. (1998). Revising the factor structure of the Juror Bias Scale: A method for the empirical validation of theoretical constructs. *Law and Human Behavior*, 22(2), 239-256.
- Narby, D.J., Cutler, B.L. y Moran, G. (1993). A Meta-Analysis of association between Authoritarianism and Jurors' Perceptions of Defendant Culpability. *Journal of Applied Psychology*, 78 (1), 34-42.
- Pennington, N. y Hastie, R. (1981). Juror decision making models: The generalization gap. *Psychological Bulletin*, 89, 246-287.
- Saks, M.J. (1997). What do jury experiments tell us about how juries (should) make decisions? *Southern California Interdisciplinary Law Journal*, 6, 1-53.
- Shaffer, D.R. y Kerwin, J. (1992). On adhering to judicial instructions: Reactions of dogmatic and nondogmatic juries to the judge's charge in an entrapment case. *Journal of Applied Social Psychology*, 22 (14), 1.133-1.147.
- Shaffer, D.R., Plummer, D., y Hammock, G. (1986). Hath he suffered enough? Effects of jury dogmatism, defendant similarity, and defendant's pretrial suffering on juridical decisions. *Journal of Personality and Social Psychology*, 50, 1.059-1.067.
- Wrightman, L.S. (2001). *Forensic Psychology*. Belmont, CA: Wadsworth Thomson.