

Gerhard Kette, Vladimir Konečni

Geschlechtsspezifisches Dekodieren und Integrieren nonverbaler Information in der richterlichen Urteilsfindung¹

15 Videoaufzeichnungen der mündlichen Streitverhandlung im Zivilprozeß wurden in vier unterschiedlichen Darbietungsmodalitäten: „Video und Ton“, „nur Video“, „nur Ton“ und „Transkript“ einer Gruppe von 72 männlichen und 72 weiblichen Vpn dargeboten. Diese skalierten nach jeweils 30 Sekunden den Betrag des Schadenersatzes, den sie dem Kläger zugestehen und ihre subjektive Urteilssicherheit. Aus der Abweichung eines individuellen Urteils vom tatsächlichen Urteil ergab sich ein Maß für die subjektive Urteilsgenauigkeit, aus der Streubreite der Urteile innerhalb der beiden Geschlechtergruppen ein Maß für die Urteilsdisparität. Zur Prüfung des Einflusses nonverbaler Information auf die Urteilsbildung wurde die Methode der Zeitreihenanalyse verwendet. Weitere 48 Vpn skalierten die Glaubwürdigkeit der Streitparteien. Es zeigt sich, daß bei männlichen Beurteilern der Ausgang der Verhandlung durch nonverbale Informationen (Gesichtsausdruck und Körpersprache der Streitparteien) nicht vorhergesagt werden kann. Die größere Sensitivität der weiblichen Beurteiler für nonverbale Information bewirkte allerdings keine größere Urteilsgenauigkeit. Weiters waren Frauen in der Bedingung „nur Video“ in ihren Urteilen homogener, jedoch unsicherer in der „Transkript“-Bedingung. Die Effekte der wahrgenommenen Glaubwürdigkeit der Streitparteien waren vom Vorhandensein verbaler Inhalte abhängig. Die Ergebnisse werden im Kontext des Widerspruchs zwischen dem rechtlich geforderten und dem tatsächlich praktizierten Urteilsverhalten (insbesondere hinsichtlich der Gewichtung nonverbaler Information) diskutiert.

The presentation of evidence in 15 videotaped civil-court cases was communicated to 72 women and 72 men in one of four modes: Video & Sound, Video Only, Sound Only, and Transcript. Every 30 sec the subjects indicated how much money the plaintiff should be awarded and their confidence in that decision. Measures of individual accuracy (deviations from the judge's verdict) and homogeneity within each group of six subjects were obtained. To examine further the relative impact of nonverbally presented emotional information, multiple time-series methodology was used. Another 48 subjects rated the relative credibility of the litigants. The findings indicate that, for men, the facial configurations and body language of the litigants did not predict the verdicts. However, women's relative sensitivity to nonverbal information was not associated with greater accuracy. Also, women were more homogeneous than men in the Video Only condition, but less confident, especially in the Transcript condition. The effects of the litigant's perceived credibility were dependent on the presence of verbal content. The discussion focused especially on the extent to which jurors' overreliance on nonverbal information (particularly of an emotional nature) may be detrimental in legal decision making.

Es wird oft behauptet, daß Richter und Geschworene neben den rechtlich zulässigen Beweisen auch Informationen, die nach der Prozeßordnung nicht als Beweise gelten, in ihr Urteil einfließen lassen. Diese „außerrechtlichen“ Einflüsse könnten sich in Gerichtsverhandlungen, insbesondere im Zivilgerichtsverfahren, in dem viele mehrdeutige „psychologische Daten“ über die Streitparteien verarbeitet werden, urteilsverzerrend auswirken.

Schon Simulationsstudien aus den 60er und 70er Jahren konnten dies bestätigen. So wurde in Laborexperimenten die physische Attraktivität des Angeklagten (Landy & Aronson, 1969; Nemeth & Sosis, 1973), die „Anständigkeit“ des Opfers (Jones & Aronson, 1973) oder die Einstellungsähnlichkeit zwischen An-

geklagten und Geschworenen (Griffith & Jackson, 1973) als zwar unzulässige aber wirksame Determinanten der richterlichen Urteilsbildung nachgewiesen.

Das Geschlecht der an der Verhandlung beteiligten Personen – als kontroverser außerrechtlicher Einflußfaktor – wurde ebenfalls eingehend untersucht. In der vorliegenden Arbeit betrachten wir das Geschlecht des „Richters“ als einen außerrechtlichen Faktor in der (öffentlichen) mündlichen Streitverhandlung; geklärt werden sollte die Frage, inwieweit ein möglicherweise geschlechtsspezifisch unterschiedliches Dekodieren, Gewichten und Verarbeiten von emotionsgeladener Information zu unterschiedlichen Entscheidungsstrategien führt. Die meisten Arbeiten dazu verwendeten das Paradigma der simulierten Richterforschung (mock jury research) und untersuchten die Effekte auf Verurteilungshäufigkeit und Strafzumessung.

Geschlechtsspezifische Unterschiede in der Urteilsbildung von Geschworenen

In Simulationsstudien ist es nicht gelungen, typisch männliche bzw. weibliche Urteilspräferenzen nachzuweisen (Gray & Ashmore, 1976; Bray & Nobel, 1978). Es zeigten sich auch keine Geschlechtsunterschiede in den Urteilen (von Studenten) in einem Mordfall (Eisen & McArthur, 1979), oder von echten Geschworenen, die vier verschiedene Serien von Prozeßmaterialien zu beurteilen hatten (Penrod & Hastie, 1979). Rose und Prell (1955; auch Griffith & Jackson, 1973) fanden jedoch Unterschiede in den Empfehlungen für die Strafzumessung. Bei der Präsentation eines Autodiebstahls gab es keine Geschlechtsunterschiede im finalen Urteil; männliche Studenten schätzten jedoch die Schuldwahrscheinlichkeit höher ein, betrachteten das Delikt als schwerwiegender und verwendeten ein strengeres (vorsichtigeres) Entscheidungskriterium für das Urteil (für das „Außer-Zweifel-Stehen“ des Urteils; Kerr, 1978). In einer anderen Studie (Bray et al., 1978) konnte gezeigt werden, daß männliche Studenten am Beginn der „Geschworenenberatung“ strengere Strafen aussprechen als weibliche, daß sich dieser Haupteffekt aber im Laufe der Beratung abschwächt und schließlich verschwindet. Eine direkte Beziehung zwischen dem Geschlecht des Urteilers und der Strafzumessung scheint nicht zu bestehen. Weibliche „Geschworene“ bestrafen weibliche Angeklagte allerdings milder (Richey & Fichter, 1969), sie sind auch weniger geneigt, diese schuldig zu sprechen (Stephan, 1974); männliche „Geschworene“ lassen sich durch die Attraktivität des Angeklagten stärker beeinflussen (Efran, 1974).

Diese Effekte gelten nicht generell, sie sind von der Art des Deliktes abhängig. Bei Gewaltdelikten (insbesondere Vergewaltigung) sind weibliche „Geschworene“ erwartungsgemäß eher geneigt, den Angeklagten schuldig zu sprechen, sie halten das Opfer für weniger mitschuldig, beschreiben dieses auch positiver und sprechen für den Angeklagten strengere Strafen aus als ihre männlichen Kollegen (Miller & Hewitt, 1978; Calhoun et al., 1976; Rumsey & Rumsey, 1977; Scroggs, 1976). Studentinnen sind im Vergleich zu Studenten sowohl vor als auch nach der Beratung eher bereit einen Schuldspruch zu fällen; keinen Unterschied gab es in den Wahrscheinlichkeitsschwellen für eine Verurteilung – unabhängig von der Definition des „Außer-Zweifel-Stehens“ der Schuld (Kerr et al., 1976; auch Davis, Bray & Holt, 1977; Nagao & Davis, 1980). Anderen Forschern wiederum (Sealy & Cornish, 1973; Hoiberg & Stires, 1973) ist es nicht

gelungen, konsistente Geschlechtsunterschiede bei erwachsenen (simulierten) Richtern nachzuweisen.

Auch in Untersuchungen mit Materialien aus Zivilverfahren sind die Ergebnisse widersprüchlich. Green (1967) fand in bezug auf die Urteilspräferenzen keinen Geschlechtsunterschied; Stephan & Tully (1977) stellten dagegen fest, daß männliche Urteiler dem Kläger einen höheren Betrag an Schadenersatz zusprachen als weibliche. Die Abhängigkeit des geschlechtsspezifischen Einflusses von der Art des Deliktes konnte überzeugend belegt werden (Feeney & Roll, 1984). Bei Leib- und Leben-Delikten (Mord, Vergewaltigung) und Desertion gibt es einen klaren Haupteffekt des Geschlechts.

Genauer untersucht wurde der urteilsverzerrende Einfluß des Geschlechts mittels Archivanalysen (insbesondere im Zivilverfahren). Auch hier sind die Ergebnisse widersprüchlich: Nagel und Weitzman (1971) legten ihren Vpn reale Fälle vor und stellten fest, daß männliche „Richter“ männlichen Klägern im Schadenersatzverfahren nach Körperverletzungen höhere Entschädigungen zubilligten als weibliche. Analoges gilt für weibliche „Richter“, diese sprachen weiblichen Klägern höhere Entschädigungen zu. Diese Ergebnisse konnten in anderen Untersuchungen nicht bestätigt werden; es zeigte sich eher ein gegenteiliger Effekt, der gegengeschlechtliche Kläger wurde bevorzugt. Die Kläger haben in ungefähr gleich vielen Fällen „gewonnen“ bzw. „verloren“, unabhängig davon ob eine Beurteilergruppe vorwiegend aus Männern oder aus Frauen bestand. Reine Männergruppen neigten im Vergleich zu Frauengruppen jedoch dazu, höhere Entschädigungen zuzusprechen, und sie bevorzugten auch Parteien mit höherem Status (Snyder, 1971).

Insgesamt gibt es aber wenig eindeutige Hinweise auf systematische Unterschiede im Urteilsprozeß von Männern und Frauen. Gesichert ist lediglich, daß Männer in der Beratung aktiver sind als Frauen und daß man damit ihren größeren Einfluß auf das finale Urteil erklären kann. Die meisten Untersuchungen zum geschlechtsspezifischen Urteilen verwenden Materialien aus Zivilverfahren. Die simulierten Richter bevorzugten Kläger des eigenen Geschlechts, sprachen aber männlichen Klägern grundsätzlich mehr Entschädigung zu als weiblichen. In den simulierten Geschworenenberatungen (zum Schuld- bzw. Freispruch eines Angeklagten) zeigen sich ebenfalls keine widerspruchsfreien Ergebnisse.

Geschlechtsunterschiede in der richterlichen Urteilsbildung

Die Beziehung zwischen dem Geschlecht des Richters und der Urteilsbildung wurde bisher nur wenig, und fast ausschließlich mit der Methode der Archivanalyse untersucht. Da jedoch die Mehrheit der Richter männlich ist, ermöglichen diese Untersuchungen keine Aufschlüsse über ein spezifisch weibliches Urteilsverhalten. Das bisherige Forschungsinteresse konzentrierte sich auf die Interaktion zwischen dem Geschlecht des Richters und drei Fallcharakteristika: die Art des Deliktes, das Geschlecht des Angeklagten sowie seine ethnische Zugehörigkeit.

Myers und Talarico (1987) erfaßten in einer großangelegten Archivanalyse 27.000 Fälle und stellten fest, daß, entgegen den Erwartungen, die ausgespro-

chene Strafe (für schwarze wie für weiße Straftäter) bei weiblichen Richtern strenger ausfiel. Die Autoren untersuchten auch die Frage, ob weibliche Richter bestimmte Tätertypen anders behandelten als männliche. Es wurde erwartet, daß weibliche Richter generell milder strafen als männliche, ausgenommen bei Vergewaltigungsdelikten. Die Ähnlichkeiten im Urteilsverhalten waren jedoch wesentlich größer als die Verschiedenheiten; dort wo es Unterschiede gab, waren Frauen nicht nur bei Gewaltdelikten, sondern auch bei Eigentumsdelikten strenger als Männer.

Trotz der sehr kleinen Stichprobe von Richterinnen in dieser Studie weisen die Daten darauf hin, daß diese Gruppe Täter bei Vergewaltigungen strenger bestraft als Richter. Diese Täter müssen, wenn sie von einer Frau abgeurteilt werden, häufiger eine Freiheitsstrafe verbüßen, und sie erhalten dann auch bedeutend längere Freiheitsstrafen. Bemerkenswert ist, daß Richterinnen häufiger „geteilte“ Strafen (Freiheitsstrafe und Bewährung) aussprechen. Weiters stellten die Autoren fest, daß Richter weibliche Angeklagte nicht milder behandelten als dies Richterinnen taten. Es zeigte sich sogar das Gegenteil: Richter bestrafte weibliche Straftäter strenger als ihre weiblichen Kollegen und es gab auch Hinweise auf den Umstand, daß männliche Richter männliche Straftäter milder bestrafte. Diese Ergebnisse sind jedoch infolge der sehr kleinen Stichprobe nur mit Einschränkungen gültig.

Die informativste Untersuchung zur geschlechtsspezifischen Strafzumessung ist eine Archivanalyse von Kritzer und Uhlmann (1977). Die Autoren prüften widersprüchliche Hypothesen über das Strafzumessungsverhalten von weiblichen Richtern: einerseits könnte sich diese Gruppe durch die Straftäter bedrohter fühlen als ihre männlichen Kollegen, insbesondere bei Delikten gegen Leib und Leben und daher auch strenger strafen; andererseits könnte sie als Folge ihrer geschlechtsspezifischen Sozialisation milder urteilen. Analoges wurde für männliche Richter vorhergesagt: aufgrund von Sozialisationsfaktoren könnten sich männliche Richter gegenüber weiblichen Straftätern „väterlich und beschützend“ verhalten und diese milder bestrafen. Diese Hypothese konnte jedoch nicht bestätigt werden: das Strafzumessungsverhalten von männlichen und weiblichen Richtern unterschied sich in den erwarteten Aspekten fast nicht. Bemerkenswert war, daß sich in dieser Untersuchung die Behandlung von weiblichen bzw. männlichen Straftätern nicht unterschied. Männliche Richter handelten gegenüber weiblichen Straftätern weder „väterlich“, noch waren weibliche Richter empathischer als männliche. Weibliche Richter bestrafte Täter von Vergewaltigungsdelikten geringfügig strenger als männliche. Weiters bemerkenswert ist der Umstand, daß Richterinnen Diebe strenger bestrafte als dies Richter taten. Weitere Unterschiede im Strafzumessungsverhalten traten als Folge rechtlich zulässiger Faktoren auf.

Ebenfalls mit der Methode der Archivanalyse fanden Gruhl et al. (1981) heraus, daß weibliche Richter im Vergleich zu männlichen tendenziell häufiger die Freiheitsstrafe aussprechen. Obwohl das Strafzumessungsverhalten in Abhängigkeit vom Delikt variierte, gab es keine schlüssigen Ergebnisse hinsichtlich der Strenge der Strafe. Entgegen den Erwartungen sprachen in dieser Untersuchung weibliche Richter bei Vergewaltigungsdelikten weniger häufig eine Freiheitsstrafe aus als männliche. In dieser Studie zeigte sich jedoch, daß Richter

weibliche Straftäter milder bestrafen als dies ihre weiblichen Kollegen taten (weniger Gefängnisstrafen, insgesamt geringere Strafdauer).

Die bisherigen Forschungsarbeiten zur Frage einer geschlechtsspezifischen Urteilsbildung geben ein sehr inkonsistentes Bild, besonders dann, wenn man auch das Geschlecht des Angeklagten als Faktor kontrolliert. Die Geschlechtsunterschiede treten (im Zivil- und Strafverfahren gleichermaßen) nur fallweise auf und sie variieren beträchtlich in ihrer Intensität. Zudem gibt es wenig Informationen zu den Ursachen dieser Effekte.

Die meisten dieser Untersuchungen haben jedoch den direkten oder indirekten Einfluß von am Prozeß beteiligten Personen außer acht gelassen; so wurde z.B. die Empfehlung des Bewährungshelfers (im anglo-amerikanischen Rechtssystem) nie systematisch kontrolliert. In einer Archivanalyse von mehr als 1200 Fällen konnten Konecni & Ebbesen (1982) zeigen, daß die Empfehlung des Bewährungshelfers mit der Entscheidung des Richters in 87% der Fälle identisch war und daß die Richtung des Einflusses eindeutig vom Bewährungshelfer auf den Richter wies. Daher wäre eine Ausweitung des Untersuchungsdesigns auf ein 2 (männlich vs. weiblich) X 3 (Empfehlung des Bewährungshelfers vs. Richter vs. Angeklagter) Design erforderlich. In dieses Design könnte man auch Faktoren wie die Art des Delikts, Überzeugungskraft der Beweise, Vorstrafen des Angeklagten oder Status des Angeklagten vor der Verhandlung (U-Haft oder auf freiem Fuß) integrieren. All diese Faktoren wurden in der Studie von Konecni und Ebbesen als ursächlich für die Empfehlung des Bewährungshelfers (und indirekt für die Entscheidung des Richters) identifiziert.

Geschlechtsspezifisches Dekodieren nonverbaler Information

In einer Metaanalyse von 75 Studien (Hall, 1978) wurde überzeugend nachgewiesen, daß weibliche Vpn nonverbales Verhalten besser dekodieren als männliche. Frauen sind in der nonverbalen zwischenmenschlichen Kommunikation besser angepaßt als Männer (Rosenthal & DePaulo, 1979). Sie sind bessere De- und Enkodierer, insbesondere in spontanen Reaktionen (Fujita, Harper & Wiens, 1980) und bei positiven Nachrichten (Noller, 1980). Eine Längs- und Querschnittsprüfung der Fähigkeit im Dekodieren nonverbaler Hinweisreize zeigt, daß Frauen mit zunehmendem Alter diese Überlegenheit verlieren, jedoch nicht in allen Kommunikationskanälen gleichermaßen (Blanck, 1981). In einer Untersuchung von Interviews konnten Christensen und Rosenthal (1982) zeigen, daß das Dekodieren nonverbaler Hinweisreize eine wichtige Determinante für den interpersonalen Erwartungseffekt in der Kommunikation darstellt. Männer werden im Vergleich zu Frauen durch ihre Erwartungen stärker beeinflusst und sie erzeugen eine stärkere verhaltensmäßige Bestätigung dieser Erwartungen beim Kommunikationspartner (in Interviews). In Folgeuntersuchungen wurde dieses Phänomen durch das Hinzufügen neuer Variablen wie Persönlichkeit (Kombos & Fourmet, 1985; Schneider & Schneider-Düker, 1984), Sozialisation und Einstellungen (Noller, 1986) oder durch methodologische Diskussion (Eisenberg & Lennon, 1983) weiter aufgeklärt.

Im Gerichtssaal könnte die nonverbale Kommunikation (Gesichtsausdruck, Gestik, Körperhaltung, paralinguistische Elemente wie Variation von Tonhöhe,

Sprechgeschwindigkeit oder Lautstärke) auf subtile und den Beteiligten nicht bewußte Weise das Verfahren und das Ergebnis der Verhandlung beeinflussen.

Hypothesen

Die vorliegende Arbeit prüft die Bedeutung des Faktors Geschlecht beim Dekodieren und Integrieren nonverbaler Information in den Urteilsprozeß bei simulierten „Richtern“, denen die Fallinformation von Zivilverfahren auf vier verschiedene Arten dargeboten wurde: „Video und Ton“, „nur Video“, „nur Ton“ und „nur Transkript“. Besonderes Augenmerk galt der Bedeutung außerrechtlicher, emotionsgeladener Information. Auf der Basis der vorhandenen Literatur wurde vorhergesagt, daß weibliche Richter beim Dekodieren und im Entscheidungsprozeß stärker von emotionalen Informationen beeinflußt werden als Männer; insbesondere, wenn der Inhalt und die emotionale Färbung der Information nicht konsistent sind (daher das Design mit 4 Informationskanälen).

Teilweise basieren diese Hypothesen auf einer Studie von Kette und Brandstätter (1990) in der – mit Hilfe der Methode der Zeitreihenanalyse – gezeigt werden konnte, daß emotionale Information, die rechtlich relevante Anteile völlig ausschließt, den Urteilsprozeß beeinflussen kann und zwar bei weiblichen (simulierten) Geschworenen stärker als bei männlichen.

Die spezifischen Hypothesen sind: (a) Frauen können im Kontext einer Verhandlung nonverbal vermittelte Information besser dekodieren als Männer, wenn die Information vorwiegend emotionaler Natur ist; (b) wenn die Information nur visuell (Bedingung „nur Video“) dargeboten wird, sind Frauen beim Vorhersagen des Ergebnisses einer Verhandlung genauer als Männer; (c) Frauen schenken den nonverbalen Hinweisreizen mehr Beachtung – dies führt zu einer stärkeren Gewichtung dieser Information im Urteilsprozeß.

Weiters sollte die Frage beantwortet werden, inwieweit in der mündlichen Streitverhandlung das Geschlecht der Streitparteien und deren (unabhängig gemessene) Glaubwürdigkeit die Urteilsprozesse von weiblichen vs. männlichen „Richtern“ beeinflußt.

Methode

Versuchspersonen waren 72 männliche und 72 weibliche Studenten der Universität von Kalifornien, San Diego im Alter von 22 bis 31 Jahren (Durchschnittsalter 24 Jahre); mit der Teilnahme am Versuch konnten die Studenten bestimmte Erfordernisse des Studienplans erfüllen. Die Vpn wurden in Gruppen von 2 bis 5 Personen getestet, wobei jede Person zufällig einer Versuchsbedingung zugeordnet wurde (die Daten von 9 Vpn wurden nicht analysiert, da ihnen die Fallmaterialien teilweise schon bekannt waren). In jeder Bedingung des 2 (Geschlecht) X 4 (Darbietungsmodus) Designs waren 18 Vpn.

Das *Versuchsmaterial* bestand aus 15 Videoaufzeichnungen von mündlichen Streitverhandlungen, die aus einer populären, über mehrere Jahre in den USA ausgestrahlten TV-Serie („The People's Court“) stammten. Jede Filmsequenz

bestand aus einer kurzen Einführung, die den Titel des Falls und eine Beschreibung der Forderungen des Klägers enthielt, darauf folgte die Darbietung der Beweise (das waren zumeist Aussagen des Klägers, des Beklagten oder von Zeugen) und das Urteil des Richters. Der Richter war in allen Fällen ein und dieselbe Person und männlich. Die Kamera fokussierte immer auf den jeweiligen Sprecher. Dadurch konnten die Vpn in den visuellen Bedingungen die nonverbalen Hinweisreize (Mimik, Gestik, Körperhaltung) genau beobachten.

Die Rolle des Richters und jene der Parteien sind insofern sehr realitätsnah, als die Akteure nicht Schauspieler sondern Personen in realen Streitverhandlungen sind. Die Entscheidung des Richters ist ebenfalls verbindlich; dies wird in einem Vertrag mit den Streitparteien vor der Verhandlung vereinbart. In den 15 verwendeten Fällen reichte die Forderung des Klägers von \$ 24,- bis \$ 1500,- (das ist die Obergrenze des Streitwertes in diesem Verhandlungstypus), wobei der Durchschnitt bei \$ 877,- lag. Die Einleitung und ein Teil der Beweisdarbietung wurde als Stimulusmaterial verwendet. Die Filme (Dauer 4-12 Minuten, Durchschnitt 7 Minuten) wurden in 30-Sekunden-Segmente zerlegt; nach jedem dieser „Beweise“ gab es eine Pause von 5 Sekunden, in der die Vpn 2 Skalen (siehe weiter unten) verwenden mußten. Aus der Gesamtmenge von 15 Fällen wurden 3 Blöcke zu je 5 Fällen (balanciert hinsichtlich der Länge des Falls, des Geschlechts und der Glaubwürdigkeit des Klägers) gebildet. Jede Vpn bearbeitete somit 5 Verhandlungen.

Die Bedingung „Video und Ton“ enthielt sowohl rechtlich relevante als auch emotionale Information; in der Bedingung „nur Video“ konnten die Vpn den Fall nur sehen, in der Bedingung „nur Ton“ nur hören. Es wurde angenommen, daß die Bedingung „nur Video“ ausschließlich emotionale Information (keine faktische) enthielt und zwar in Form von Hinweisreizen aus dem Verhalten und aus dem Gesichtsausdruck. Die Bedingung „nur Ton“ mußte demnach neben den rechtlich relevanten Beweisen auch paralinguistische Hinweisreize auf Emotionen enthalten. Die Bedingung „Transkript“ enthielt nur insofern emotionale Information, als der Inhalt selbst Emotionen auslösen konnte.

Somit war in den vier Darbietungsmodalitäten der Anteil emotionaler bzw. faktischer Information graduell abgestuft. Einige Forschungsfragen zu diesen unterschiedlichen Kommunikationskanälen sind schon untersucht worden, z.B. wurde der Gesichtsausdruck zur Vermittlung von Emotionen eingehend geprüft (Fujita, Harper & Wiens, 1980; Rosenthal & DePaulo, 1979; Harris & Rosenthal, 1985). Mehrere Studien haben gezeigt, daß visuelle Hinweisreize allein die Erwartungen des Experimentators vermitteln können (Blanck & Rosenthal, 1984). Kraus et al. (1981) wiesen auf die große Bedeutung, die paralinguistische und visuelle Faktoren bei der Wahrnehmung von Emotionen anderer Personen haben, hin (siehe auch Ekman, 1980).

Versuchsplan, Versuchsdurchführung und abhängige Variablen

In dem 2 (Geschlecht des „Richters“) X 4 (Darbietungsmodus: „Video und Ton“, „nur Video“, „nur Ton“, „nur Transkript“) Design wurde jede Vpn zufällig einer Serie von fünf Fällen zugeteilt; eine Vpn beurteilte die Falldarbietung in jeweils nur einer Bedingung (in nur einem Darbietungsmodus). Den Vpn wurde mitge-

teilt, daß der Versuch der Frage galt, „wie genau eine juristisch ungeschulte Person das Ergebnis einer zivilgerichtlichen Verhandlung vorhersagen kann, und zwar in Abhängigkeit von der Informationsdarbietung“. Die Aufgabe der Vpn bestand darin, auf einer 120 mm langen graphischen Skala anzugeben, wieviel der Kläger von seiner Forderung berechtigterweise erhalten sollte. Die Skala reichte von 0 \$ bis zu dem Betrag, den der Kläger forderte. Die Vpn wurden gebeten rasch zu urteilen und alle verfügbaren Informationen zu berücksichtigen.

Als Maß für die Genauigkeit in den Vorhersagen des tatsächlichen Ergebnisses der Verhandlung wurde die Differenz zwischen dem tatsächlichen Urteil und dem Urteil der Vpn gebildet. Eine Standardisierung (durch Transformation) aller Werte auf eine 120 mm Skala war erforderlich, da die Forderungen der Kläger (in \$-Beträgen) uneinheitlich waren. Ebenfalls analysiert wurde die *Urteilshomogenität* (die Streubreite der Urteile) in den beiden Geschlechtergruppen. Beide Maße ergaben sich aus den jeweils letzten Skalierungen der Vpn; es mußte alle 30 Sekunden ein „Urteil“ abgegeben werden. Eine andere abhängige Variable (in einer multiplen Zeitreihenanalyse) war der Verlauf der Urteile in der Bedingung mit der vollständigsten Information („Video und Ton“); die Urteilssequenzen in den Bedingungen „nur Transkript“ und „nur Video“ waren die unabhängigen Variablen.

Ebenfalls geprüft wurde der Einfluß des Geschlechts des Klägers auf die Urteile der Vpn, und zwar als *post hoc* unabhängige Variable; nicht möglich war hingegen eine Prüfung des Faktors „Geschlecht des Beklagten“, da es nur in 2 der 15 Fälle weibliche Beklagte gab. Dies war eine Konsequenz der methodologischen Erfordernisse, die Fälle so aufzunehmen, wie sie in der TV-Serie ausgestrahlt worden waren.

Eine weitere Analyse galt der Frage nach dem Einfluß der *Glaubwürdigkeit* der Streitparteien auf das Urteil der simulierten Richter. Um eine Konfundierung der Aufgabe der 144 Vpn im Experiment mit der Einschätzung der Glaubwürdigkeit zu vermeiden, beurteilten 24 weibliche und 24 männliche Rater diese Eigenschaft der Streitparteien in den 15 Fällen auf einer 120 mm Skala (die Skala reichte von „überhaupt nicht glaubwürdig“ bis „völlig glaubwürdig“). Diese Skalierungen wurden für eine Teilung der 15 Fälle in eine Gruppe mit „relativ glaubwürdigen“ und in eine Gruppe mit „relativ unglaubwürdigen“ Klägern verwendet; maßgeblich war der relative Abstand der Glaubwürdigkeitseinschätzung zwischen den Streitparteien.

Keine Hypothesen gab es zur Frage nach möglichen Unterschieden in der Urteilssicherheit zwischen den beiden Geschlechtergruppen. Aus explorativen Überlegungen mußten die Vpn nach jeder Einstufung des Geldbetrages auch ihre subjektive Sicherheit beim Urteilen abgeben, wiederum auf einer 120 mm graphischen Skala, die von „ich bin überhaupt nicht sicher“ bis „ich bin absolut sicher“ reichte.

Ergebnisse

Urteilsgenauigkeit

Die Urteilsgenauigkeit wurde definiert als die absolute Differenz (in mm der graphischen Skala) zwischen dem Urteil einer bestimmten Vpn und dem tatsächlichen Urteil des Richters. Die Urteilsgenauigkeit innerhalb einer Gruppe von 6 männlichen bzw. 6 weiblichen Vpn wird als Disparität ihrer Urteile definiert (innerhalb eines Falls und innerhalb nur eines Darbietungsmodus). Das individuelle Genauigkeitsmaß gibt an, wieviel das Urteil eines simulierten Richters vom Urteil des echten Richters abweicht. Die Disparität spiegelt die Homogenität der Urteile innerhalb der beiden Geschlechtergruppen wider. Die Korrelationen zwischen diesen beiden Maßen zeigt die Tabelle 1. Es ist bemerkenswert, daß die Korrelationen für weibliche Richter höher sind als für männliche und daß sie in den „schwierigsten“ Bedingungen („nur Video“ und „nur Ton“) am geringsten sind.

Tab. 1: Korrelationen zwischen der durchschnittlichen individuellen Urteilsgenauigkeit und der Urteilsdisparität als Funktion des Geschlechts des „Richters“ und der experimentellen Bedingung

	Video u. Ton	nur Video	nur Ton	Transcript	Alle
Weibliche „Richter“	-.69**	-.58*	-.63*	-.65**	-.62**
Männliche „Richter“	-.50	-.40	-.37	-.67**	-.50**
Männliche + Weibliche	-.59**	-.49**	-.47**	-.66**	-.55**

Anmerkung: ** $p < .01$; * $p < .05$; die Korrelationen ergeben sich aus den Skalierungen von 15 Fällen und Gruppen von jeweils sechs (männlichen, weiblichen) „Richtern“.

Eine Varianzanalyse der individuellen Urteile ergab einen Haupteffekt des Darbietungsmodus $/F(3,719) = 2.78, p = 0.04^2$; in der Bedingung „nur Video“ war es am schwierigsten, das tatsächliche Urteil vorherzusagen. Die durchschnittliche absolute Abweichung vom Ausgang der realen Verhandlung – für beide Geschlechtergruppen zusammen – war 43.8 (mm der graphischen Rating-Skala) oder 37%. Die restlichen drei Bedingungen unterschieden sich nicht (die Abweichungen reichten von 40 mm in der Bedingung „Video und Ton“ bis 42 mm in der Bedingung „nur Ton“).

In der Urteilsgenauigkeit zeigt sich weder ein signifikanter Haupteffekt des Geschlechts der simulierten Richter $/F(1,719) = 1.48, p = 0.23$ / noch eine signifikante Wechselwirkung $/F(3,719) = 0.55, p = 0.65$ /. Reduziert man diesen Vergleich auf nur zwei Bedingungen, dann sind männliche Vpn in der Bedingung „Transkript“ tendenziell genauer als weibliche, in der Bedingung „nur Video“ können weibliche den Ausgang der realen Verhandlung besser vorhersagen als männliche (siehe Tabelle 2).

Tab. 2: Mittelwerte und Standardabweichungen der Urteilsgenauigkeit (Abweichung vom tatsächlichen Urteil) als Funktion des Geschlechts des „Richters“ und der experimentellen Bedingung

	Video u. Ton	nur Video	nur Ton	Transcript
Weibliche „Richter“				
M	42.7	49.7	45.7	44.1
SD	36.6	41.0	40.5	41.6
Männliche „Richter“				
M	37.6	52.7	38.9	38.4
SD	36.7	44.9	43.5	37.8

Anmerkung: Die Abweichung zwischen dem individuellen Urteil und dem tatsächlichen ist die absolute Differenz der beiden Skalierungen (in mm der 120 mm graphischen Skala). Hohe Werte entsprechen einer großen Abweichung vom tatsächlichen Urteil, also geringer Genauigkeit.

Ein signifikanter Unterschied zwischen männlichen und weiblichen Vpn in bezug auf die *Urteilsdisparität der Gruppe* würde entweder auf einen geschlechtsspezifischen Unterschied in der Verarbeitung faktischer bzw. emotionaler Information hinweisen oder auf ein geschlechtsspezifisch unterschiedliches Reagieren auf Aufgabenschwierigkeit, Entscheidungsstrategie und/oder mögliche andere Faktoren (zu deren Analyse die vorliegende Untersuchung jedoch nicht geplant war) oder in bezug auf beide Ursachen. Die Urteilsdisparität war definiert als die Standardabweichung der Urteile von männlichen bzw. weiblichen Vpn, die für jeden der 15 Fälle gesondert ermittelt wurde /120 Standardabweichungen: 2 (männliche bzw. weibliche Vpn in 6er-Gruppen) X 4 (Darbietungsmodus) X 15 (Fälle). Auch hier gab es keinen Haupteffekt des Geschlechts, $F(1,119) = 0.10$, $p = 0.10$, aber wieder einen signifikanten Haupteffekt des Darbietungsmodus, $F(3,119) = 2.84$, $p = 0.05$; die Wechselwirkung war nicht signifikant $F(3,119) = 1.25$, $p = 0.10$. In einer theoretisch wichtigen Analyse zeigte sich aber eine marginal signifikante Wechselwirkung zwischen diesen beiden Faktoren: 2 (Geschlecht) X 2 (3 Kanäle in denen faktische Informationen kommuniziert werden, also „Video und Ton“ + „nur Ton“ + „Transkript“ vs. nur nonverbale Information, also „nur Video“): $F(1,119) = 3.49$, $p = 0.06$. Diese Interaktion (siehe Tabelle 3) kann auf die größere Urteilsvariabilität der männlichen Vpn in der Bedingung „nur Video“ zurückgeführt werden und zwar im Kontext des Haupteffektes des Darbietungsmodus, der hauptsächlich durch die Bedingung „Video mit Ton“ (hier gab es die geringste Urteilsvariabilität für männliche wie auch für weibliche Vpn) entstand.

Tab. 3: Mittelwerte und Standardabweichungen der Urteilsdisparität als Funktion des Geschlechts des „Richters“ und der experimentellen Bedingung

	Video u. Ton	nur Video	nur Ton	Transcript
Weiblicher „Richter“				
M	31.6	36.5	35.3	37.4
SD	13.2	8.9	16.1	14.2
Männlicher „Richter“				
M	28.9	46.0	35.6	33.8
SD	13.1	8.0	19.8	11.7

Anmerkung: Die Urteilsdisparität ist die Standardabweichung der individuellen Urteile (in mm der 120 mm graphischen Skala) innerhalb einer Zelle und eines Falls; d.h., daß sich die Mittelwerte in der Tabelle aus 15 Standardabweichungen ergeben – entsprechend den 15 Fällen.

Einschätzung der Glaubwürdigkeit

Die von den Vpn wahrgenommene Glaubwürdigkeit der Streitparteien ist für die Aufklärung des Entscheidungsprozesses ebenfalls aufschlußreich. In 9 der 15 Fälle wurde der Kläger von unabhängigen Beurteilern als glaubwürdiger eingeschätzt als der Beklagte; der durchschnittliche Unterschied über alle 15 Fälle betrug nur 7.5 mm zugunsten des Klägers (auf der 120 mm langen Skala); die Streubreite reichte von 62 mm zugunsten des Klägers bis 42 mm zugunsten des Beklagten. Verwendet man die Glaubwürdigkeit des Klägers (relativ zur Glaubwürdigkeit des Beklagten) in einem Design mit 3 Faktoren – 2 (Geschlecht des simulierten Richters) X 2 (Glaubwürdigkeit des Klägers) X 4 (Darbietungsmodus) – als *post hoc* unabhängige Variable, zeigt sich ein Haupteffekt der Glaubwürdigkeit auf die Urteilsgenauigkeit $F(1,719) = 7.25, p < 0.01/$ und auf das letzte Urteil, also auf den letztmalig zugesprochenen Schadenersatzbetrag $F(1,719) = 41.82, p < 0.01/$. Ist der Kläger glaubwürdig, sind die Urteile genauer; es überrascht nicht, daß glaubwürdige Kläger von den simulierten Richtern einen größeren Betrag zugesprochen bekommen als unglaubwürdige. Es gab allerdings keinen Effekt der Glaubwürdigkeit auf die Urteilsdisparität $F(1,119) = 0.89, p > 0.10/$. Weiters ergaben sich signifikante Interaktionseffekte zwischen der Glaubwürdigkeit des Klägers und dem Faktor des Darbietungsmodus und zwar sowohl mit der Urteilsgenauigkeit als abhängiger Variable $F(3,719) = 2.56, p = 0.05/$ wie auch mit der Urteilsdisparität $F(3,119) = 3.89, p < 0.05/$. Keines dieser Maße interagierte mit dem Geschlecht. Aus den Tabellen 4 und 5 ist ersichtlich, daß die (schwachen) Effekte hauptsächlich auf die Wahrnehmung der Glaubwürdigkeit in der Bedingung „nur Video“ bzw. „Transkript“ zurückgeführt werden können: die wahrgenommene Glaubwürdigkeit beeinflusst nur dann den Urteilsprozeß, wenn eine möglichst reichhaltige Informationsbasis vorhanden ist.

Tab. 4: Mittelwerte der Urteilsgenauigkeit als Funktion der (relativen) Glaubwürdigkeit des Klägers und der experimentellen Bedingung

Relative Glaubwürdigkeit des Klägers	Video u. Ton	nur Video	nur Ton	Transcript
gering	43.3	53.1	50.9	56.0
hoch	37.8	51.2	37.6	31.1

Tab. 5: Mittelwerte der Urteilsdisparität als Funktion der (relativen) Glaubwürdigkeit des Klägers und der experimentellen Bedingung

Relative Glaubwürdigkeit des Klägers	Video u. Ton	nur Video	nur Ton	Transcript
gering	34.6	38.4	31.5	41.8
hoch	27.4	40.3	39.7	29.3

Einschätzung der Urteilssicherheit

Die Urteilssicherheit einer Vpn (oder genauer, ihre Sicherheit das Urteil des wirklichen Richters nachzuvollziehen) könnte für die introspektive Beschreibung ihrer Beziehung zur Aufgabe und der verwendeten Entscheidungsstrategie aufschlußreich sein. Die Ergebnisse sind in Tabelle 6 dargestellt. Männliche „Richter“ zeigen in allen vier Darbietungsmodalitäten größere Urteilssicherheit als weibliche $F(1,719) = 14.25, p < 0.01/$, und die signifikante Wechselwirkung F

(3,719) = 3.20, $p < 0.05$ / entsteht hauptsächlich aus dem geschlechtsspezifischen Unterschied in der Bedingung „Transkript“. Es überrascht, daß es keinen Haupteffekt des Darbietungsmodus gibt $F(1,719) = 1.33$, $p > 0.10$).

Tab. 6: Mittelwerte der Sicherheitseinschätzung als Funktion des Geschlechts des „Richters“ und der experimentellen Bedingung

	Video u. Ton	nur Video	nur Ton	Transkript
Weibliche „Richter“	76.5	79.7	77.0	65.3
Männliche „Richter“	83.4	81.9	81.4	84.9

Anmerkung: Die Einschätzung der Urteilssicherheit erfolgt auf einer 120 mm graphischen Skala die von „überhaupt nicht sicher“ bis „absolut sicher“ reicht.

Die Korrelationen zwischen Genauigkeit und Urteilssicherheit zeigt die Tabelle 7. Bei weiblichen Vpn gibt es keinen Zusammenhang zwischen Urteilssicherheit und Urteilsgenauigkeit. Bei Männern hingegen ist die Korrelation in jenen Bedingungen signifikant oder annähernd signifikant, in denen der faktische Inhalt im Vordergrund steht („Transkript“ bzw. „nur Ton“).

Tab. 7: Korrelationen zwischen der Urteilsgenauigkeit und der Urteilssicherheit als Funktion des Geschlechts des „Richters“ und der experimentellen Bedingung

	Video u. Ton	nur Video	nur Ton	Transkript
Weibliche „Richter“	-.07	-.07	.12	-.09
Männliche „Richter“	-.03	-.06	.20	.23*

Anmerkung: * $p < .05$. Ein positiver Wert zeigt an, daß hohe Urteilssicherheit mit hoher Urteils-genauigkeit zusammenhängt.

Das *Geschlecht des Klägers* wurde als *post hoc* unabhängige Variable in einer 2 (Geschlecht des Klägers; sieben männlich acht weibliche) X 2 (Geschlecht des „Richters“) X 4 (Darbietungsmodus) Varianzanalyse untersucht; die abhängige Variable war das individuelle Urteil (das ist der am Ende der dargebotenen Sequenz von „Beweisen“ zugestandene Schadenersatz als \$-Betrag). Es ergab sich ein (starker) Haupteffekt des Geschlechts des Klägers $F(1,719) = 12.56$, $p > 0.01$, und zwar so, daß weiblichen Klägern generell und unabhängig vom Kommunikationskanal ein höherer Betrag sowohl von männlichen als auch von weiblichen „Richtern“ zuerkannt wurde. Dieses Ergebnis ist jedoch wegen der geringen Zahl von Fällen (in nur zwei der 15 Fälle waren die Beklagten weiblichen Geschlechts) sowie infolge der Unmöglichkeit, die Fälle sinnvoll zu klassifizieren, wenig aussagekräftig.

Die Gewichtung der emotionalen Information

Die Frage nach der Gewichtung der emotionalen Information, insbesondere unter der Annahme, daß emotionale Information den Urteilsprozeß geschlechtsspezifisch unterschiedlich beeinflußt, läßt sich mit der Methode der *Zeitreihenanalyse* untersuchen (vgl. Möbus & Nagel, 1983). Es wurden zwei „unabhängige“ Variablen („nur Video“ und „Transkript“) und eine „abhängige“ („Video und Ton“) definiert.³ Dieser Analyse lag weiters die Annahme zugrunde, daß die „nur Video“-Bedingung vorwiegend emotionale und die „Transkript“-Bedingung vorwiegend faktische Information enthielt. In der Bedingung „Video und Ton“ wurde die

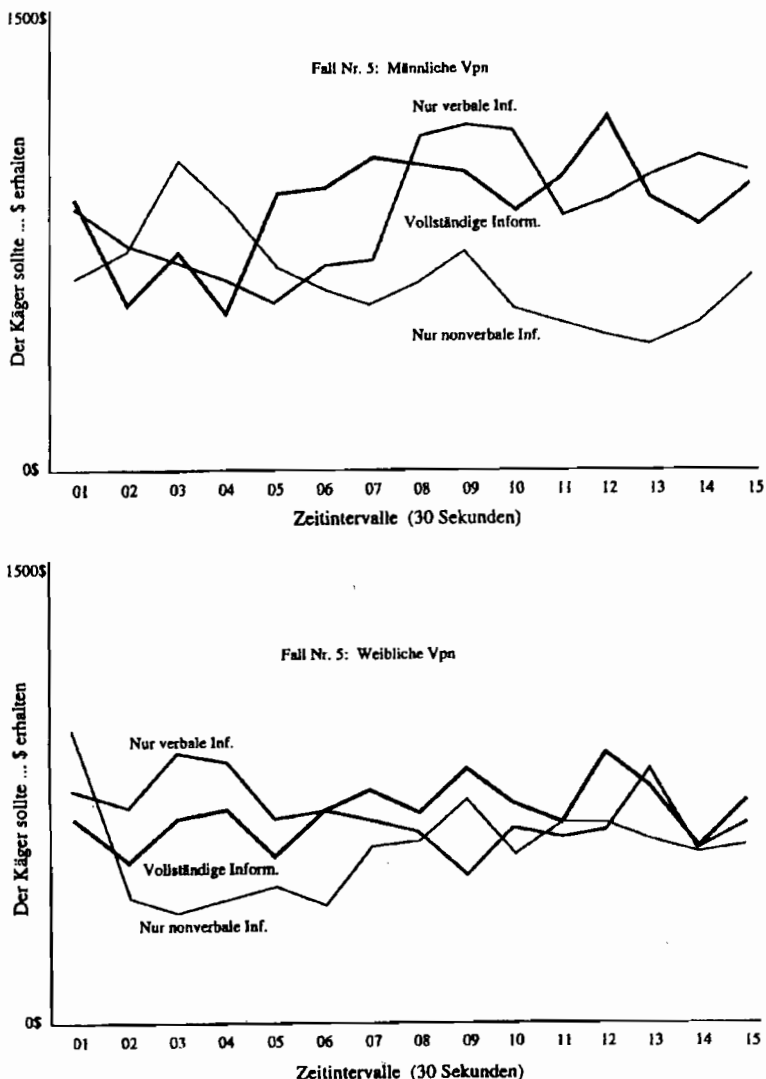


Abb. 1: Beispiele von Zeitreihen männlicher und weiblicher Vpn in einem Fall und drei Darbietungsmodalitäten; unabhängige Zeitreihen: nur nonverbale und nur verbale Information; abhängige Zeitreihe: Bedingung mit vollständiger Information („Video und Ton“).

die vollständigste (faktische und emotionale) Information kommuniziert. Die Daten für die Zeitreihenanalyse stammten aus den (getrennt für männliche und weibliche Vpn) nach jedem 30-Sekunden-Intervall gemittelten Urteilsverläufen – in allen 15 Fällen. Oder anders formuliert: die individuellen Zeitreihen sind die in den drei genannten Kanälen separat für Männer und Frauen gemittelten Zeitreihen.

In Abbildung 1 sind die statistischen Verfahren am Beispiel eines Falls veranschaulicht. Der obere Teil der Abbildung zeigt die durchschnittliche Zeitreihe (im Verlauf der 15 30-Sekunden-Segmente dieses Falls) für männliche Vpn in der Bedingung „nur Video“, „Transkript“ und „Video und Ton“. Im unteren Teil der Abbildung sind die Daten für weibliche Vpn dargestellt. Die Analyse vergleicht die Vorhersagekraft (β -Koeffizienten) der beiden „unabhängigen“ Variablen getrennt nach den beiden Geschlechtergruppen.

Bei weiblichen Vpn sind die Bedingungen „nur Video“ und „Transkript“ gleich gute Prädiktoren der abhängigen Variable, die Urteilsprozesse der männlichen Vpn können hingegen nur durch die Bedingung „Transkript“, also aufgrund der faktischen Information vorhergesagt werden; dies ist in Abb. 1 durch den Verlauf der Zeitreihen „nur Video“ veranschaulicht. In 11 von 15 Fällen gewichteten Frauen die nonverbale, emotionale Information stärker als Männer. Abb. 2 zeigt die β -Koeffizienten und die t-test-Statistik.

Autokorrelationen der Zeitreihen

Es wurde auch eine Analyse der 720 Autokorrelationen (144 Vpn beurteilen jeweils 5 Fälle) durchgeführt. Die Autokorrelation (ACF) oder partielle Autokorrelation (PACF, vgl. Möbus & Nagel, 1983) einer Zeitreihe kann man als „Gedächtnis“ für bestimmte Ereignisse in der Zeitreihe betrachten (zum Zeitpunkt t , t minus 1, t minus 2, ... t minus n), d.h. als Gedächtnis für bereits dargebotene „Beweise“ (oder bereits abgegebene Beurteilungen dieser „Beweise“). Eine solche Analyse könnte eine mögliche Interaktion von Geschlecht X Darbietungsmodus weiter aufklären; und zwar dann, wenn in den relevanten Zellen unterschiedliche Muster von Autokorrelationen auftreten.

Hohe Autokorrelationen können selbstverständlich auch durch Art und Reihenfolge der Beweisdarbietung verursacht sein. (Zum Beispiel könnte die in der Verhandlung dargebotene Sequenz von Beweisen die Position nur einer der beiden Parteien stärken; somit würde jedes der vorgebrachten Beweismaterialien einen kleinen Ausschlag der Zeitreihe in die eingeschlagene Richtung bedeuten und einen Pseudo-Trend der Zeitreihe verursachen).⁵ Um jene Gedächtniseffekte, die aus irrelevanten Effekten des Falltyps stammen, statistisch zu beseitigen, muß die Zeitreihe durch Differenzenbildung transformiert werden. Wenn dann noch statistisch signifikante Autokorrelationen übrigbleiben, weist dies darauf hin, daß vorangehende Urteile spätere beeinflussen; schon abgegebene Urteile werden dann noch einige Zeit „mitgezogen“. Man kann daraus folgern, daß ein gutes Gedächtnis für eine bestimmte Art von bereits dargebotener Information (z.B. in der Bedingung „nur Video“) den finalen Urteilsprozeß wesentlich beeinflussen kann.

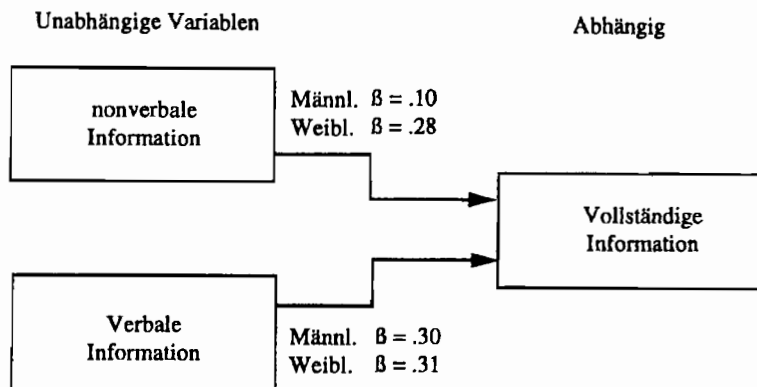


Abb. 2: Einfluß der Bedingung „nur Video“ (nonverbale Information) und „nur Transkript“ (verbale Information) auf die Bedingung „Video und Ton“ (vollständige Information) für männl. und weibl. Vpn. Für die nonverbale Information besteht ein signifikanter Unterschied (t-Test) zwischen den beiden Gruppen: $t = 2.19$, $df = 13$, $p < .05$; nicht für die verbale Information: $t = .18$, $df = 13$, $p > .10$.

In der Analyse der 720 Autokorrelationen in einem 2 (Geschlecht der Vpn) X 4 (Darbietungsmodus) Design zeigten sich keine signifikanten Effekte. In einer zweiten Auswertung wurden nur jene Kanäle untersucht, die auch in anderen Analysen wiederholt zwischen männlichen und weiblichen Vpn differenzierten. In einer 2 (Geschlecht der Vpn) X 2 (Bedingung „nur Video“ vs. Bedingung „Transkript“) Analyse konnten 360 Autokorrelationen verrechnet werden; diese ergab keinen Haupteffekt, jedoch eine signifikante Wechselwirkung $F(1,359) = 3.82$, $p < .05$. Bei Männern waren die Autokorrelationen in der Bedingung „nur Video“ geringer als für Frauen; das Gegenteil war in der Bedingung „Transkript“ der Fall (siehe Tabelle 8). Diese Ergebnisse bestätigen die weiter oben dargestellten Analysen zur zentralen theoretischen Frage.

Tab. 8: Mittelwerte der Autokorrelationen der individuellen Zeitreihen (Skalierungen der $\$$ -Beträge) als Funktion des Geschlechts des „Richters“ und der experimentellen Bedingung

		nur Video	Transcript
Weibliche „Richter“	M	.39	.28
	SD	.29	.29
Männliche „Richter“	M	.33	.34
	SD	.31	.33

Anmerkung: Die Autokorrelationen (time lag minus 1) entsprechen dem „Gedächtnis“ der individuellen Zeitreihen. $N = 90$ (18 Vpn pro Zelle, jede Vp skaliert fünf Fälle).

Diskussion

Gesichtsausdruck, Gestik und Körpersprache der Streitparteien waren bei männlichen Vpn im Vergleich zu weiblichen keine Prädiktoren für den Urteilsprozeß und das finale Urteil (siehe Urteilsverläufe; Abb. 1 und Abb. 2). Die relativ stärkere Gewichtung der nonverbalen Information durch Frauen führte jedoch nicht zu einer größeren Genauigkeit beim Dekodieren: wenn überhaupt etwas aus den Daten über Genauigkeitsunterschiede (Tabelle 2) verallgemeinert werden kann, dann der Umstand, daß Männer in drei von vier Darbietungsmodalitäten genauer urteilten als Frauen, und zwar dann, wenn die Information verbale/faktische Information enthielt – die Unterschiede waren jedoch nicht signifikant. Es ist anzunehmen, daß die Streitparteien in einer mündlichen Verhandlung relativ deutliche nonverbale (emotionale) Signale kommunizieren; die größere (bewußte oder unbewußte) Sensitivität und Aufmerksamkeit von Frauen für diese Signale könnte für die etwas geringere Urteilsgenauigkeit verantwortlich sein.

Weiters: Frauen waren bei der Darbietung der Verhandlung als Videoaufzeichnung ohne Ton in ihrem Urteil homogener (Tabelle 3); die größere Urteils-homogenität stand nur in den „schwierigen“ Darbietungsbedingungen mit größerer Urteilsgenauigkeit in Verbindung („nur Video“ bzw. „nur Ton“; Tabelle 1). Diese geschlechtsspezifische Eigentümlichkeit des Entscheidungsprozesses von Frauen führte jedoch nicht zu einer größeren Urteilssicherheit bzw. zu einem größeren Vertrauen in die Genauigkeit beim Urteil. Das Gegenteil war der Fall: Männer waren in allen vier Darbietungsmodalitäten in ihrem Urteil sicherer, insbesondere in der Transkript-Bedingung (Tabelle 6). Das größere Vertrauen in das eigene Urteil bei Männern steht mit deren größerer Urteilsgenauigkeit, insbesondere in jenen Bedingungen, die vermehrt faktenbezogene Informationen enthalten, in Beziehung (Tabelle 7).

Es gab allerdings keine Hinweise darauf, daß die größere Homogenität und die größere Sensitivität für nonverbale Hinweisreize (ungeachtet der geringeren Sicherheit) in den Urteilen von weiblichen Vpn zu einer objektiv „besseren“ Entscheidung (im Sinne von Fairneß oder „equity“) führten. Man könnte sogar schlußfolgern, daß die stärkere Gewichtung der emotionsgeladenen Information durch weibliche „Richter“, die vorwiegend nonverbal übermittelt wird, zu einer im Sinn der Zivilprozeßordnung „unfairen“ Haltung gegenüber den Streitparteien führt. Diese Behauptung würde aber voraussetzen, daß der in der TV-Serie gezeigte Richter „fair“ urteilt; und zwar gemessen an objektiven Kriterien – diese Annahme ist unter den gegebenen Bedingungen nicht haltbar. Weiters war der TV-Richter männlich und immer ein und dieselbe Person; es ist durchaus möglich, daß die größere Abhängigkeit der Urteile von emotionaler Information bei Frauen zu einer größeren Genauigkeit geführt hätte, wäre als Vergleichsstandard ein weiblicher TV-Richter verfügbar gewesen.

Die hier gezeigten geschlechtsspezifischen Urteilsdifferenzen sollten eine weitere Prüfung der Frage nach der unterschiedlichen Verwendung emotionaler (vs. rechtlich relevanter, faktischer) Information in rechtlichen Entscheidungsverfahren anregen. Die Klärung der Frage der postulierten unterschiedlichen Gewichtung nonverbaler (emotionaler) Information im Urteilsprozeß könnte auch

die in der Einleitung beschriebene Inkonsistenz der Forschungsergebnisse zur richterlichen Urteilsbildung erklären.

Die Tatsache, daß ein Maß für die Urteilsgenauigkeit im Kontext eines „quasi realen rechtlichen settings“ möglich war, verdient insofern Aufmerksamkeit, als in der Literatur darauf hingewiesen wird, daß sich Beurteiler (auch in rechtlichen Entscheidungssituationen) generell mehr an nonverbalen Hinweisreizen orientieren, insbesondere dann, wenn paralinguistische und verbale Inhalte widersprüchlich sind (Domangue, 1978; Efran, 1974; Ekman & Friesen, 1974; Kulka & Kessler, 1978; Littlepage & Pineault, 1981; Mehrabian & Wiener, 1967). Billigt man dem TV-Richter (und allgemein einem Richter im Straf- bzw. Zivilverfahren) den Status eines fairen und objektiven Urteilers zu (oder zumindest den eines erfahrenen), dann ist die stärkere Berücksichtigung von nonverbalen oder emotionalen Hinweisreizen bei Laienrichtern dem Erfordernis eines fairen Verfahrens gewiß abträglich (Maier & Thurber, 1968). Dann müßte man fragen, ob man diesen Mißstand (etwa durch eine verbesserte Rechtsbelehrung der Geschworenen) nicht beseitigen könnte.

In diesem Zusammenhang ist auch die Anmerkung interessant, daß der Einfluß der wahrgenommenen Glaubwürdigkeit der Streitparteien auf die Disparität wie auch auf die Genauigkeit der Urteile vom Vorhandensein verbaler Inhalte abhängig war (Tabelle 4). Am schwächsten waren diese Effekte in der Bedingung „nur Video“ und am stärksten in der Bedingung „Transkript“; es gab auch keine Interaktion mit dem Geschlecht des Beurteilers. Man wundert sich, worauf weibliche Beurteiler bei nonverbalen Signalen eigentlich reagieren, wenn diese Hinweisreize nichts zur wahrgenommenen Glaubwürdigkeit beitragen.

Angesichts der Unterschiede in den Bedingungen „nur Video“ vs. „nur Ton“ vs. „Transkript“ fällt einem die Darstellung der Justitia ein: sie hält ein Schwert in ihrer Rechten, eine Waage in der Linken und sie trägt eine Augenbinde – sie kann nur *hören*.

Anmerkungen

- 1 Diese Untersuchung wurde vom Österreichischen Fonds zur Förderung der wissenschaftlichen Forschung (Projekt Nr.: J70 an Gerhard Kette) und von einem Projekt an Vladimir Konecni finanziert. Die Autoren danken Mag. Gernot Fillip für seine Unterstützung bei der statistischen Datenanalyse und Univ. Prof. Dr. Hermann Brandstätter für wertvolle methodologische und theoretische Anregungen.
- 2 Mixed-design ANOVA: in dem 2 (Geschlecht) X 4 (Darbietungsmodus) Design beurteilen 18 Vpn in jeder Zelle fünf Episoden.
- 3 Als Maß für die Reliabilität der Skalierungen wurde die durchschnittliche Interkorrelation, die als „intra-class correlation“ (Guilford, 1954, S. 395) bekannt ist, verwendet. Die Reliabilität der Skalierungen in allen vier Darbietungsmodi war zufriedenstellend. Die Reliabilität der Urteile eines Raters reichte von $r\ 1/1 = .30$ bis $.54$, die Reliabilität der durchschnittlichen Urteile der 18 Rater pro Zelle von $r\ 18/18 = .61$ bis $.90$. Dadurch war es möglich, die individuellen Zeitreihen in eine Gruppen-Zeitreihe zu integrieren.
- 4 Die für solche Daten charakteristischen Trends, die künstlich hohe Korrelationen zwischen den unabhängigen und abhängigen Zeitreihen schaffen, und die zumeist

- nur Merkmale des Falls oder der Reizdarbietung anzeigen, wurden durch Differenzbildung, einem in der Zeitreihenanalyse üblichen Verfahren, beseitigt.
- 5 Um einen zu großen (fälschlichen) Effekt jedes einzelnen Beweismaterials auf die Reaktionen zu vermeiden, erhielten die Vpn bei jeder Skalierung die schriftliche Instruktion: „Consider all available information thus far, the plaintiff will receive“.

Literaturverzeichnis

- Blanck, P.D. (1981). Sex differences in eavesdropping on nonverbal cues: Developmental changes. *Journal of Personality & Social Psychology*, 41 (2) 391-396.
- Blanck, P.D., & Rosenthal, R. (1984). The mediation of interpersonal expectancy effects: The counselor's tone of voice. *Journal of Educational Psychology*, 76, 418-423.
- Bray, R.M., & Noble, A.M. (1978). Authoritarianism and decisions of mock juries: Evidence of jury bias and group polarization. *Journal of Personality and Social Psychology*, 36, 1424-1430.
- Bray, R.M., Struckman-Johnson, C., Osborne, M.D., McFarlane, J.B., & Scott, J. (1978). The effects of defendant status on the decisions of student and community juries. *Social Psychology*, 41, 256-260.
- Calhoun, L.G., Selby, J.W., & Warring, L.J. (1976). Social perception of the victims causal role in rape: An explorative examination of four factors. *Human Relations*, 29, 517-526.
- Christensen, D., & Rosenthal, R. (1982). Gender and nonverbal decoding skill as determinants of interpersonal expectancy effects. *Journal of Personality & Social Psychology*, 42 (1) 75-87.
- Davis, J.H., Bray, R.M., & Holt, R.W. (1977). The empirical study of social decision processes in juries. In J. Tapp, & F. Levine (eds.), *Law, justice, and the individual in society: Psychological and legal issues* (S. xxx). New York: Holt, Rinehart, & Winston.
- Domangue, R. (1978). Decoding effects of cognitive complexity, tolerance of ambiguity, and verbal-nonverbal inconsistency. *Journal of Personality*, 46, 519-522.
- Efran, M.G. (1974). The effect of physical appearance on the judgement of guilt, interpersonal attraction and severity of recommended punishment in a simulated jury task. *Journal of research in personality*, 8, 45-54.
- eisen, S.V., & McArthur, L.Z. (1979). Evaluating and sentencing a defendant as a function of his salience and the perceiver's set. *Personality and Social Psychology Bulletin*, 5, 48-52.
- Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, 94 (1), 100-131.
- Ekman, P. (1980). Relative importance of face, body, and speech in judgements of personality and affect. *Journal of Personality and Social Psychology*, 38, 270-277.
- Ekman, P., & Friesen, W.V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 29, 288-298.
- Feeney, S.K., & Roll, S. (1984). Sex as an extralegal factor in juridical decision-making: an analogue study. *American Journal of Forensic Psychology*, 2, 65-78.
- Fujita, B.N., Harper, R.G., & Wiens, A.N. (1980). Encoding-decoding of nonverbal emotional messages: Sex differences in spontaneous and enacted expressions. *Journal of Nonverbal Behavior*, 4 (3), 131-145.
- Gray, D.B., & Ashmore, R.D. (1976). Biasing influence of defendants' characteristics on simulated sentencing. *Psychological Reports*, 38, 727-738.
- Green, E. (1967). The reasonable man: Legal fiction or psychosocial reality? *Law and Society Review*, 2, 241-257.

- Griffitt, W. & Jackson, T. (1973). Simulated jury decisions: The influence of jury-defendant attitude similarity-dissimilarity. *Social Behavior and Personality*, 1 (1), 107-111.
- Gruhl, J., Sphohn, C., & Welch, S. (1981). Women as policymakers: the case of trial judges. *American Journal of Political Science*, 25, 308-322.
- Guilford, J.P. (1954). *Psychometric Methods*. 2nd ed. New York: McGraw.
- Hall, J.A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85 (4), 845-857.
- Harris, A., & Rosenthal, R. (1985). The mediation of interpersonal expectancy effects. *Psychological Bulletin*, 97, 363-382.
- Hoiberg, B., & Stires, L. (1973). The effect of several types of pretrial publicity on the guilt attributions of simulated jurors. *Journal of Applied Social Psychology*, 3, 267-275.
- Jones, C. & Aronson, E. (1973). Attribution of fault to a rape victim as a function of respectability of the victim. *Journal of Personality and Social Psychology*, 26, 415-419.
- Kerr, N.L. (1978). Severity of prescribed penalty and mock jurors' verdicts. *Journal of Personality and Social Psychology*, 36, 1431-1442.
- Kerr, N.L., Atkin, R., Stasser, G., Meek, D., Holt, R., & Davis, J. (1976). Guilt beyond a reasonable doubt: Effects of concept definition and assigned rule on the judgements of mock jurors. *Journal of Personality and Social Psychology*, 34, 282-294.
- Kette, G. & Brandstätter, H. (1990). Time series analysis of a murder trial. *Journal of Social Behavior*, 5, 15-31.
- Kombos, N.A., Fournet, G.P. (1985). Effects of dominance- submissiveness and gender on recognition of nonverbal emotional cues. *Educational & Psychological Research*, 5 (1), 19-28.
- Konečni, V.J., & Ebbesen, E.B. (1982). An analysis of the sentencing system. In V.J. Konečni & E.B. Ebbesen (Eds.), *The criminal justice system: A social-psychological analysis* (S. xxx). San Francisco: W.H. Freeman.
- Kraus, R.M., Apple, W., Moency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal, and visible factors in judgements of another's affect. *Journal of Personality and Social Psychology*, 40, 312-320.
- Kritzer, H.M., & Uhlman, T.M. (1977). Sisterhood in the courtroom: sex of judge and defendant in driminal case disposition. *Social Science Journal*, 14, 77-88.
- Kulka, A., & Kessler, xxx (1978). Is justice really blind? - The influence of litigant physical attractiveness on juridical judgement. *Journal of Applied Social Psychology*, 8, 366-379.
- Landy, D., & Aronson, E. (1969). The influence of the character of the criminal and his victim on the decision of simulated jurors. *Journal of Experimental Social Psychology*, 5, 141-152.
- Littlepage, F., & Pineault, xxx (1981). Detection of truthful and deceptive interpersonal communications across information transmission modes, *Journal of social Psychology*, 114, 57-65.
- Maier, E., & Thurber, G. (1968). Accuracy of judgements of deception when an interview is watched, heard, and read, *Personnel Psychology*, 21, 23-40.
- Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology*, 6, 1081-1083.
- Miller, M., & Hewitt, J. (1978). Conviction of a defendant as a function of juror-victim racial similarity. *Journal of Social Psychology*, 105, 159-160.
- Möbus, C. & Nagl, W. (1983). Messung, Analyse und Prognose ovn Veränderungen. In J. Breidenkamp und H. Feger (Hrsg.), *Hypothesenprüfung. Enzyklopädie der Psychologie* B15 (S. 239-470). Göttingen: Hogrefe.

- Myers, M.A., & Talarico, S.M. (1987). *The Social Context of Criminal Sentencing*. New York: Springer.
- Nagao, D.H., & Davis, J.H. (1980). The effects of prior experience on mock juror case judgements. *Social Psychology Quarterly*, 43, 190-199.
- Nagel, S.S., & Weitzman L.J. (1971). Women as litigants. *Hastings Law Journal*, 23, 171-198.
- Nemeth, C., & Sosis, R.H. (1973). A simulated jury study: Characteristics of the defendant and the jurors. *Journal of Social Psychology*, 990 (2), 221-229.
- Noller, P. (1980). Misunderstanding in marital communication: A study of couples' nonverbal communication. *Journal of Personality & Social Psychology*, 39 (1-6), 1135-1148.
- Noller, P. (1986). Sex differences in nonverbal communication: Advantage lost or supremacy regained? *Australian Journal of Psychology*, 38 (1), 23-32.
- Penrod, S.D., & Hastie, R. (1979). Models of jury decision making. *Psychological Review*, 87, 133-159.
- Richey, M.H., & Fichter, J.J. (1969). Sex differences in moralism and punitiveness. *Psychonomic Science*, 16, 185-186.
- Rose, A., & Prell, A. (1955). Does the punishment fit the crime? A study in social valuation. *American Journal of Sociology*, 61, 247-259.
- Rosenthal, R., & DePaulo, B.M. (1979). Sex differences in eavesdropping on nonverbal cues. *Journal of Personality & Social Psychology*, 37 (2), 273-285.
- Rumsey, M.G., & Rumsey, J.M. (1977). A case of rape: Sentencing judgements of males and females. *Psychological Reports*, 41, 459-465.
- Schneider, J., & Schneider-Düker, M. (1984). Sex roles and nonverbal sensitivity. *Journal of Social Psychology*, 122 (2), 281-282.
- Scroggs, J.R. (1976). Penalty for rape as a function of victim provocativeness, damage, and resistance. *Journal of Applied Social Psychology*, 6, 360-368.
- Sealy, A.P., & Cornish, W.R. (1973). Jurors and their verdicts. *Modern Law Review*, 36, 496-508.
- Snyder, E.C. (1971). Sex role differential and juror decisions. *Sociology and Social Research*, 55, 442-448.
- Stephan, C. (1974). Sex prejudice injury simulation. *Journal of Psychology*, 88, 305-312.
- Stephan, C., & Tully, J.C. (1977). The influence of physical attractiveness of a plaintiff on the decisions of simulated jurors. *Journal of Social Psychology*, 101, 149-150.

Zu den Autoren: Dr. Gerhard Kette, Ass. Prof., Univ. Doz. an der Johannes Kepler Universität Linz, Institut für Pädagogik und Psychologie, A-4040 Linz.
Prof. Vladimir Konečni, University of California, San Diego.