APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **22**: 789–801 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1483

Study Space Analysis for Policy Development

ROY S. MALPASS[1]*, COLIN G. TREDOUX[2], NADJA SCHREIBER COMPO[3], DAWN McQUISTON-SURRETT[4], OTTO H. MACLIN[5], LAURA A. ZIMMERMAN[6] and LISA D. TOPP[7]

[1]*University of Texas at El Paso, El Paso, USA*
[2]*University of Cape Town, Rondebosch, South Africa*
[3]*Florida International University, Miami, USA*
[4]*Arizona State University, Glendale, USA*
[5]*University of Northern Iowa, Cedar Falls, USA*
[6]*Klein Associates Division, Applied Research Associates, Fairborn, USA*
[7]*University of Texas at El Paso, El Paso, USA*

SUMMARY

Transforming research findings into policy recommendations requires evaluative criteria beyond traditional academic review. Policy development involves entire literatures, and criteria for examining adequacy of the underlying research as a policy base are needed. At the level of the studies many are obvious: high quality studies, well reported and replicable, consensus on their validity, and ecological validity for application. At the research literature level the distribution of important variables in the literature is important. We discuss policy adequacy criteria and present the *Study Space* concept for evaluating breadth of coverage and gaps in our knowledge in policy research domains. Copyright © 2008 John Wiley & Sons, Ltd.

Lineup construction and administration policies are being scrutinized in jurisdictions across Canada, England, Germany and the United States, stemming from concern for faulty eyewitness identifications and wrongful conviction. Calls for reform of identification procedures have a long history in North America, but recent proposals derive chiefly from two publications: the American Psychology – Law Society (APLS) 'white paper' (Wells et al., 1998) and the National Institute of Justice Research Report 'Eyewitness Evidence: A Guide for Law Enforcement' (Technical Working Group, 1999). Variations of these recommendations have been carried forward in subsequent policy statements by justice organizations (e.g. The Innocence Project, 2005, The Justice Project, 2007). Sequentially presented lineups, together with blind administration have been added to many reform initiatives.

The reform effort raises new questions about the conditions under which a scientific literature can serve as the basis for public policy and implementation of new procedures in the application environment (McQuiston-Surrett, Malpass, & Tredoux, 2006), and occasions questions about the process through which research findings are, or ought to be,

*Correspondence to: Roy S. Malpass, Department of Psychology, University of Texas at El Paso, El Paso, TX 79968, USA. E-mail: rmalpass@utep.edu

transformed into policy. With researchers vigorously advocating for their ideas (Lindsay, 1999; Wells, 2001), the kind of base that policy recommendations have in the empirical literature is worthy of examination. Below we examine criteria for judging the adequacy of a research literature as a policy base and propose a new procedure, *Study Space Analysis*, as an addition to our toolbox for making such assessments.

PRECURSOR

The 1980s produced a series of discussions about the appropriateness of a precursor form of participation in public policy: expert testimony in the courts (Loftus, 1983a,b; McCloskey & Egeth, 1983a,b; McCloskey, Egeth, & McKenna, 1986). This is somewhat removed from the public policy question but there are many similarities, and it is a good place to start. While the discussion contained many dimensions and nuances, Bermant (1986) captures the main thread of the critique:

> 'This article responds to concerns about expert testimony in experimental psychology by conjecturing that disagreements about the propriety of the testimony are camouflaged arguments about the strength of psychological knowledge. Differences between proponents and opponents of expert testimony are about the state of psychological knowledge and certainty, rather than about the proper standard for psychologists to use when deciding whether to testify' (p. 97).

How we determine when the literature is sufficiently strong is important for deciding about expert testimony and also for evaluating the scientific adequacy of a literature for development and implementation of public policy proposals.

While no clear criteria of strength of knowledge emerged from the 1980s era discussions, two central positions have been identified: the *Best Practices* (BP) and *Well Established Knowledge* (WEK) models. The WEK model suggests that the studies forming the research base should be scientifically respectable, well established (Yarmey, 1986), and based on 'settled' science (Kargon, 1986). This implies that: (a) the question to be evaluated has been extensively studied, and (b) matters with disagreement in the research community are inappropriate as a basis for policy recommendations. This is a high criterion position, requiring that policy formation wait until a well established research base is available. No specific criteria for assessing adequacy are available, beyond extensive study and general agreement.

The BP model states that it is acceptable to make conclusions based on the best evidence available at the moment. This approach potentially could be taken as a guide to policy development at any level of empirically based knowledge, in contrast to WEK. However, BP contains no criteria for assessing the strength of the empirical base or the attributes important to examine in assessing some form of adequacy. An implication is that policy recommendations may change fairly rapidly as new work becomes available. If the application environment is not one in which scientists are routinely involved in implementing policy changes in new procedures, their ability to revise and apply the best practice 'snapshot' in light of new findings may be difficult. This implies that policy recommendations should be carefully drawn and their empirical base carefully assessed. It seems appropriate for policy recommendations to come packaged with a caveat indicating

DOI: 10.1002/acp

the strength and stability of the science on which the policy is based. There is a need for the development of criteria for evaluating policy adequacy.

ADEQUACY CRITERIA: HOW MUCH? WHAT KIND?

Whether the field is in a position to offer policy proposals to the justice and law enforcement agencies in the United States depends on what one thinks is required of a scientific literature as a policy basis, and on the frequency and nature of our access to policy revision. If the current knowledge in a field has developed in the absence of studies that include important variables then a best practice statement would be knowably inadequate.

WHAT CRITERIA SHOULD BE USED IN DECIDING THAT A RESEARCH LITERATURE IS AN ADEQUATE BASE FOR POLICY DEVELOPMENT?

Here are some possibilities.

What counts as research?

Deffenbacher, Bornstein, Penrod, and McGorty (2004) said it well as they described their inclusion criteria for their meta-analysis of the effects of stress on eyewitness memory: 'No unpublished studies were included, because the legal standards for proffered scientific testimony established by the U.S. Supreme Court in *Daubert* ... have strengthened the preference by the legal system for meta-analytic conclusions based on a body of well conceived, well executed, and easily retrievable studies.' (pg. 692).

Consistency of findings

Consistency of findings can be examined in a number of ways.

Replication
Research should have been replicated across a range of laboratories and study conditions to avoid confounding construct with method (Campbell & Fiske, 1959). Reporting standards for methods and procedures are important (McQuiston-Surrett et al., 2006).

Meta-analysis
Meta-analysis is an important method of reviewing and summarizing the findings of a research literature, so that the consistency of the findings can be directly and quantitatively measured.

Diversity of methods and techniques

Diversity of methods and techniques can be examined, including independent variable manipulations, operational definitions, and dependent variables, to assess the sufficiency of their coverage of the events and indicators seen in the policy domain.

DOI: 10.1002/acp

Agreement among scientific peers

Agreement among scientific peers can be assessed in a number of ways.

Multi-authored scholarly reviews
These include Wells et al. (1998) and McCloskey, Egeth, and McKenna (1986).

Consensus panels
Examples include the NIJ report (Technical Working Group, 1999) and two generations of policy reviews and implementations on the treatment of child and other vulnerable witnesses in the UK (Home Office in conjunction with Department of Health, 1992; Home Office, 1999, 2000, 2001, 2002).

Surveys of qualified experts
Results of surveys such as Kassin, Ellsworth, and Smith (1989) and Kassin, Tubb, Hosch, and Memon (2001) can provide important information.

The amount of research available

The amount of research available on a topic is an important criterion, but how much is enough?

A single study
Is a single study enough? We cannot think of a policy domain in which a single study would be sufficient.

A series of studies
A series of studies showing the superiority of one policy option over another is a more satisfying basis for policy, but the studies ought to reflect the range and complexity of the policy options studied and important variations in the environment of application.

Comprehensive exploration of the research domain
This sort of strategy would be advantageous but this seems out of reach.

A series of studies would seem to be sufficient, provided the series is large enough for stability of findings, broad enough to cover important variations dictated by theoretical considerations, represents the ecology of the application environment and attends to conventional methodological standards (e.g., counterbalancing of stimulus presentation, disaggregation of confounded variables).

The relation between science and implementation

The relationship between the scientific knowledge base and the nature of the policy application environment has important implications for the completeness and elaboration of scientific knowledge required before specific policies are recommend and implemented. The rationale for this lies in the nature of the application environment.

Expert domains
Expert policy domains are characterized by scientists participating fully and frequently in both the development of the knowledge base that informs policy and its implementation.

DOI: 10.1002/acp

This can be seen in medical policies such as age and frequency for mammograms (Lee, 2002), PAP smears (Wai, Ferrier, Collings, & Laverty, 1996), colonoscopy (Swaroop & Larson, 2002), and changing recommendations concerning interpretation and treatment following prostate specific antigen (PSA) tests (McDougall, Weber, Dziuk, & Heneghan, 2000). Eyewitness identification policy in the UK exhibits many attributes of expert domains.

Customary domains

Some policy domains appear isolated from scientific knowledge. Law enforcement has traditionally employed customary knowledge, and until recently has not contained research and development operations (Geller, 1997). With respect to lineup construction, Wogalter, Malpass, and McQuiston's (2004) findings are consistent with customary systems where technical training is largely unavailable and uncodified practices handed down through generations pass for policy. There is little evidence of ongoing and institutionalised dialog between law enforcement and the scientific knowledge available. Importing policy proposals into an organization operating in a customary environment may not provide a structure conducive to continuing research, development and revision of practice and procedure.

Serial access and policy revision

An expert domain more easily accommodates serial access of science to policy and implementation. The importance of serial access is considerable. If policy is changed outside of normal regulatory processes the welcome mat may not be out for each new round of policy change.

Coverage of important variables

Traditional literature review approaches focus on what has been found rather than on what has been studied. For policy development, however, what has been studied has real importance considered in relation to the attributes and needs of the policy application environment. The idea that a certain amount of research (usually 'more') should be done on a topic before policy recommendations are formed does not guide researchers to the particular questions needing further investigation – nor to the questions on which existing knowledge is inadequate or missing altogether. The WEK model implies consideration of coverage but contains no procedural specifics. Likewise, the BP model extracts recommendations from existing research, but does not measure available knowledge against any standard of coverage, balance or adequacy – it is not part of the BP model to think about the inadequacies of the literature in the context of factors not studied.

Traditional evaluative review is an important tool because it can provide close critical attention to the studies and their cumulative impact. And it generally has as its purpose an assessment of what conclusions are reasonably made on the basis of the existing studies and their methodological and analytical characteristics. Meta-analysis is also an important tool because it is sensitive to the quantitative findings of the literature and can code and analyse the variations among the studies comprising the literature. But the technique itself is necessarily insensitive to variables that have not been studied, or variables that do not appear together in the same studies – essentially, missing data. No tool yet developed has a

DOI: 10.1002/acp

specialized sensitivity to what variables have and have not been studied, their distribution across the literature, and their association or dissociation. A technique for doing just that is described below, and some reasons for using it to characterize a literature for policy development are shown.

THE STUDY SPACE

The idea that a policy-adequate literature should provide extensive coverage of the research domain is related to some basic statistical sampling ideas. Small samples are known to be prone to error as estimates of the distribution of elements in a population. Larger literatures should be more likely to represent the variables of relevance than smaller literatures. However, this assumes that variables measured and manipulated in a research domain are randomly determined. This seems a remote possibility. Researchers appropriately follow their own interests, studying the subsets of variables and effects relevant to their line of investigation and theory development. This is a good thing, and from the perspective of scientific investigation it is not to be constrained except under exceptional circumstances.

Policy development is one of those exceptional circumstances. Examining variables and effects that derive from ecological analysis of the policy environment, or the institutional environment in which policy will be implemented may be forced upon investigators on the way to policy implementation. Systematic and representative coverage of the universe of variables and effects unstudied, under- or over-represented in the research domain can be important in the policy development process.

The study space concept uses the idea of a space in a way similar to the idea of a 'sample space', in which elements are identified, and on the basis of which the probabilities of the elements and subsets of elements can be developed. The elements of the study space are defined by the intersection of the levels of study attributes – independent (or classification) variables, methodological and procedural strategies and dependent (measured) variables used in and across studies. Examining the study space, using the variables, methods and procedures present in an existing literature, can assist in identifying regions of concentration and inattention, alerting investigators to territories that have been well worked over and to others where new contributions can be made. Such an empirical analysis can also help to define confounds and unfortunate strategic decisions, as we will show below.

A generic study space model is shown in Table 1. A study space can be developed at 'low resolution' by using only variable names, while a 'high resolution' study space would contain the levels of variables and specific variations in method and procedural variations at the level of the literature. The study attributes define the dimensions of the space and the frequencies in the cells may be taken to reflect the relative population of regions of the space.

These are the steps in constructing a study space matrix.

1. Identify the studies in the literature of interest, a step shared with narrative and meta-analytic reviews.
2. Construct for each study a matrix showing the independent, method, procedural and dependent variables. Independent variables could come from writing the equation (in the linear model sense) for each of the major dependent variables in the research

DOI: 10.1002/acp

Table 1. Generic study space structure

		IV / CV				MV / PV				DV / MV			
		1	2	.	n	1	2	.	n	1	2	.	n
	1												
IV / CV	2												
	.												
	n												
	1												
MV / PV	2												
	.												
	n												
	1												
DV / MV	2												
	.												
	n												
Where:													
IV / CV = Independent Variables / Classification Variables													
MV / PV = Methodological variables / procedural variables													
DV / MV = Department Variables / Measured Variables													

Where: IV/CV, independent variables/classification variables; MV/PV, methodological variables/procedural variables; DV/MV, department variables/measured variables.

domain. In the eyewitness identification field any 'independent variable' that could reasonably be thought to influence choosing, correct identification and false identification rates would be a reasonable candidate for inclusion.

3. Identify and include in the study matrix constant 'variables'. These are factors that are controlled and held constant in a given study but which may vary across studies. Examples from the eyewitness identification literature are target-suspect similarity, lineup position, target criminality, ecological sampling of faces, stimulus persons and scenarios (MacLin & Malpass, 2002), etc. These cross-study variables should be coded in each study matrix.
4. Enter a '1' in every cell in the matrix corresponding to an intersection of study attributes in each individual study.
5. Merge the individual study matrices into an overall matrix.

The structure in a study space may reveal aspects of knowledge in the research domain related to both adequacy for policy purposes and needs for research and research support and allows us to ask questions that go beyond whether there have been enough studies done on a topic to produce stable cross – study, cross – laboratory results. It allows us to ask whether the distribution of empirical elements includes or omits certain classes of variables, for example lineup size, bias and similarity structure. Quite apart from the size of the study space and the number of elements representing empirical data that populate it, the distribution of empirical elements in the space is a matter of great interest. The size or consistency of a literature matters little if important dimensions of the study space are not adequately represented in the published work. Just doing more studies (or comparing the same variables, in the same labs, with the same researcher's procedural conventions and conceptualisations) will not necessarily improve the quality of a literature as a framework for policy. A few, well designed studies accompanied by a strong theory could lead to as

DOI: 10.1002/acp

much understanding as dozens of not well directed empirical pieces. Study space analysis leads to identifying such studies.

STUDY SPACE ANALYSIS APPLIED TO SIMULTANEOUS AND SEQUENTIAL LINEUPS

A selected subset of independent variables from the studies included in the McQuiston-Surrett et al. (2006) meta analysis is included in the limited study space analysis shown in Table 2.

Some interesting things can be observed in this analysis.

- Low similarity target absent lineups have not been used.
- Simultaneous presentation has never been associated with either backloading or with asking *n* identification questions of the witness.
- Sequential lineups are studied without backloading half as frequently as they are with backloading, but never in association with asking only one identification question.

This reveals a set of confounds in the research design, such that backloading and asking an identification question for each member of a lineup are uniquely associated with sequential lineups. Simultaneous and sequential lineups are not discrete entities. They are packages of variables that take one level for simultaneous lineups and another for sequential lineups.

Simultaneous lineups are usually associated with these administration practices:

- The photos, suspect and fillers, are shown together, simultaneously.
- One omnibus identification question is asked of the entire lineup.
- The photos shown in the display are all there are.

Table 2. A minimalist study space for simultaneous and sequential lineups (dependent variables omitted)

		Backloading		# Questions		Target	Target Absent		
		Yes	No	1	*n*	Present	Hi Sim	Mod. Sim	Low Sim
# Questions	1	**0**	**38**						
	n	**16**	**8**						
Target Present		**9**	**22**	**22**	**14**				
Target Absent	Hi Sim	**4**	**9**	**9**	**4**				
	Mod. Sim	**3**	**10**	**7**	**6**				
	Low Sim	**0**	**0**	**0**	**0**				
Sim. Presentation		**0**	**38**	**38**	**0**	**22**	**9**	**7**	**0**
Seq Presentation		**16**	**8**	**0**	**24**	**14**	**4**	**6**	**0**

Additional studies use a non-backloading procedure with sequential lineups, but are difficult to categorise into this structure because the similarity of the suspect to the culprit is unknown.

DOI: 10.1002/acp

Sequential lineups are usually associated with these administration practices:

- Photos are shown one at a time.
- One identification question is asked for each photo displayed.
- Additional photos are placed after the photos of interest (backloaded).

Apart from the actual presentation mode, then, the two procedures are differentiated on (at least) these two other variables: number of identification questions and backloading. These factors are exchangeable. Backloading can be accomplished for simultaneous lineups by instructing witnesses that they will be shown more than one lineup. Only one of these, normally the first, will contain the suspect. One omnibus question can be asked for sequential lineups by asking witnesses to view a set of photos and then declare their response. As with simultaneous lineups the procedure can be applied in groups of six (or whatever lineup size is being used) to achieve comparability. Asking an identification question for each lineup member can be implemented in simultaneous lineups by identifying the position of each face and asking if that is the offender.

Confounds in research design are troublesome, especially in policy development where cost benefit analyses are also important. We do not know the active ingredient in 'sequential' lineups, and cannot discover it through meta-analysis. Perhaps singling out each face for identification is a powerful procedure. Perhaps criterion manipulation using backloading is a powerful contributor. These have implications for the components of 'sequential lineup' policy. Perhaps sequential presentation is a minor part of 'sequential lineups', but we cannot make that determination until the confound is discovered, and new research is done.

Zimmerman, Malpass, and MacLin (2005) studied this question. They used PC_Eyewitness, a computer program developed to manage eyewitness identification studies (MacLin, Meissner, & Zimmerman, 2005; MacLin, Zimmerman, & Malpass, 2005), to administer (by video) a staged crime experiment using two 'culprits'. The two lineups were constructed using similarity ratings to render them fair in both the size and bias sense (Malpass, Tredoux, & McQuiston-Surrett, 2007). Traditional simultaneous and sequential lineups were used and compared with conditions in which the package variables were exchanged. The traditional simultaneous lineup was not backloaded and witnesses were asked one omnibus identification question. The traditional sequential lineup was backloaded, and one identification question was asked of each face in the sequence. Non-traditional simultaneous lineups were backloaded and one identification question was asked of each face in the display, while non-traditional sequential lineups were not backloaded and witnesses were asked one omnibus identification question. Zimmerman et al. (2005) replicated the finding of higher percentages of correct rejections for traditional sequential culprit absent lineups compared with traditional simultaneous lineups. However there were no differences between non-traditional simultaneous and non-traditional sequential lineups. Further, the percentage of correct rejections of culprit absent lineups dropped significantly from traditional to non-traditional sequential lineups. Their conclusion is that the claimed superiority of sequential lineups is strongly dependent on the package variables, and they dominate the effectiveness of the traditional sequential lineup.

The confound was discovered through a study space analysis, although it must be said that it could easily have been detected beforehand. The total absence of low similarity target absent lineups would not be so easy to discover without a systematic frequency analysis.

DOI: 10.1002/acp

STUDY SPACE ANALYSIS APPLIED TO ALCOHOL EFFECTS ON WITNESS MEMORY

The study space concept can also be applied to legal psychology research areas that currently lack research attention. One such area deals with the effect of alcohol consumption on aspects of eyewitness memory (Table 3). Despite the fact that witnesses are often intoxicated when encountered by the legal system (Evans & Schreiber, 2007; National Institute of Justice, 2003), very little is known about their cognitive capabilities as eyewitnesses. Two published studies have investigated the effects of intoxication on eyewitness identification performance (Dysart et al., 2002; Yuille & Tollestrup, 1990), one has examined eyewitness recall (Yuille & Tollestrup) and one has examined the effects of alcohol intoxication on suspects' subsequent eyewitness identification and recall performance (Read, Yuille, & Tollestrup, 1992).

What are the most urgent research questions within these areas? And, outside of these areas, where else might the effects of intoxication emerge in the criminal justice system (e.g., in suspect interrogations or jury decision making)? Here, the dimensions of the study space are both determined by the existing literature and guided by policy concerns and legal psychology theory and research.

Although a survey of eyewitness experts revealed that 90% felt eyewitness memory impairment due to alcohol intoxication is reliable enough for courtroom testimony (Kassin, Tubb, Hosch, & Memon, 2001), application of the study space concept indicates that most variables have in fact never been studied. It also becomes apparent that no central eyewitness concept (e.g. TA vs. TP identification) has been empirically addressed in combination with intoxication in more than one study.

Table 3. Study space for alcohol and eyewitness memory

A. INDEPENDENT VARIABLES														
1. INTOXICATION LEVEL	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	2.6	3.1	3.2	3.3
1.1 Alcohol , unvaried														
1.2 moderate														
1.3 extreme intoxication														
1.4. Placebo	1													
1.5 Sober	2			2										
2.1 EYEWITNESS ID: TP (Lineup)	1			1	1									
2.2 TA (Lineup)	1			1	1									
2.3 Sequential														
2.4 Simultaneous	1			1	1									
2.5 TP (Showup)	1				1									
2.6 TA (Showup)	1				1									
3.1 WITNESS INT: Recall format														
3.2 Recall delay	1			1	1									
3.3 Misinformation				1	1									
B. DEPENDENT VARIABLES														
4 EYEWITNESS IDs														
4.1 Correct IDs	2			1	2	1	1		1	1	1			
4.2 Filler IDs	2			1	2					1	1			
4.3 No IDs												1		
4.4 ID confidence	1			1	1									
5 WITNESS INTERVIEWING														
5.1 Recall accuracy	1			2	2									
5.2 Recall quantity	1			1	1									
5.3 Recall confidence				1	1									
5.4 Ratings of Credibility														

DOI: 10.1002/acp

Study space analysis reveals that the empirical investigation of the role of intoxication in eyewitness memory has (a) neglected a major legal psychology research area, (b) is not meeting the needs of the 'real world', (c) is not meeting the requirements of 'replicated research', and (d) is not meeting the requirements of 'cross-laboratory' research/ investigation.

CONCLUSION

Study space analysis differs from both narrative review and meta-analysis. Meta-analysis is limited to the studies at hand, while study space analysis deals also with studies that have not been done, identified by empty combinations of variables. A study space analysis identifies the variables of potential importance that could be/need to be studied for some policy (or other) purpose, and makes visible those things that have eluded study.

Study space analysis provides a guide for further research, and allows an investigator to target specific areas and relationships that are neglected in the current distribution of research. Targeted studies can be identified and conducted so that important questions not yet addressed can be examined prior to the formation of policy recommendations. The study space can be assessed and 'filled in' before claims are made that the topic of interest is mature and well understood. This enhances credibility of researchers and buffers against 'best practices recommendations' that change over short periods of time and 'well established knowledge' that has been blind to important questions and failures of the existing literature. The residual between the relationships that have been studied and those that have not may be large or small, and may or may not be important, but a residual will probably always be present because no research literature is likely to exhaust its domain. However, assessing the importance of this residual is more easily accomplished through techniques that render it visible.

The science will continue to grow and change. New perspectives will appear. Old theories will be applied in new ways. Things we used to know will be discarded. The answers we have to important questions will be refined, and may even change direction. The point to attack, then, is not whether psychologists should contribute their knowledge for application. Some procedures for making identifications and offering the results as evidence will be used. The question is, will they be the product of folklore and historical practice or will they be based on empirical research. We think the latter is preferable. But then how does practice keep up with changing knowledge? There are many items in existing policy proposals and consensus panel documents that are based on stable science and which have been shown to be effective on the basis of collections of research without important confounds in research design and unstudied combinations of variables. These should be the focus of policy development. Others will come into focus when the policy adequacy of their respective empirical literatures mature and cover the important issues.

The study space concept offers a methodology for revealing unstudied questions in research domains, and may be of use to scientists working in policy areas. The study space takes on special importance when the serial access to policy development is not an institutionalised avenue of change. One can always expect progress and reframing of policy options to occur. In order to cope with this, serial access is important. If it is not available then the responsibilities are very great for policy framers to get it right the first time, based on a policy-adequate literature that represents the study space.

DOI: 10.1002/acp

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation through grant number SBR—9730937. We thank Kim Gaitens for her contributions to the manuscript.

REFERENCES

Bermant, G. (1986). Two conjectures about the issue of expert testimony. *Law and Human Behavior: The Ethics of Expert Testimony*, *10*, 97–100.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

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.

Dysart, J. E., Lindsay, R. C. L., MacDonald, T. K., & Wicke, C. (2002). The intoxicated witness: Effects of alcohol on identification accuracy from show-ups. *Journal of Applied Psychology*, *87*, 170–175.

Evans, J., & Schreiber, N. (2007). Intoxicated witness and victims of crime: How common are they? Poster presented at off the witness stand: Using psychology in the practice of justice, John Jay College of Criminal Justice, New York City.

Frye v. U.S., 293F. 1013, 34 A.L.R. 145 (App D.C. 1923).

Geller, W. A. (1997, December) Suppose we were really serious about police departments becoming ''learning organizations''? *National Institute of Justice Journal*, *234*, 2–7.

Home Office. (1999). Action for justice: Implementing the speaking up for justice report on vulnerable or intimidated witnesses in the criminal justice system in England And Wales. Home Office Communication Directorate: London, UK http://www.homeoffice.gov.uk/cpd/pvu/pvu.htm.

Home Office. (2000). Consultation paper. Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children. Communication Directorate: London, UK. http://www.homeoffice.gov.uk/cpd/pvu/pvu.htm.

Home Office. (2001). Vulnerable Witnesses: A Police Service Guide. Home Office Communication Directorate: London, UK http://www.homeoffice.gov.uk/cpd/pvu/pvu.htm.

Home Office. (2002). Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children. Home Office Communication Directorate: London, UK http://www.homeoffice.gov.uk/cpd/pvu/pvu.htm.

Home Office in conjunction with Department of Health. (1992). *Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings*. London: HMSO.

Kargon, R. (1986). Expert testimony in historical perspective. *Law and Human Behavior*, *10*, 15–27.

Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The ''general acceptance'' of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist*, *44*, 1089–1098.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the ''general acceptance'' of eyewitness testimony research. *American Psychologist*, *56*, 405–416.

Lee, C. H. (2002). Screening mammography: Proven benefit, continued controversy. *Radiological Clinics of North America*, *40*, 395–407.

Lindsay, R. C. L. (1999). Applying applied research: Selling the sequential line-up. *Applied Cognitive Psychology*, *13*, 219–225.

Loftus, E. F. (1983a). Silence is not golden. *American Psychologist*, *38*, 564–572.

Loftus, E. F. (1983b). Whose shadow is crooked? *American Psychologist*, *38*, 576–577.

MacLin, O. H., & Malpass, R. S. (2002). Towards an ecological approach in examining the diagnosticity of sequential and simultaneous line-up presentations. Paper presented at the meetings of the American Psychology – Law Society, Austin, TX.

MacLin, O. H., Meissner, C. A., & Zimmerman, L. A. (2005). PC_Eyewitness: A computerized framework for the administration and practical application of research in eyewitness psychology. *Behavior Research Methods*, *37*, 324–334.

MacLin, O. H., Zimmerman, L. A., & Malpass, R. S. (2005). PC_Eyewitness and the sequential superiority effect: Computer based lineup administration. *Law and Human Behavior*, *29*, 303–321.

DOI: 10.1002/acp

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. E. (2007). Lineup construction and lineup fairness. In R. Lindsay, D. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology memory for people* (Vol. 2, pp. 155–178). Mahwah, New Jersey: Lawrence Erlbaum & Associates.

McCloskey, M., & Egeth, H. E. (1983a). Eyewitness identification: What can a psychologist tell a jury. *American Psychologist*, *38*, 550–563.

McCloskey, M., & Egeth, H. E. (1983b). A time to speak, or a time to keep silence? *American Psychologist*, *38*, 573–575.

McCloskey, M., Egeth, H. E., & McKenna J. (Eds.). (1986). The ethics of expert testimony. Law and Human Behavior, 10.

McDougall, G. J., Weber, B. A., Dziuk, T. W., & Heneghan, R. (2000). The controversy of prostate screening. *Geriatric Nursing*, *21*, 245–248.

McQuiston-Surrett, D., Malpass, R. S., & Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of methods, data and theory. *Psychology, Public Policy and Law*, *12*, 137–169.

National Institute of Justice. (2003). Sourcebook of criminal justice statistics (31st ed.). Washington, DC: Author.

Read, J. D., Yuille, J. C., & Tollestrup, P. (1992). Recollections of a robbery: Effects of arousal and alcohol upon recall and person identification. *Law and Human Behavior*, *16*, 425–446.

Swaroop, V. S., & Larson, M. V. (2002). Colonoscopy as a screening test for colorectal cancer in average-risk individuals. *Mayo Clinic Proceedings*, *77*, 951–956.

Technical Working Group. (1999). Eyewitness evidence: A guide for law enforcement. U.S. Department of Justice, National Institute of Justice. (Document NCJ 178240) Washington DC. Retrieved 3/18/08 from http://www.ojp.usdoj.gov/nij/pubs-sum/178240.htm.

The Innocence Project. (2005). Model Legislation, 2005 State Legislative Sessions. Retrieved 12/26/05 from http://www.innocenceproject.org/docs/Model_Statute_Eyewitness_ID.pdf.

The Justice Project. (2007). Eyewitness Identification – A Policy Review. Washington, D.C.: The Justice Project.

Wai, D., Ferrier, A., Collings, S., & Laverty, C. (1996). Have the most recent Pap smear guidelines affected GP practices? *Australian Family Physician*, (Suppl 1), S44–S48.

Wells, G. L. (2001). Police lineups: Data, theory, and policy. *Psychology, Public Policy, And Law*, *7*, 791–801.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *23*, 603–647.

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of police on preparation and conduct of identification lineups. *Psychology, Crime and Law*, *10*, 69–82.

Yarmey, A. D. (1986). Ethical responsibilities governing the statements experimental psychologists make in expert testimony. *Law and Human Behavior*, *10*, 101–116.

Yuille, J. C., & Tollestrup, P. (1990). Some effects of alcohol on eyewitness memory. *Journal of Applied Psychology*, *75*, 268–273.

Zimmerman, L. A., Malpass, R. S., & MacLin, O. H. (2005). The Simultaneous v. Sequential Lineup Comparison is Confounded: Fair Comparison Eliminates Sequential Superiority. Unpublished manuscript, University of Texas at El Paso.

DOI: 10.1002/acp