
Expert Testimony on Eyewitness Evidence: In Search of Common Sense

Kate A. Houston*, Lorraine Hope[†], Amina Memon[‡] and J. Don Read[§]

Surveys on knowledge of eyewitness issues typically indicate that legal professionals and jurors alike can be insensitive to factors that are detrimental to eyewitness accuracy. One aim of the current research was to assess the extent to which judges, an under-represented sample in the extant literature, are aware of factors that may undermine the accuracy and reliability of eyewitness evidence (Study 1). We also sought to assess the knowledge of a jury-eligible sample of the general public (drawn from the same population as the judges) and compared responses from a multiple choice survey with a scenario-based, response-generation survey in order to investigate whether questionnaire format alters the accuracy of responses provided (Study 2). Overall, judges demonstrated a reasonable level of knowledge regarding general eyewitness memory issues. Further, the jury-eligible general public respondents completing a multiple choice format survey produced more responses consistent with experts than did participants who were required to generate their own responses. The results are discussed in terms of the future training requirements for legal professionals and the ability of jurors to apply the knowledge they have to the legal context. Copyright © 2013 John Wiley & Sons, Ltd.

Mistaken eyewitness testimony is considered by many to be responsible for 75% of 301 cases of wrongful imprisonment in the US (The Innocence Project, 2012; Wells, Memon, & Penrod, 2006; Wise & Safer, 2004). In many of these cases, the judge and/or juries were convinced by testimony from a witness which implicated the defendant but which was, in fact, inaccurate (e.g. Wells et al., 2006). Research in psychology has been directed at identifying the factors that may affect the accuracy and reliability of eyewitness memory. However, the extent to which scientific findings align with the “common sense” of legal professionals and jurors is less clear. Investigations to date of the knowledge transfer among academics, potential jurors and legal professionals have presented a somewhat bleak picture (e.g. Magnussen et al., 2008; Wise & Safer, 2004; Wise, Safer, & Maro, 2011).

Knowledge Base of Legal Professionals Regarding Eyewitness Testimony

In a survey of 160 US judges’ knowledge and beliefs about eyewitness testimony, Wise and Safer (2004) identified a number of areas in which judges’ beliefs did not reflect

*Correspondence to: Dr. Kate A. Houston, Department of Psychology, University of Texas at El Paso, El Paso, TX 79968, U.S.A. E-mail: KHouston@utep.edu

[†]University of Portsmouth.

[‡]Royal Holloway, University of London.

[§]Simon Fraser University.

current research evidence. For instance, judges' knowledge differed from research evidence in relation to: the optimum methods for line-up presentation and administration; memory decline; how the presence of a weapon might affect memory for the perpetrator; the effects of disguise on the ability to accurately describe and identify the perpetrator; and the potential effects of post-event information on eyewitness testimony (Wise & Safer, 2004). Further surveys outside the US suggest similar low levels of knowledge regarding eyewitness memory issues by legal professionals. Granhag, Strömwall, and Hartwig (2005) found that the responses of Swedish police, lawyers and judges were in line with expert opinion on issues such as the possible effects of weapon presence during a crime, but not on issues such as line-up construction and administration. Furthermore, Granhag *et al.* (2005) found that the legal professionals surveyed seldom agreed with each other on whether certain factors might affect the reliability and accuracy of eyewitness testimony. They reported that the judges, in particular, were more likely than police officers or lawyers to respond "don't know" to questionnaire items. Indicative of this pattern of responding, Granhag *et al.* (2005) also reported that the legal professionals who were surveyed felt they were not up-to-date with research on the reliability of eyewitness testimony. In a further survey, Magnussen *et al.* (2008) found comparable low levels of knowledge regarding eyewitness memory among a Norwegian judicial sample.

Taken together, the findings from the Innocence Project and surveys such as Wise and Safer (2004) and Magnussen *et al.* (2008) suggest that the knowledge base of legal professionals on eyewitness testimony is inadequate. However, in spite of these findings, the acceptance of expert testimony on the quality of witness evidence in court is relatively uncommon. One argument is that an understanding of the potential weaknesses or inaccuracies of eyewitness testimony falls within the domain of common sense of both legal professionals and jurors. The extant literature, on the other hand, suggests that the workings and vulnerabilities of eyewitness memory fall largely outwith the realms of "common sense".

Knowledge Base of Jurors Regarding Eyewitness Testimony

Surveys of mock jurors' knowledge of eyewitness issues are more numerous than those of judges and legal professionals. However, the findings of mock-juror surveys are similar to surveys of legal professionals, as respondents typically demonstrate an understanding of eyewitness testimony that is at odds with research findings (Benton *et al.*, 2006; Brigham & Wolfskeil, 1983; Deffenbacher & Loftus, 1982; McConkey & Roche, 1989; Noon & Hollin, 1987). Mock jurors tend to be insensitive to biased procedures used by law enforcement, such as poorly constructed line-ups, misleading feedback or biased instructions (Shaw, Garcia, & McClure, 1999). Potential jurors also find it difficult to distinguish between accurate and inaccurate witnesses (e.g. Lindsay, Wells, & O'Connor, 1989; Lindsay, Wells, & Rumpel, 1981).

Part of the underlying cause of poor juror understanding regarding eyewitness memory issues may be a lack of knowledge regarding the ability and performance parameters of memory more generally. For instance, Simons and Chabris (2012) surveyed the general public regarding common memory myths, such as events being recorded in the memory akin to a video tape and able to be reviewed and inspected at a later date, and that once a memory has been formed for an event it will not change. Simons and Chabris (2012) found that their respondents claimed these memory myths

were truths over 50% of the time. Furthermore, previous experience as a witness does not appear to be related to knowledge of eyewitness issues (Noon & Hollin, 1987).

Desmarais and Read (2011) conducted a meta-analysis of 23 experiments that documented the knowledge of the general public about factors that may affect the reliability of eyewitness testimony. Desmarais and Read (2011) found that the responses of the general public matched the general consensus expert opinion in the field for two-thirds of questions. However, the meta-analysis also revealed inconsistencies in the ways in which “expert agreement” is reached on a topic and the topics that are generally assessed across studies (Desmarais & Read, 2011). Owing to variability across studies in terms of the topics on which potential jurors were assessed and the differences in methods used to measure knowledge, it is difficult to evaluate the precise nature of potential juror understanding.

Current Research

Research has typically investigated either the knowledge of legal professionals or the knowledge of a jury-eligible general public regarding eyewitness evidence. However, no research to date has assessed the knowledge of judges and the knowledge of a jury-eligible general public from the same population. Consequently, an accurate representation of the level of knowledge regarding eyewitness issues present in a given courtroom as a function of either common sense (jurors) or professional experience (judges) is absent from the extant literature. The current series of studies assessed the levels of knowledge regarding eyewitness testimony among judges and jurors from the same population, that is, individuals whose home nation is Scotland, UK. Both groups (judges and jurors) completed the same questionnaire.

STUDY 1

The commonly asserted assumption by judges that eyewitness memory is a matter of common sense contradicts research findings demonstrating that potential jurors (Cutler & Penrod, 1995; Cutler, Penrod, & Dexter, 1990; Kassin & Barndollar, 1992; Shaw et al., 1999) and legal professionals (Granhag et al., 2005; Wise & Safer, 2004) are typically rather limited in their understanding of factors affecting eyewitness accuracy. Therefore, in an attempt to assess the knowledge base of judges and jurors within the same population, our first experiment assessed judges’ knowledge regarding eyewitness testimony.

Method

Participants

Ninety-nine judges took part in our survey.¹ Judges were recruited to take part in the research during a routine professional development seminar run by the Judicial Studies

¹ The judges were informed that the provision of all information was voluntary. All of the judges surveyed declined to complete the demographic information sheet provided to them. Therefore, the demographic information presented here represents demographic information available from the Judiciary of Scotland website: <http://www.scotland-judiciary.org.uk>

Committee. The judges surveyed were all at the Scottish rank of “sheriff,” which means they had at least 10 years’ experience as an advocate, solicitor or lawyer as well as considerable court experience. Sheriffs deal with the majority of criminal court cases in Scotland and must retire from the bench on the day of their 70th birthday (Sheriffs’ Association, 2012). Of the current 142 sheriffs in Scotland, 112 are male and 30 are female (Sheriffs’ Association, 2012). The sample recruited for the current study represents 70% of all active sheriffs in the jurisdiction.

Survey Development and Administration

A pool of statements concerning eyewitness identification issues, including statements used by Kassin *et al.* (2001); Read and Desmarais (2009a) and Deffenbacher and Loftus (1982), was generated. From this initial pool a multiple choice questionnaire was developed. The eyewitnesses topics selected for inclusion were based on ratings of reliability supplied in the Kassin *et al.* (2001) expert survey in response to the question, “Do you think this phenomenon is reliable enough for psychologists to present in courtroom testimony?” Topics with levels of agreement among experts below 90% with respect to reliability and/or research basis were not included in the survey, with the exception of a trained observers question. The trained observer item in Kassin *et al.* (2001) only received 39% consensus amongst experts and a 75% agreement that there was a research basis for the conclusion that trained observers are no better or worse than untrained eyewitnesses. However, given that police officers frequently deliver eyewitness evidence in court, the perceived accuracy of police (and other trained observers) as eyewitnesses was included as an item, based upon Read and Desmarais (2009a) and Kassin *et al.* (2001). The question was as follows: police officers often witness various crimes and have to report their memories of them. In your opinion, police officers: (a) make more accurate witnesses than the average person; (b) are as accurate as the average person; (c) make less accurate witnesses than the average person; or (d) I don’t know. Response option (b) was designated as consistent with expert opinion in line with Kassin *et al.* (2001).

The multiple choice (MC) questionnaire adopted a choice response format such that respondents were required to complete the statement by circling their preferred response option. Response options were provided for each question and all included an “I don’t know” option. Each set of response options also included a response consistent with the current understanding of that particular phenomenon within the literature as determined by expert agreement within the Kassin *et al.* (2001) survey and evaluation by the authors. One such example [drawn from the Read and Desmarais (2009a) survey] was: “Sometimes witnesses experience crimes under the influence of alcohol. In your opinion, alcohol intoxication: (a) improves a witness’s ability to later recall people/crimes; (b) has no influence on a witness’s ability to later recall people/crimes; (c) reduces a witness’s ability to later recall people/crimes; or (d) I don’t know.

Initial piloting of the questionnaire resulted in the rewording of a number of statements to improve respondents’ understanding of the issues. As in Read and Desmarais (2009a), this typically required some rewording to make the statements more comprehensible to a non-academic audience (i.e. ensure the meaning of the question was clear and did not contain unfamiliar jargon). Following piloting, the final questionnaire comprised 11 topic areas (see Table 1). For 10 of the topics selected, expert agreement according to the Kassin *et al.* (2001) results indicated an overall mean agreed reliability

of 90%, while an overall mean of 94% agreed that there was a research basis for this conclusion, with the exception of the trained observers item.

An introductory paragraph at the outset of the questionnaire informed respondents that the aim of the study was to examine how they believed a typical witness would behave in particular circumstances. Respondents were asked to indicate which response they believed was most accurate. They were also instructed not to guess and to use the “I don’t know” response if they could not identify the accurate response. Respondents were instructed to read and complete each question carefully in the order presented and in their own time.

The survey was administered during a training course. Judges were requested to complete the survey individually without discussion and were allocated as much time as necessary to complete it. Responses to the questionnaire were either consistent or inconsistent with expert opinion, in line with the findings of Kassin et al. (2001). For example, for the question on post-event information “When witnesses are asked to report about a crime they saw, their report generally: (a) includes not only what they actually saw but also information they learned after the crime; (b) includes only what they actually saw; or (c) I don’t know, response (a) would be coded as consistent with expert opinion, response (b) would be coded as inconsistent with expert opinion and response (c) would be coded as a “don’t know” response. Therefore, a coding packet, designating responses as either consistent or inconsistent with expert opinion as above, was generated in line with the Kassin et al. (2001) findings. Thus, given that only one of the MC responses per topic was consistent with expert opinion and therefore could be objectively coded, responses were categorized by one independent coder.

Results and Discussion

Initial analysis

Initial coding of the data resulted in three response options: consistent with expert opinion, inconsistent with expert opinion, and I don’t know. In order to identify questions for which the distribution of responses did not differ from chance, we ran multiple

Table 1. Percentage of judges’ responses that were consistent with expert opinion, inconsistent with expert opinion or “don’t know” responses

	Consistent	Inconsistent	Don’t know
Estimator variables			
Exposure duration	42.7	53.1	4.2
Alcohol intoxication	96.9	1.0	2.1
Weapon focus	40.2	42.3	17.5
Cross-race bias	61.2	21.4	17.4
Accuracy confidence	64.9	28.9	6.2
Unconscious transference	83.7	11.2	5.1
Trained observers	52.6	45.4	2.1
System variables			
Post-event information	90.6	2.1	7.3
Wording of questions	83.7	16.3	0.0
Child suggestibility	77.3	13.4	9.3
Mugshot-induced bias	45.4	26.8	27.8
Totals	67.2	23.8	9.0

chi-squared analysis with all three response options included for each item. This analysis showed that every distribution differed from chance, $\chi^2(2) \geq 98.0$, $p < 0.001$, Cramer's $\phi = 1.00$. Consistent with Read and Desmarais (2009a), we then removed the "I don't know" responses from our analysis and re-ran the chi-squared analyses. Removal of all "I don't know" responses did little to change the results, with all distributions differing from chance. As can be seen from Table 1, the judges only selected the "I don't know" option on average 9% of the time.

Overall consistency with expert opinion

Judges provided responses that were consistent with expert opinion 67% of the time (see Table 1). Although it is difficult to compare this percentage with those reported in other articles, due to differences in question topics/response options, this percentage does appear to fall within the 19–94% agreement reported by Wise and Safer (2004). This result is similar to that of Granhag *et al.* (2005) in that judges answered two-thirds of the questions posed to them correctly. The lowest percentage of responses that were consistent with expert opinion was 40%, which was in response to the question on mugshot bias. However, mugshot bias also resulted in the largest don't know response, comprising 28% of responses.

Item-specific consistency with expert opinion

Of all the questions asked, judges displayed the highest level of consistency with expert opinion in relation to the effects of alcohol intoxication on a witness at the time of the crime (97% of responses consistent with expert opinion). Judges also demonstrated knowledge that an eyewitness's statement may contain post-event information, with 91% of responses to this question being consistent with expert opinion. This is not surprising, given that these two factors are probably commonly experienced in the courtroom. Judges were less well informed about the effects of exposure duration on memory, weapon focus, and mugshot bias, with less than 50% of responses to these questions consistent with expert opinion.

Within our sample, responses that were consistent with expert opinion ranged from a low of 40% for weapon focus to a high of 97% for alcohol intoxication, showing large variability. It is clear from Table 1 that while the research on certain variables that affect eyewitness memory are being successfully communicated to the Scottish courts, more work needs to be done on topics such as exposure duration, weapon focus, the use of police officers as witnesses, the cross-race effect, mugshot bias and the relationship between confidence in memory and the accuracy of testimony.

Judges beliefs regarding juror knowledge

A further aim of our questionnaire was to assess the beliefs of judges regarding the abilities of the jury. According to the revised *Jury Instruction Manual* published by the Judicial Studies Committee (2012), judges are required to provide directions to a jury regarding the role of the jury, the role of the judge, the verdicts available to them and the innocent until proven guilty code. The judge may also instruct the jury on the evidence presented at hand, but it is entirely up to judges to determine how much and how specific their instructions on the evidence are (Judicial Studies Committee, 2012).

Table 2. Percentage of agree, disagree and neither agree nor disagree responses to the statements regarding the ability of jurors and the use of expert evidence

Statement	Agree	Neither	Disagree
Experts required to provide guidance on the reliability of eyewitness testimony	28.3	14.1	56.6
Experts are not needed for matters of common sense	74.7	11.1	11.1
Eyewitness testimony can be evaluated by common sense alone	72.8	17.2	8.1
Jurors are able to tell accurate from inaccurate eyewitnesses	63.6	25.3	11.1
More training is required in the reliability of eyewitness testimony as evidence	58.6	19.2	21.2

Note. Some of these figures will not total 100% as some questions were left blank.

Therefore, if judges believe that jury members are able to distinguish between an accurate and an inaccurate witness on their own and/or that weaknesses in the testimony of a witness are a matter of common sense, they may not instruct the jury regarding factors that affect the reliability of eyewitness memory.

As can be seen from Table 2, an overwhelming 73% of judges surveyed indicated that the reliability of eyewitness testimony is a matter of common sense, with 75% also responding that experts are not required to inform the court of matters of “common sense”.

Consistency with previous research

Judges sampled for this research provided responses consistent with expert opinion 67% of the time, on average. However, across our questions, judges’ responses demonstrated a large degree of variability in their consistency with expert opinion (ranging from a low of 40% to a high of 97%). Our findings, therefore, are similar to those of Wise and Safer (2004) and Granhag et al. (2005) in that there is a degree of variability in the level of knowledge exhibited by the judges. Taken together, the findings from the current survey add to the message of previous work suggesting that knowledge transfer between judges and academics/researchers is incomplete, the result of which is that judges appear able to maintain specialized knowledge in some, but not all, aspects of eyewitness memory.

Furthermore, the majority of judges also believed that jurors would be able to tell the difference between accurate and inaccurate witnesses on their own. This is at odds with the findings of Wise and Safer (2004), who found that only 46% of US judges responded that they were confident in the abilities of their jury-eligible citizens to recognize factors that affect identification accuracy. Thus, Scottish judges appear more confident in the ability of jury-eligible citizens in Scotland to discriminate between reliable and unreliable eyewitnesses than judges surveyed in the US. We wondered to what extent this confidence was justified and conducted Study 2 to assess the “common sense” beliefs of a jury-eligible general public derived from the same nation as the judges.

STUDY 2

Convictions that originally relied heavily on eyewitness testimony but are now known to have been in error illustrate quite clearly that jurors and legal professionals alike are often unable either to generate or to apply the common sense expected of them

by the courts. For example, Shaw *et al.* (1999) reported that while jury-eligible participants were aware of factors such as length of exposure to the perpetrator, age of witness and delay between encountering the crime and making an identification, they were unaware of how the different test procedures and interview tactics of the police can influence the accuracy of eyewitness testimony.

However, having knowledge of the right response and applying that knowledge appropriately could be conceptualized as rather different tasks. A closer examination of the literature reveals that some methodologies assess the ability of participants to recognize the correct response whilst others assess the ability of participants to apply that knowledge appropriately (for a review, see Read & Desmarais, 2009b). Read and Desmarais (2009b) suggested that the use of different methods may give an inaccurate representation of the level of juror understanding regarding eyewitness testimony evidence. For instance, some surveys have used MC format questions to evaluate juror knowledge (e.g. Deffenbacher & Loftus, 1982; Noon & Hollin, 1987), effectively relying upon the participants to recognize the correct answer. Others have used a more evaluative response format of Likert-type scales with agree–disagree anchors (Kassin & Barndollar, 1992). An evaluative format has also been used in one of the more recent investigations of juror knowledge in the US, which used the Kassin *et al.* (2001) survey items with limited modifications (Benton *et al.*, 2006).

Previous research has shown that self-constructed response formats (where the respondent generates the response unaided) measure higher-order reasoning abilities while MC taps lower-level cognitive processes (such as familiarity or recognition responses) or factual knowledge (Katz, Bennet, & Berger, 2000). Furthermore, MC items can provide unintended hints and these hints may be a potential source of construct-irrelevant variance (Messick, 1989). It is, of course, also the case that questions that require evaluation or inference may promote responses or articulate thinking in ways that would not otherwise occur spontaneously (Chan & Kennedy, 2002).

In the light of these findings, it might be asked whether MC-type surveys of juror knowledge simply overestimate what jurors actually know about eyewitness issues. However, to date, there are no empirical data on this topic (Read & Desmarais, 2009b). Therefore, in order to address this question, the current experiment evaluated mock juror understanding of factors influencing the reliability of eyewitness testimony with both MC and response generation (RG) questionnaires. Our aim was to determine whether jurors could spontaneously generate “common sense” responses in relation to various eyewitness issues – as opposed to relying on potential cues inherent in some (but not all) MC alternatives. To achieve this, we adopted a methodological strategy employed by Chan and Kennedy (2002), who matched MC and RG questions with identical stems such that our potential jurors were required to answer equivalent questions but in different formats. In the MC version, respondents were required to choose their answer from either three or four response alternatives (identical to Study 1), while in the RG condition they were required to produce the answer on their own.

Methods

Participants

A total of 192 potential jurors (81 males and 115 females) completed the survey in Scotland with 96 participants completing the MC questionnaire, and 96 participants

completing the RG questionnaire. Scotland is one of four constituent countries within the UK (the others being England, Wales, and Northern Ireland) and whilst not an independent country, Scotland has its own legal system and Parliament (The National Archives, 2003). Within this jurisdiction, jurors must be between 18 and 65 years of age, be cited on the electoral register and have lived in the UK for a period of at least 5 years since the age of 13 years. In the current study, respondents ranged in age from 18 to 65 years ($M = 37.70$; $SD = 15.24$). Almost all respondents (98%) had obtained some level of educational qualification. Most had completed secondary education (61%), while a further 38% indicated that they had completed undergraduate university-level education. Only 10% of the sample reported previous experience of jury service. All respondents spoke English as their first or main language and were UK residents. It should be noted that jurors in Scotland and the rest of the UK are randomly selected and there is no *voir dire* process.

Survey development and administration

The same MC questionnaire used in Study 1 was also employed in this study. All question topics were the same for both the RG and the MC questionnaires. The stems used for the MC questionnaire were turned into scenarios for the purposes of the RG questionnaire. Thus, for the RG questionnaire, respondents were first of all required to indicate a categorical response to the statement (yes, no, I don't know) and were then asked to explain their answer in free text. For example, "People sometimes witness crimes under the influence of alcohol. If a witness were intoxicated at the time of the crime, would this affect their ability to remember what they saw? Respondents then ticked either a yes/no/I don't know response before moving to the second part of the question: "If yes, what exactly do you think the effect might be and why? If no, why do you think this is the case?"

Researchers managed a volunteer desk at a shopping mall in a busy city centre in Scotland, UK, for a period of 14 days. Respondents were invited to participate on a voluntary basis. Respondents were randomly assigned to complete either the RG or MC version of the questionnaire. The questionnaires were completed individually and respondents took between 20 and 30 min, with the RG questionnaire taking longer on average to complete. On completion, respondents were debriefed and thanked for their contribution.

As in Study 1, responses to the MC questionnaire were coded as either consistent or inconsistent with expert opinion using the coding packet based upon Kassin et al. (2001). Responses for the RG questionnaire were independently coded by the second and third authors as either consistent or inconsistent with expert opinion as identified by the Kassin et al. (2001) survey. Responses that were illegible or in other ways unable to be coded (such as completely off-topic or irrelevant responses) were eliminated as "not codable". Inter-coder correlations indicated a mean agreement of $r = 0.72$. Where there was divergence of opinion between coders, the response was re-analyzed and discussed by both coders until an agreed code was reached. The re-coded responses from both questionnaires were then combined to form a final dataset comprising the following response categories: "consistent with expert opinion", "inconsistent with expert opinion" and "I don't know" responses.

Table 3. Percentage of consistent (C), inconsistent (I) and don't know (DK) responses by questionnaire type.

	Multiple choice			Response generation		
	C	I	DK	C	I	DK
Estimator variables						
Exposure duration	67.7	30.2	2.1	68.8	17.7	13.5
Alcohol intoxication*	91.7	3.1	5.2	62.5	30.2	7.3
Weapon focus*	49.0	42.7	8.3	26.0	57.3	16.7
Cross-race bias	60.4	30.2	9.4	35.4	33.3	31.3
Accuracy confidence	31.3	63.5	4.2	21.9	43.8	18.8
Unconscious transference*	76.0	19.8	4.2	36.5	38.5	25.0
Trained observers*	37.5	48.3	4.2	11.5	76.0	12.5
System variables						
Post-event information	62.5	26.0	11.5	72.9	15.6	11.5
Wording of questions*	66.7	29.2	4.2	75.0	9.4	15.6
Child suggestibility	63.5	24.0	12.5	59.4	25.0	15.6
Mugshot-induced bias	64.6	18.8	16.7	58.3	18.8	22.9
Totals	61.1	31.4	7.5	48.0	33.2	18.7

* $p \leq 0.004$.

Results and Discussion

The aim of the current study was to assess juror knowledge of eyewitness issues elicited by two different questionnaire formats whereby respondents completed either an MC format survey or an RG survey (wherein they were required to generate their own responses). Responses in both surveys were coded as either consistent or inconsistent with documented expert opinion (from Kassin *et al.*, 2001). Overall results are presented in Table 3 (including the "I don't know" responses).

Overall consistency with expert opinion

Participants scored an average of 61% consistency with expert opinion on the MC questionnaire. However, only 48% of participants' responses on the RG questionnaire were consistent with expert opinion (see Table 3). In comparison to the previous literature, both these consistency scores are below the overall two-thirds agreement reported by Desmarais and Read (2011) in their recent meta-analysis.

Item-specific consistency with expert opinion across questionnaire formats

Consistent with Read and Desmarais (2009a), responses to the MC and RG questionnaire formats were analyzed with and without the "I don't know" responses. Inclusion of the "I don't know" responses resulted in significant differences in the consistency of responses with expert opinion between the MC and RG questionnaires on eight topics. However, removal of the "I don't know" response option reduced the number of significant differences to five topics.² Therefore, the more conservative analysis with

² Topics of alcohol intoxication, weapon focus, trained observers, wording of questions, and unconscious transference.

the “I don’t know” responses removed will be reported. Furthermore, owing to multiple comparisons (11), a critical value of $\alpha = 0.004$ was employed.

As can be seen from Table 3, averaged across questionnaire formats and question topics, our community sample provided responses consistent with the expert opinions expressed in Kassin et al. (2001) 55% of the time. However, consistency of response with expert opinion appeared to be associated with questionnaire format for the topics of alcohol intoxication, weapon focus, unconscious transference, trained observers, and wording of questions [$\chi^2(1) \geq 8.0$, $p < 0.004$, Cramer’s $\phi \geq 0.20$] (full statistics for each comparison can be found in Table 4).

On the topic of alcohol intoxication, 92% of responses were consistent with expert opinion in the MC questionnaire and this dropped to 62% when the same question was asked in an RG format. For weapon focus, 49% of MC responses were consistent with expert opinion, compared with 26% of RG responses. The pattern evident for trained observers was similar: there was a higher percentage of responses consistent with expert opinion for the MC questionnaire (37%) than for the RG questionnaire (11%), as was the case for unconscious transference, with 76% of MC responses consistent with expert opinion, compared with 36% of RG responses. However, for the wording of questions topic, the pattern was reversed, with a higher percentage of responses to the RG questionnaire being consistent with expert opinion (75%) than was the case for the MC questionnaire (67%).

Consistency with previous research

Read and Desmarais (2009a) report a 67% consistency rate between a Canadian community sample and items of the Kassin et al. (2001) survey on which experts had reached consensus. The current survey reported a 61% consistency rate with experts. However, this relatively high rate of consistency with expert opinion was only found when participants completed the MC and not the RG version of the survey. For all estimator factors, juror knowledge appeared to be reasonably high across a number of topics when multiple-response alternatives were provided and respondents were simply required to choose what they believed to be the correct response. Yet when participants

Table 4. Comparison of consistent (C) responses by questionnaire type

	MC	RG	χ^2	p	ϕ
Estimator variables					
Exposure duration	67.7	68.8	2.464	.116	.12
Alcohol intoxication*	91.7	62.5	26.403	<.001	.38
Weapon focus*	49.0	26.0	8.402	.004	.22
Cross-race bias	60.4	35.4	3.594	.058	.15
Accuracy confidence	31.3	21.9	0.691	.708	.07
Unconscious transference*	76.0	36.5	16.969	<.001	.32
Trained observers*	37.5	11.5	15.206	<.001	.29
System variables					
Post-event information	62.5	72.9	3.269	.071	.14
Wording of questions*	66.7	75.0	9.567	.002	.23
Child suggestibility	63.5	59.4	0.102	.749	.02
Mugshot-induced bias	64.6	58.3	0.071	.789	.02

MC, multiple choice; RG, response generation.

* $p \leq 0.004$.

were required to spontaneously generate that same knowledge, performance was significantly poorer for over half of the topics.

Although these findings produced rates of agreement with expert opinion that are lower than those established in the previous literature, the discrepancy between MC and RG questionnaire formats was as predicted. Chan and Kennedy (2002) suggested that respondents produce less accurate responses on RG-type questions than on equivalent MC questions because closed questions (such as MC) restrict the frame of reference, focus attention on the available alternatives and allow a deliberate choice, decision or process of deduction (Schuman & Scott, 1987). Bearing this in mind, it is possible that MC surveys overestimate the ability of potential jurors to correctly identify potential threats to eyewitness accuracy when quizzed about their knowledge in an MC format survey. The results of the current study certainly support such an argument. These findings are also consistent with those reported by Alonzo and Lane (2010), who examined the relationship between expressed beliefs regarding eyewitness factors and the ability to use this knowledge when evaluating a witness. Alonzo and Lane (2010) found that survey responses did not predict subsequent evaluations of eyewitness evidence. Therefore, our conclusion is also consistent with previous arguments that the use of questionnaires might overestimate the extent to which potential jurors will apply their knowledge of the vulnerabilities of eyewitness testimony (e.g. Alonzo & Lane, 2010; Cutler, Penrod, & Stuve, 1988; Read & Desmarais, 2009b; Laub & Bornstein, 2008).

GENERAL DISCUSSION

The findings of the current studies reveal that judges and members of the public in the sample jurisdiction shared comparable rates of knowledge regarding factors that could affect the reliability of eyewitness testimony. However, judges appear to overestimate the ability of jurors to distinguish between accurate and inaccurate eyewitnesses and thus overestimate jurors' common sense. In contrast to our findings, Wise and Safer (2004) found that judges who were more knowledgeable about eyewitness testimony were more likely to believe that jurors may have only limited knowledge about factors likely to affect eyewitness accuracy. Consequently, Wise and Safer (2004) strongly recommended the education of judges as a means to reduce the impact of mistaken eyewitness testimony. Encouragingly, 58% of the judges surveyed in Study 1 believed that they required additional training in the area of eyewitness testimony and its reliability as evidence in court. Study 2 identified a number of topics where juror knowledge might be less consistent with expert opinion than previously documented, and this could possibly be due to the response formats used in earlier surveys. However, data from the current survey and sample suggest that there are topics where potential jurors were able to generate a response consistent with that generally agreed upon by experts.

There are, of course, a number of important caveats associated with the current findings. The very nature of the research questions means that there was a fundamental difference in the tasks faced by respondents depending upon which questionnaire they were required to complete. On average, the RG questionnaire took longer to complete and exerted greater task demands on respondents than the MC questionnaire, in that they were required to generate a response (as opposed to selecting a response from

alternatives). As in all survey-based research, it may be that respondents both between and within the MC and RG questionnaire conditions in both studies interpreted the questions in different ways and tailored their responses accordingly, despite our pilot testing to ensure question clarity.

It could also be argued that the participants in the RG condition were at a greater disadvantage than those in the MC condition, as no additional information was available to them to contextualize their responses. Indeed, this criticism could be leveled at surveys of this type in general. Completing a questionnaire is very different from being fully immersed in a court setting in the role of a juror, where rich and detailed contextual (and other case-relevant) material will be readily available. However, it is important to note that both Cutler et al. (1988) and Cutler et al. (1990) also found significant shortcomings in juror knowledge regarding eyewitness issues using a mock juror court-based paradigm. Therefore, it might be argued that enhancing the context does not necessarily stimulate the spontaneous generation of knowledge regarding eyewitness topics.

Secondly, although we developed the current questionnaire using a selection of items from several published sources, including the much-cited Kassin et al. (2001) survey, it is possible that some survey items were problematic in their construction. For instance, Wright (2006), in an examination of the Kassin et al. (2001) survey items, suggests that many of the survey items were presented as associative hypotheses but have the potential to be interpreted as causal. This could be problematic because an alternative interpretation would have resulted in a different response and this could, in particular, have affected our sample of judges. For example, a causal hypothesis about the effect of emotion on memory might state that a highly emotional event will negatively affect eyewitness accuracy and completeness. An expert might qualify this statement by saying it is based on the sample of events included in the Deffenbacher et al. (2004) meta-analysis concerning the effects of high stress on eyewitness memory. However, if an expert is asked whether emotional events are remembered more accurately than non-emotional events, the answer would involve comparing emotional and non-emotional events. The latter involves testing an associative hypothesis. Wright (2006) urged caution with respect to the general acceptance of questionnaire responses, given that it is not certain that respondents interpreted the survey statements consistently. Clearly, the same caution applies to the current studies.

In addition, the more general point made by Read and Desmarais (2009a, 2009b), that presentation of the Kassin et al. (2001) items (which were designed to be answered by experts) to laypersons may similarly introduce difficulties in interpretation, needs to be kept in mind when direct comparisons are made between experts and jurors on identical items. The items herein were modified to ameliorate this issue, but we may not have been entirely successful in doing so. Nonetheless, the eyewitness topics targeted in the current survey were those that exhibited the highest rates of expert agreement in the original survey.

CONCLUSIONS

Overall, the findings of the current set of studies suggest that there is a place within the legal system for researchers to present scientific findings to inform the court: first, to assist in the training of judges and other legal professionals in factors potentially

affecting the accuracy and reliability of eyewitness testimony more generally; and secondly, to present relevant research to a jury in a courtroom. Our findings also suggest that caution should be taken, particularly with respect to methodology, when evaluating the level of knowledge relating to eyewitness topics. In particular, the results suggest that the response format in previous surveys may have overestimated the understanding or knowledge of the vulnerabilities of eyewitness testimony in relation to certain topics. However, the current results also demonstrate that potential jurors can, for some topics, be relied upon to self-generate (rather than recognize) a response consistent with that generally agreed among experts. Previous research suggests that there may be a relationship between jurors' knowledge of eyewitness memory evidence vulnerabilities and the verdicts they render in court (e.g. Neal, Christianson, Bornstein & Robicheaux, 2012). Therefore, an avenue of future research could be to investigate whether responses to MC or RG questionnaires differentially influence the verdicts delivered by judges and/or jurors.

An interesting question raised by these findings is whether simplistic MC-style questionnaires are the best method for assessing expert knowledge against which judge and juror knowledge is compared. An important next step could be to present experts with a questionnaire where they self-generate (rather than recognize) the response; it may very well be that self-generated responses of experts could differ from the established Kassin *et al.* (2001) MC survey. Future research should also focus more closely on the ability of both judges and jurors to *spontaneously* apply a "common sense" understanding of eyewitness issues encountered in court.

REFERENCES

- Alonzo, J. D., & Lane, S. M. (2010). Saying versus judging: Assessing knowledge of eyewitness memory. *Applied Cognitive Psychology, 24*(9), 1245–1264. doi: 10.1002/acp.1626
- Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: comparing jurors, Judges and Law Enforcement to eyewitness experts. *Applied Cognitive Psychology, 20*, 115–129. DOI: 10.1002/acp.1171
- Brigham, J. C., & Wolfskeil, M. P. (1983). Opinions of Attorneys and Law-enforcement personnel on the accuracy of eyewitness identifications. *Law and Human Behavior, 7*, 337–349.
- Chan, N., & Kennedy, P. E. (2002). Are multiple-choice exams easier for Economics students? A comparison of multiple-choice and "equivalent" constructed-response exam questions. *Southern Economic Journal, 68*, 957–971.
- Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identifications: The eyewitness, Psychology, and the Law*. New York, NY: Cambridge University Press.
- Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14*, 185–191. DOI: 10.1007/BF01062972
- Cutler B., Penrod S., & Stuve T. (1988). Juror decision making in eyewitness Identification cases. *Law and Human Behavior, 12*, 41–55.
- Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior, 6*, 15–30. DOI: 10.1007/s10979-004-0565-x
- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28*, 687–706.
- Desmarais, S. L., & Read, J. D. (2011). After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. *Law and Human Behavior, 35*, 200–210. doi: 10.1007/s10979-010-9232-6
- Granhag, P. A., Strömwall, L. A., & Hartwig, M. (2005). Eyewitness testimony: Tracing the beliefs of Swedish legal professionals. *Behavioral Sciences & the Law, 23*, 709–727. DOI: 10.1002/bsl.670
- Judicial Studies Committee (2012). *Jury Manual*. Scotland, UK: Judicial Studies Committee for Scotland.
- Kassin, S. M., & Bardollar, K. A. (1992). The psychology of eyewitness testimony – a comparison of experts and prospective jurors. *Journal of Applied Social Psychology, 22*, 1241–1249.
- Kassin, S., Tubbs, A., Hosch, H. M., & Memon, A. (2001). On the 'general acceptance' of eyewitness testimony research: A new survey of experts. *American Psychologist, 56*, 405–416. DOI: 10.1037//0003-066X.56.5.405

- Katz, I. R., Bennet, R. A., & Berger, A. E. (2000). Effects of response format on difficulty of SAT mathematics items: It's not the strategy. *Journal of Educational Measurement*, *37*, 36–57.
- Laub, C., & Bornstein, B. H. (2008). Juries and eyewitnesses. In Cutler B. L. (Ed.), *Encyclopedia of Psychology and Law* (Vol. 1). California: USA: Sage Publications, Inc. doi: 10.4135/9781412959537
- Lindsay, R. C. L., Wells, G. L., & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses: A replication and extension. *Law & Human Behavior*, *13*, 333–339. DOI: 10.1007/BF01067033
- Lindsay, R. C. L., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations. *Journal of Applied Psychology*, *66*, 79–89.
- Magnussen, S., Wise, R. A., Raja, A. Q., Safer, M. A., Pwelenko, N., & Stridbeck, U. (2008). What judges know about eyewitness testimony: A comparison of Norwegian and US judges. *Psychology, Crime and Law*, *14*(3), 177–188. doi: 10.1080/10683160701580099
- McConkey, K. M., & Roche, S. M. (1989). Knowledge of eyewitness memory. *Australian Psychologist*, *24*, 377–384. DOI: 10.1080/00050068908259576
- Messick, D. M. (1989). Review of decision analysis and behavioral research, by von Winterfeldt & Edwards (Eds.) (book review). *Psychometrika*, *54*, 363–364.
- Neal, T. M. S., Christiansen, A., Bornstein, B. H., & Robicheaux, T. R. (2012). The effects of mock juror's beliefs about eyewitness performance on trial judgments. *Psychology, Crime and Law*, *18*(1), 49–64. doi: 10.1080/1068316X.2011.587815
- Noon, E., & Hollin, C. R. (1987). Lay knowledge of eyewitness behaviour: A British Survey. *Applied Cognitive Psychology*, *1*, 143–153. DOI: 10.1002/acp.2350010207
- Read, J. D., & Desmarais, S. L. (2009a). Lay knowledge of eyewitness issues: A Canadian survey. *Applied Cognitive Psychology* *23*, 301–326. DOI: 10.1002/acp.1459
- Read, J. D., & Desmarais, S. L. (2009b). Expert psychology testimony on eyewitness identification: A matter of commonsense? In Cutler B. L. (Ed.), *Expert Testimony on the Psychology of Eyewitness Identification* (pp. 115–142). New York: USA, Oxford University Press.
- Schuman, H., & Scott, J. (1987). Problems in the use of survey questions to measure public opinion. *Science*, *236*, 957–959.
- Shaw, J. S., III, Garcia, L. A., & McClure, K. A. (1999). A lay perspective on the accuracy of Eyewitness testimony. *Journal of Applied Social Psychology*, *29*, 52–71. DOI: 10.1111/j.1559-1816.1999.tb01374.x
- Sheriffs' Association (2012). *The Office of Sheriff*. Scotland, UK: Judiciary of Scotland.
- Simons, D. J., & Chabris, C. F. (2012). Common (Mis)Beliefs about Memory: A Replication and Comparison of Telephone and Mechanical Turk Survey Methods. *PLoS ONE*, *7*(12): e51876. doi:10.1371/journal.pone.0051876
- The Innocence Project (2012). Understanding the causes: Eyewitness misidentification. Retrieved August 2012 from: <http://www.innocenceproject.org/understand/Eyewitness-Misidentification.php>
- The National Archives (2003). Countries within a country. Retrieved July 2013 from: <http://webarchive.nationalarchives.gov.uk/+http://www.number10.gov.uk/Page823>
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*(2), 45–75.
- Wise, R. A., & Safer, M. A. (2004). What US Judges know and believe about eyewitness testimony. *Applied Cognitive Psychology*, *18*, 427–443. DOI: 10.1002/acp.993
- Wise, R. A., Safer, M. A., & Maro, C. A. (2011). What U.S. law enforcement officers know and believe about eyewitness factors, eyewitness interviews and identification procedures. *Applied Cognitive Psychology*, *25*, 488–500. doi: 10.1002/acp.1717
- Wright, D. B. (2006). Causal and associative hypothesis in psychology: Examples from eyewitness testimony research. *Psychology, Public Policy, and Law*, *12*, 190–213. DOI: 10.1037/1076-8971.12.2.190