Costs and Benefits of Eyewitness Identification Reform

Steven E. Clark

Department of Psychology
Presley Center for Crime and Justice Studies
University of California, Riverside

- Trade-off between false identifications avoided (benefits) and correct identifications lost (costs)
- Blind lineup administration
- Confidence and accuracy

Eyewitness Identification Reform

1. Instruct the witness that the perpetrator may *or may not* be in the lineup. (unbiased instructions)
2. Present the lineup sequentially rather than simultaneously
3. Present a fair, unbiased lineup: The foils should be selected so that the suspect does not stand out.
4. Lineup administrator should not influence the witness (blind lineup administration).

THE FUNDAMENTAL CLAIM: These reforms increase the accuracy of eyewitness identification

- Accuracy = A high correct identification rate (of suspects who are guilty) and a low false identification rate (of suspects who are innocent).

These reforms increase the accuracy of eyewitness identification

- Strong version (No Cost Claim). Reforms increase accuracy, either by reducing the false identification rate, with little or no loss of correct identifications, or by increasing the correct identification rate, with little or no increase in false identifications.

	CORRECT	FALSE
NON-RECOMMENDED	.50	.20
RECOMMENDED	.50	.10

These reforms increase the accuracy of eyewitness identification

- Strong version (No Cost Claim). Reforms increase accuracy, either by reducing the false identification rate, with little or no loss of correct identifications, or by increasing the correct identification rate, with little or no increase in false identifications.

	CORRECT	FALSE
NON-RECOMMENDED	.50	.10
RECOMMENDED	.70	.10

These reforms increase the accuracy of eyewitness identification

- Weak version (Low cost claim). The reforms have an effect on both correct and false identification rates, but the effects are disproportional, thus increasing overall accuracy.

	CORRECT	FALSE
NON-RECOMMENDED	.50	.20
RECOMMENDED	.48	.10

No Cost Claim
(Strong Accuracy Claim)

- Recommended procedures produce only benefits and no costs.
- The policy decision is uncomplicated.
- The recommended procedures are objectively correct *and policy is specified by the data.*
- To not implement the recommended procedures would be irrational.

The No Cost View is widely held.

- By researchers,
- Legal scholars
- Policy-makers
- Textbooks
- Popular media

The No Cost View is widely held.

- Researchers:
  - "We have taken great care to recommend procedures that do not serve to reduce the chances that the guilty ... will be identified." Wells et al. (1998).
  - "...decades of laboratory research showing that the sequential procedure reduces mistaken identifications with little or no reduction in accurate identifications." (Wells et al., 2011)

Basis of the No Cost Claim

- Early data, some misintrepreted data, and a theory (based on a distinction between absolute and relative judgments) that appeared to account for it.

Signal Detection Theory

- Correct and false identification rates should covary.

Data

	CORRECT	FALSE	d’
Biased Ins.	.59	.15	1.40
Unbiased Ins.	.50	.09	1.38

	CORRECT	FALSE	d’
Simultaneous	.54	.15	1.27
Sequential	.43	.09	1.24

	CORRECT	FALSE	d’
More influence	.58	.21	1.14
Less Influence	.45	.11	1.25

	CORRECT	FALSE	d’
Less Similar	.67	.31	1.09
More Similar	.59	.16	1.36

	CORRECT	FALSE	d’
Susp. Match	.46	.07	1.38
Desc. Match	.53	.15	1.11

Reconciling No Cost and Low Cost claims with extant data

- Sequential > Simultaneous: Steblay et al. (2011)
- Sequential = Simultaneous: Clark (2012); Palmer & Brewer (2012)
- Sequential < Simultaneous: Mickes et al. (2012); Gronlund et al. (2012); Dobolyi & Dodson (2013)
  - Analysis of the wrong data
  - Statistical artifacts and problems
  - Inclusion and exclusion criteria
  - Selective publication

Focusing on the wrong data

Psychology, Public Policy, and Law
2011, Vol. 17, No. 1, 99–139

1076-8971/11/$12.00 DOI: 10.1037/a0021650

SEVENTY-TWO TESTS OF THE SEQUENTIAL LINEUP SUPERIORITY EFFECT: A Meta-Analysis and Policy Discussion

Nancy K. Steblay
Augsburg College

Jennifer E. Dysart
John Jay College of Criminal Justice

Gary L. Wells
Iowa State University

A decade ago, a meta-analysis showed that identification of a suspect from a sequential lineup versus a simultaneous lineup was more diagnostic of guilt (Steblay, Dysart, Fulero, & Lindsay, 2001). Since then, controversy and debate regarding sequential superiority has emerged. We report the results of a new meta-analysis involving 72 tests of simultaneous and sequential lineups from 23 different labs involving 13,143 participant-witnesses. The results are very similar to the 2001 results in showing that the sequential lineup is less likely to result in an identification of the suspect, but also more diagnostic of guilt than is the simultaneous lineup. An examination of the full diagnostic design dataset (27 tests that used the full simultaneous/sequential × culprit-present/culprit-absent design) showed that the average gap in correct identifications favoring the simultaneous lineup over the sequential lineup—8%—is smaller than the 15% figure obtained from the 2001 meta-analysis (and from the current full 72-test dataset). The lower error rate incurred for culprit-absent lineups with use of a sequential format remains consistent across the years, with 22% fewer errors than simultaneous lineups. A Bayesian analysis shows that the posterior probability of guilt following an identification of the suspect is higher for the sequential lineup across the entire base rate for culprit presence/absence. New ways to think about policy issues are discussed.

Statistical Artifacts and Problems

- Ceiling effects in Target-Present (guilty suspect) lineups (Clark, 2005)
- Measurement of accuracy
  - Diagnosticity Ratio of Correct and False ID rates
  - ROC analyses
  - d'
  - Wixted: Correlation between d' and pAUC is stronger than the correlation between C/F ratio and pAUC.

Inclusion and Exclusion Criteria

- Included by SDW (2011); Excluded by Clark (2012) and by Palmer & Brewer (2012)
  - Lindsay et al. (1991). Sim/Seq comparison confounded with biased instructions and biased lineup composition.
- Excluded by SDW (2011); Included by Clark (2012)
  - Haw & Fisher (2004)
  - Smith et al. (2001)
  - Gronlund et al. (2009)*
  - Douglass & McQuiston-Surrett (2006)
  - Steblay et al. (2011)
- Lindsay Lab Effect (McQuiston-Surrett et al. 2006)

Selective Publication

- Steblay, Dysart, & Wells (2011)
  - Comparing sequential and simultaneous lineups

	CORRECT IDENTIFICATIONS	FALSE IDENTIFICATIONS
UNPUBLISHED	-.17	-.19
PUBLISHED	-.10	-.25

from Clark, Moreland, & Gronlund 2014

Long-run Expected Utility

- Base rates. How often are innocent suspects presented to witnesses?
  - 1/100?
  - 20/100?
- Costs and benefits associated with outcomes. What are the utilities and disutilities associated with eyewitness identification outcomes?
  - What is the cost of a false identification error?
  - What is the cost of a false non-identification error?

Expected Utility

$$E[U] = [p(CID)u(CID) - p(FN)u(FN)]\, p(g) + [p(CN)u(CN) - p(FID)u(FID)]\, [1-p(g)]$$

$$E[U]_{SIM} = [p(CID_{SIM})u(CID) - p(FN_{SIM})u(FN)]\, p(g) + [p(CN_{SIM})u(CN) - p(FID_{SEQ})u(FID)]\, [1-p(g)]$$

$$E[U]_{SEQ} = [p(CID_{SEQ})u(CID) - p(FN_{SEQ})u(FN)]\, p(g) + [p(CN_{SEQ})u(CN) - p(FID_{SEQ})u(FID)]\, [1-p(g)]$$

Key: CID = Correct ID (guilty suspect), CN= Correct non-ID (of innocent suspect)
FID = False ID (of innocent suspect), FN = False non-ID (of guilty suspect)

Ceci & Friedman
(Cornell Law Review, 2000)

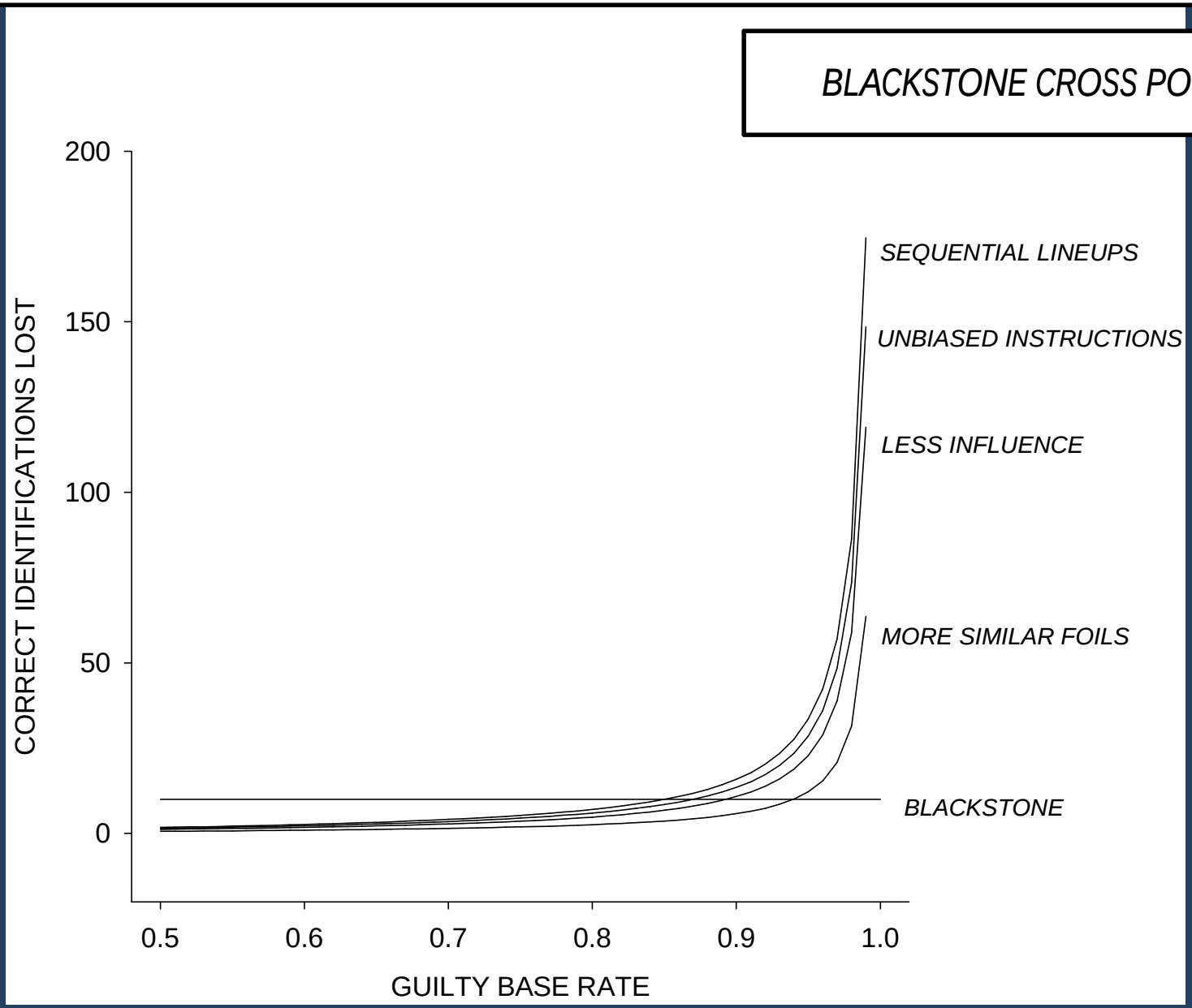
For the comparison of two lineup procedures, A and B,
where the correct and false identification rates are lower for B than for A,

Procedure B should be preferred over Procedure A if the following inequality holds:

$$\frac{p(FN_B)p(G) - p(FN_A)p(G)}{p(FID_A)p(I) - p(FID_B)p(I)} < \frac{u(CN) - u(FID)}{u(CID) - u(FN)}$$

$$\frac{\text{The loss of correct identifications, for B - A}}{\text{The decrease in false identifications, for A - B}} \qquad \frac{\text{the cost of a false ID}}{\text{the cost of a false non-ID}}$$

Number of correct identifications lost in exchange for each false identification avoided



Blackstone (1769). "It is better that ten guilty men escape than that one innocent suffer."

- Utility analyses force policy assumptions into the daylight.
- “The” guilty base rate is unknown.
- How should base rates contribute to policy?
  - Wells (2006) has argued that police should implement procedures to increase the guilty base rate and also implement more conservative identification procedures (i.e., sequential lineup).
  - However, that combination might put criminal justice at the “costly” end of the utility function.
  - Other problems…

Blind Lineup Administration

Blind Lineup Administration

- *If one is concerned that police might deliberately or inadvertently leak their expectations regarding the lineup, a reasonable solution is to prevent the police from having expectations, a solution that would be achieved through blind lineup administration. (Clark, 2012)*

Blind Lineup Administration

NJ v. Henderson (2011)

- [A] non-blind procedure can affect the reliability of a lineup because even the best-intentioned, non-blind administrator can act in a way that inadvertently sways an eyewitness trying to identify a suspect. An ideal lineup administrator, therefore, is someone who is not investigating the particular case and does not know who the suspect is.
- We find that the failure to perform blind lineup procedures can increase the likelihood of misidentification.

Data

Blind versus Non-Blind Lineup Administration

Published

- Greathouse & Kovera (2009), *Law & Human Behavior (ssci 19, gs 57)*
- Phillips , McCauliff, Kovera, & Cutler (1999) *Jo. Applied Psychology (ssci 53, gs 120)*
- Perlini & Silvaggio (2007) *Psych Reports (ssci 1,gs 4)*

Unpublished

- Beaudry (2008)
- Dysart & Fugal (2006)
- Dysart et al. (2008)
- Haw, Mitchell, & Wells (2003)
- Russano, Dickinson, Cass, Kovera & Cutler (2002)

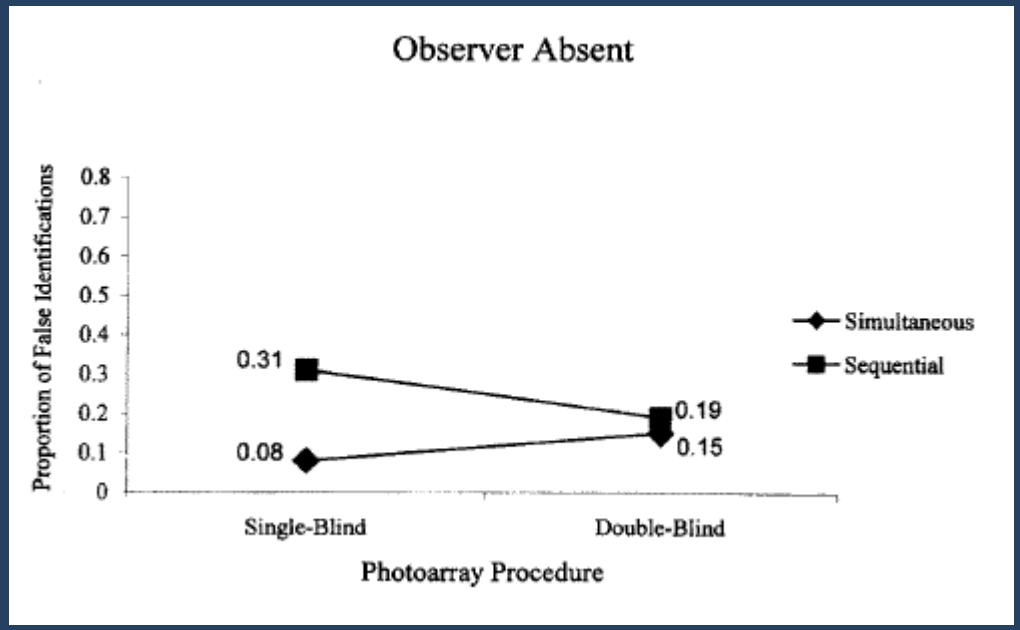
Blind Lineup Administration

- Blind lineup administration will increase the rate of false identifications, decrease the reliability of eyewitness identification evidence, and increase the rate of false convictions.

Table 2 Proportion of identification decisions by target presence, instruction bias, lineup presentation, and administrator knowledge

	Target-present		Target-absent		Collapsed across target presence	
	Simultaneous	Sequential	Simultaneous	Sequential	Simultaneous	Sequential
Identifications of suspects						
Biased Instructions						
Single-blind	.86 ($n = 14$)	.57 ($n = 14$)	.33 ($n = 15$)	.21 ($n = 14$)	.60 ($n = 29$)	.39 ($n = 28$)
Double-blind	.64 ($n = 14$)	.50 ($n = 14$)	.00 ($n = 15$)	.07 ($n = 15$)	.32 ($n = 29$)	.28 ($n = 29$)
Unbiased Instructions						
Single-blind	.47 ($n = 15$)	.79 ($n = 14$)	.14 ($n = 14$)	d'= 1.005	.31 ($n = 29$)	.46 ($n = 29$)
Double-blind	.43 ($n = 14$)	.56 ($n = 16$)	.19 ($n = 16$)	d= 0.702	.31 ($n = 30$)	.32 ($n = 31$)
Foil identifications						
Biased Instructions						
Single-blind	.14 ($n = 14$)	.36 ($n = 14$)	.47 ($n = 15$)	.64 ($n = 14$)	.31 ($n = 29$)	.50 ($n = 28$)
Double-blind	.29 ($n = 14$)	.43 ($n = 14$)	.87 ($n = 15$)	.93 ($n = 15$)	.59 ($n = 29$)	.69 ($n = 29$)
Unbiased Instructions						
Single-blind	.27 ($n = 15$)	.14 ($n = 14$)	.79 ($n = 14$)	.67 ($n = 15$)	.52 ($n = 29$)	.41 ($n = 29$)
Double-blind	.43 ($n = 14$)	.38 ($n = 16$)	.81 ($n = 16$)	.67 ($n = 15$)	.63 ($n = 30$)	.52 ($n = 31$)
Rejections of the lineup						
Biased Instructions						
Single-blind	.00 ($n = 14$)	.07 ($n = 14$)	.20 ($n = 15$)	.14 ($n = 14$)	.10 ($n = 29$)	.11 ($n = 28$)
Double-blind	.07 ($n = 14$)	.07 ($n = 14$)	.13 ($n = 15$)	.00 ($n = 15$)	.10 ($n = 29$)	.03 ($n = 29$)
Unbiased Instructions						
Single-blind	.27 ($n = 15$)	.07 ($n = 14$)	.07 ($n = 14$)	.20 ($n = 15$)	.17 ($n = 29$)	.14 ($n = 29$)
Double-blind	.14 ($n = 14$)	.06 ($n = 16$)	.00 ($n = 16$)	.27 ($n = 15$)	.07 ($n = 30$)	.16 ($n = 31$)

from Greathouse & Kovera (2009)



Phillips et al. 1999

Blind Lineup Administration

(NJ Pattern Jury Instruction)

- *A lineup administrator who knows which person or photo in the lineup is the suspect may intentionally or unintentionally convey that knowledge to the witness. That increases the chance that the witness will identify the suspect, even if the suspect is innocent. For that reason, whenever feasible, live lineups and photo arrays should be conducted by an officer who does not know the identity of the suspect.*
- *In this case, it is alleged that the person who presented the lineup knew the identity of the suspect. It is also alleged that the police did ... not compensate for that by conducting a procedure in which the officer did not see the photos as the witness looked at them.*
- *You may consider this factor when you consider the circumstances under which the identification was made, and when you evaluate the overall reliability of the identification.*

Blind Lineup Administration

- Potentially a dangerous one-two punch:
  - Blind lineup administration may not actually reduce false identifications or increase accuracy.
  - But jury instructions tell jurors that it does.
  - There may be more (not fewer) false identifications, but jurors will put more trust in them.
  - More false IDs + Greater Trust = More False Convictions

problems with experimental comparisons of blind and non-blind lineup administration

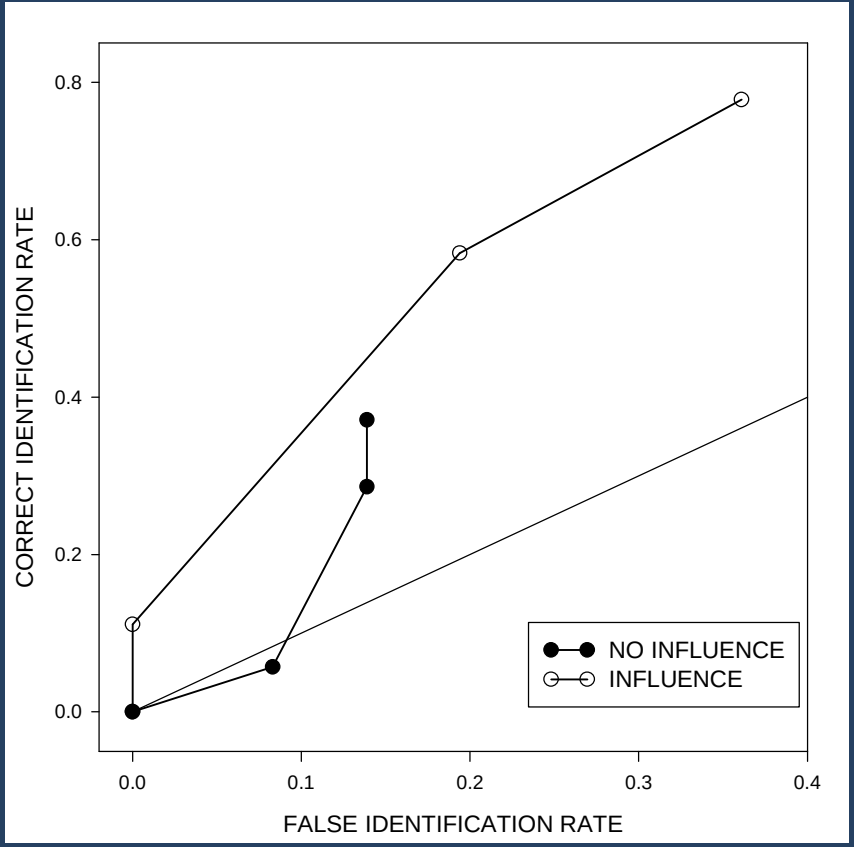
- Very small n's (14-16 subjects per cell, Greathouse & Kovera, 2009).
- Unskilled lineup administrators – who may have incentives to obtain suspect identifications, but may not know how to obtain them.

Trained Manipulative Lineup Administrators

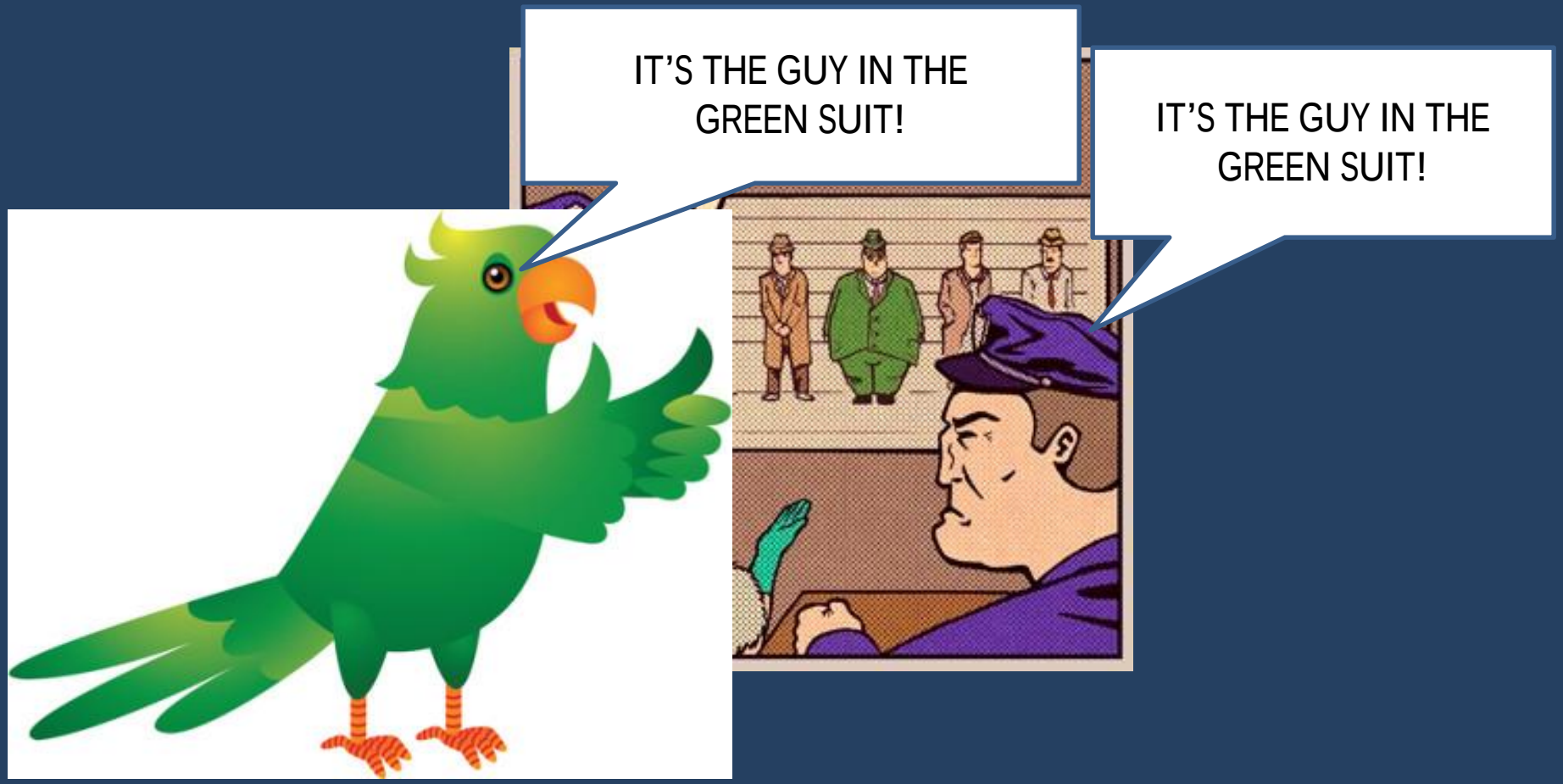
Clark, Brower, Rosenthal, Hicks, & Moreland (2013), *JARMAC*

- The witness…
  - mentions the suspect:
    - *That's an ID. Period.*
  - mentions a foil (tentatively):
    - Clarification: *Are you saying that number two is the person who you saw commit the crime, or are you saying that number two looks similar …?*
  - is non-responsive.
    - *Take your time … look at each photograph carefully…*
    - *Anyone in the lineup look more like him than anyone else?*

	Correct	False	d’
No Influence	.36	.14	0.72
Influence	.78	.36	1.13



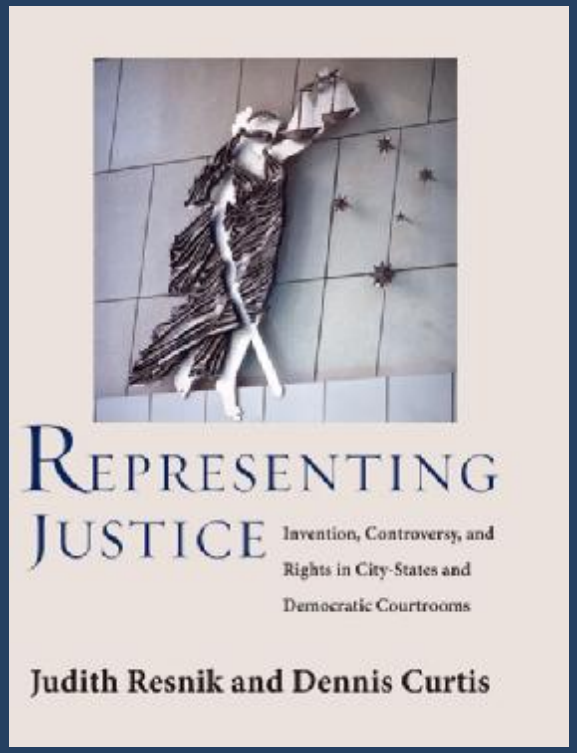
I’m not advocating witness nudging, manipulation, and steering.



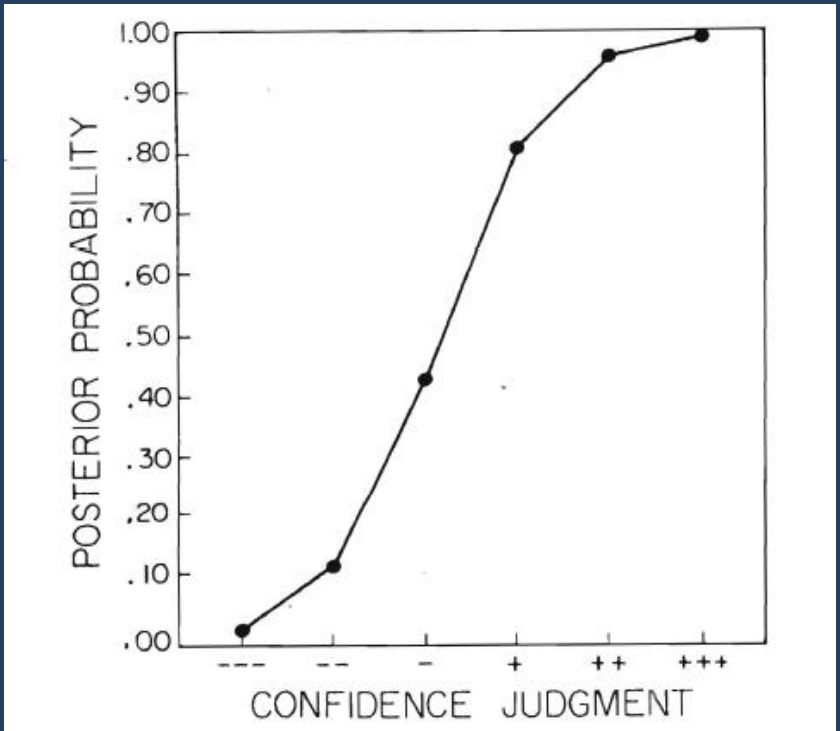
However,

- The interactions between witnesses and lineup administrators are likely to involve a complex mix of memory compliance and memory retrieval – and there is a lot that we don't know.
- If we are going to make a policy recommendation about blind lineup administration, we should be able to indicate what the likely outcome of that policy will be.

Normative foundations for blind lineup administration



Confidence and Accuracy



Murdock, 1974

- "When one is correct 975-995 times in 1000, one comes to feel that one can believe the subject when he is confident." (Murdock, 1974, p. 33)

Confidence and Accuracy

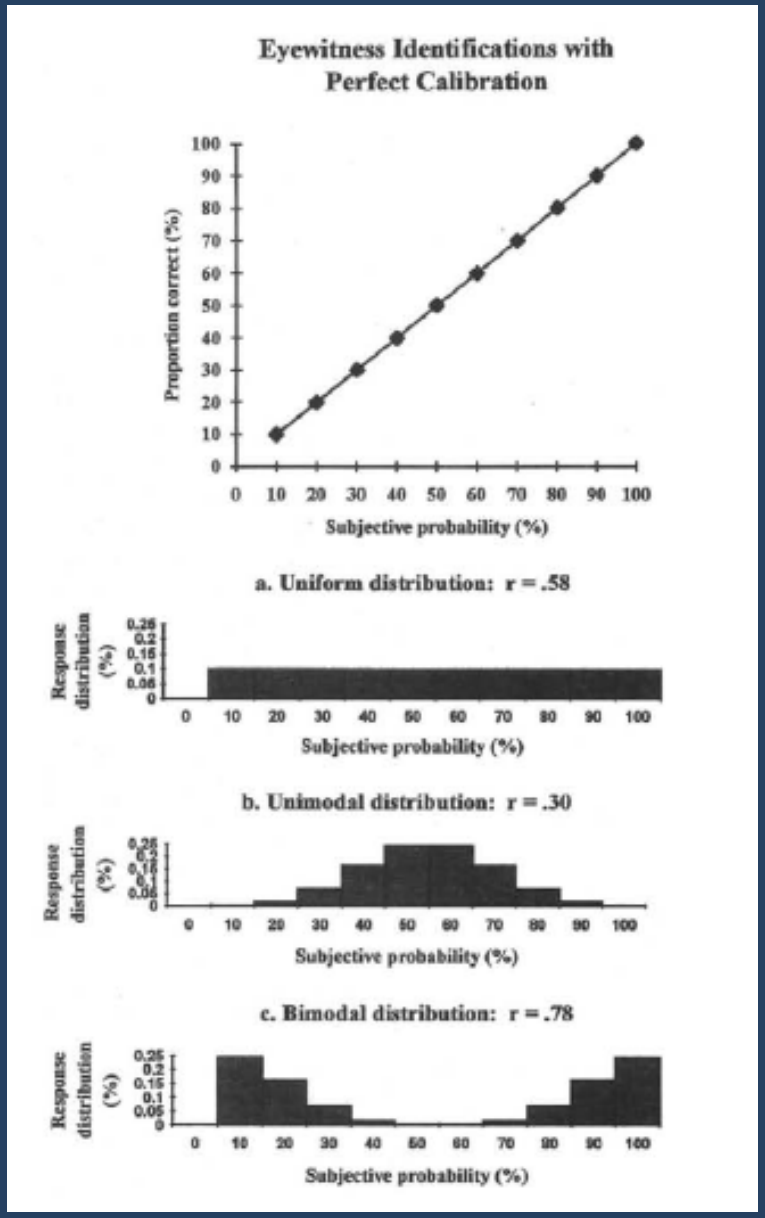
- Confidence "is a weak indicator of eyewitness accuracy even when measured at the time an ID is made and under relatively "pristine" laboratory conditions" Penrod & Cutler (1995, p. 830).
- "...of limited utility ..." (Wells & Quinlivan 2009).

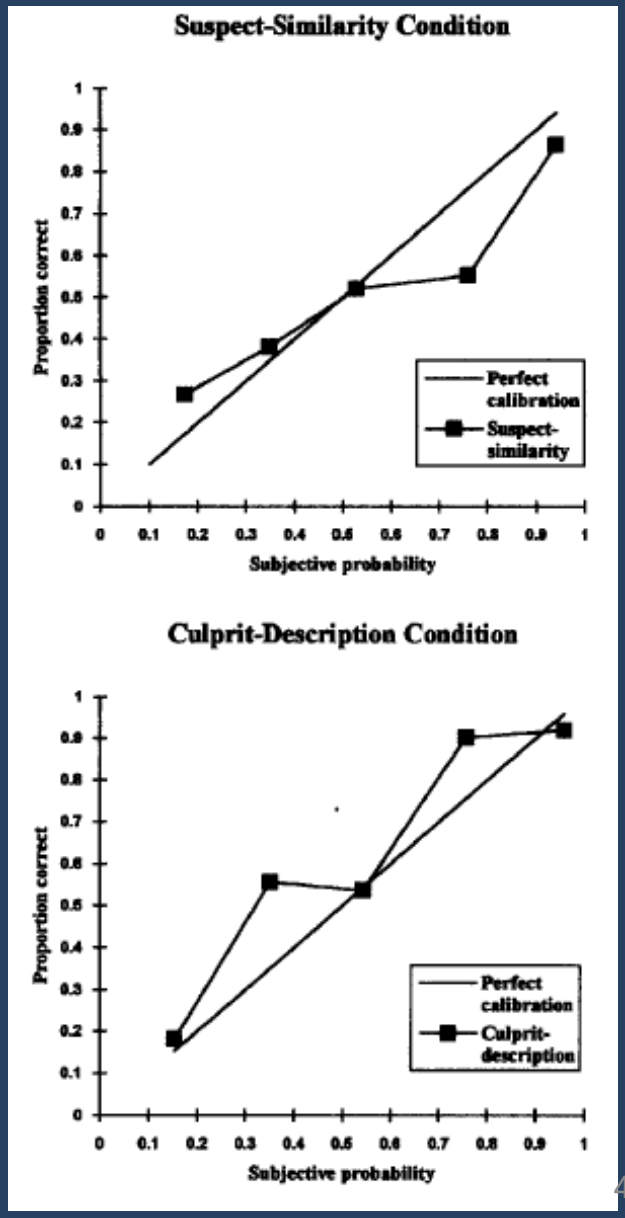
Confidence and Accuracy

Brewer et al (2013); Juslin et al. (1996); Wixted et al. (submitted)

- What is the source of this controversy?
  - properties of the point-biserial correlation
  - Interpretation of the point-biserial correlation

from Juslin, Olson & Winman (1996)





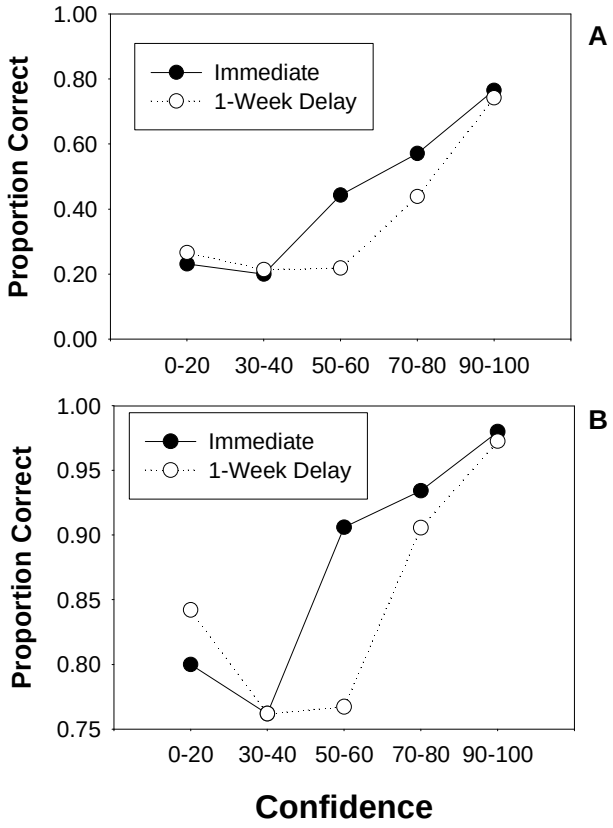


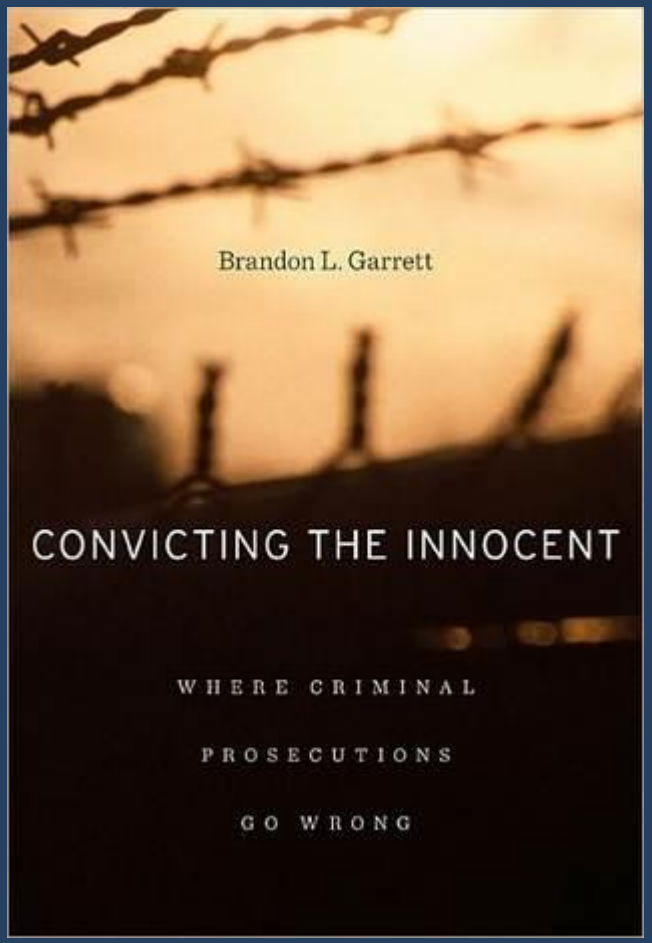
Figure 5. Observed relationship between proportion correct and confidence (**A**) and observed relationship between proportion correct and confidence for suspect identifications only (B) for the two retention-interval conditions from Experiment 1 of Palmer et al. (2013). Note the restricted range for the proportion correct scale in panel B.

From Wixted et al., Adapted from Palmer et al. 2013

Interpretation of r

Study	r
Salk vaccine trials (1954)	.01
Aspirin (heart attacks)	.03
AZT (AIDS)	.23
Myelin and Multiple Sclerosis	.31
Interpersonal Expectancy Effects	.33
Confidence and Accuracy	.41

Table adapted from Rosenthal & Rosnow (2008)



57 percent: initial identification was made with low confidence.

Some caveats about confidence

- Feedback can inflate confidence (Wells & Bradfield, 1998).
- Wells & Quinlivan (2009) have raised the question as to whether the same suggestive procedures that increase the risk of false identifications also increase witness confidence in those false identifications.
  - Biased lineup composition (Ross et al., 2007)
  - Lineup administrator influence (Greathouse & Kovera, 2009; Clark, et al. 2013).

Conclusions

- Trade-off between false identifications avoided and correct identifications lost.
- Trade-off becomes more costly if GBR is high.
- No cost view is not supported by data.
- Low cost view is not supported in some cases.
- Policy decisions are not specified by the data.
- Other considerations – due process, fairness, social justice, procedural justice, and normative social values.
- The foundations of policy must be clear – To what extent is policy driven by social science versus social values?

end.

Other considerations

- Due process, fairness, and social justice.
- Policy recommendations may be based on social science and social values

Data and Due Process

- To what extent are recommendations based on (or supported by) data?
- To what extent are they based on other considerations, namely due process, beliefs about social justice, procedural justice, etc?
- We must not confuse the two.

Extras

Basis of the No Cost View

Early 1980's Data

- One early study (Malpass & Devine, 1984) did show the No Cost pattern.
- Two others (Lindsay & Wells, 1980, 1985) were interpreted as showing the No Cost pattern (although they did not).

Theory: Relative versus Absolute Judgments

(Wells, 1984)

- Absolute judgments
  - "Witnesses identify the person in the lineup whose match to memory exceeds some cut-off or threshold."
- Relative judgments
  - "[W]itness seems to be choosing the lineup member who most resembels the witness's memory relative to the other lineup members."
  - A "useful and unflawed strategy" if the suspect is guilty.
  - "fallacious", "dysfunctional", and "dangerous" if the suspect is innocent.

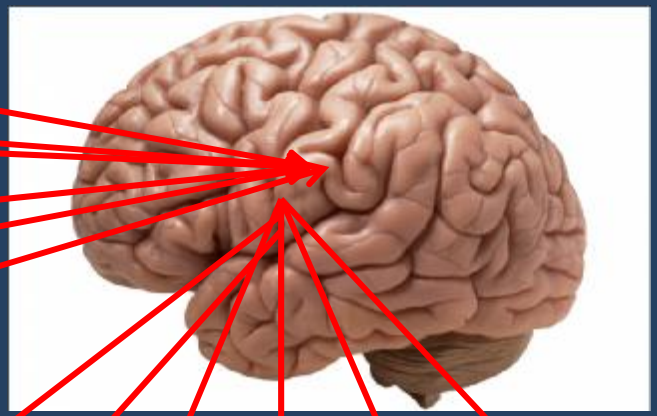
Theory: Relative versus Absolute Judgments

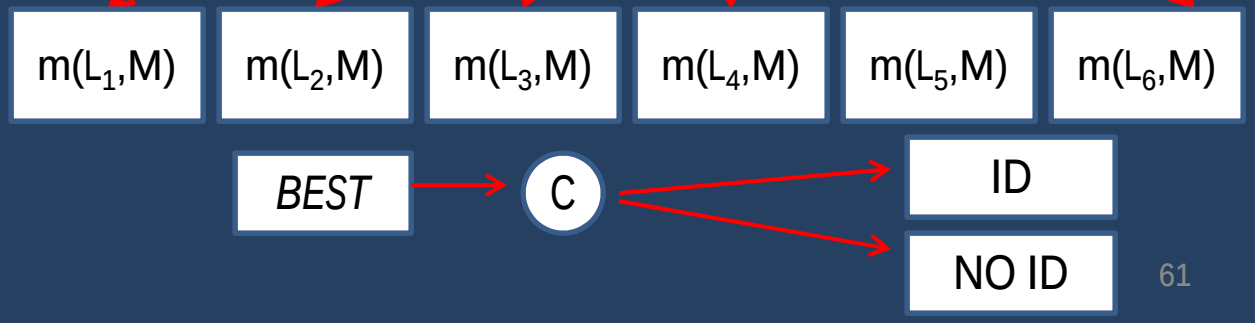
(Wells, 1984)

- Reforms are often viewed as shifting witnesses from relative judgment strategies to absolute judgment strategies.
  - Which, according to theory, should reduce false identification rates, but have no effect on correct identification rates.
  - The Absolute-relative judgment framework specifies the pattern of results that *should* be obtained.

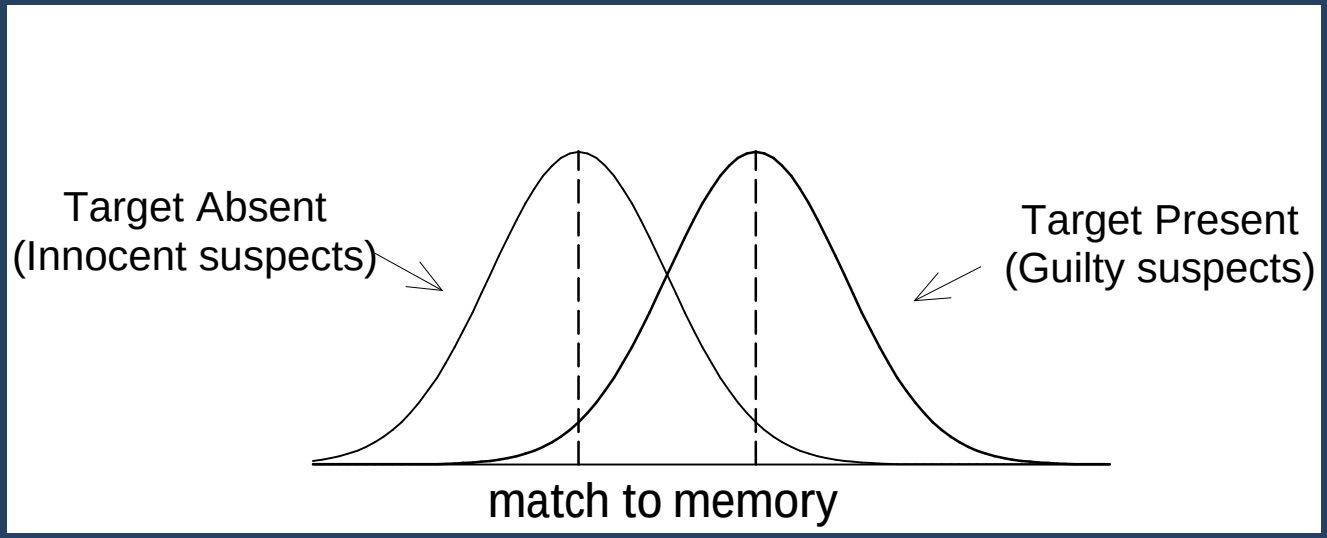
Matching Model of Eyewitness ID

Clark (2002); Clark, Erickson, & Breneman (2011); Wixted & Mickes (in press)

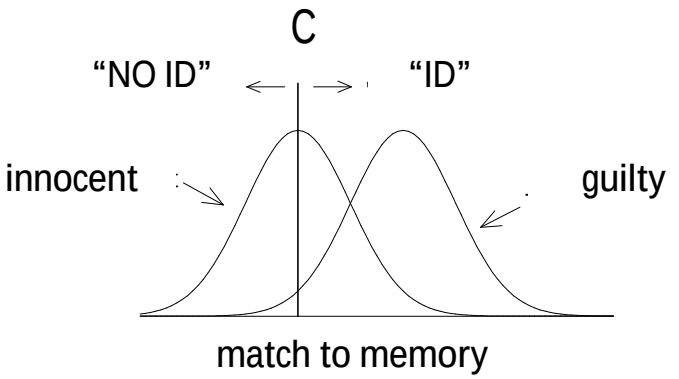


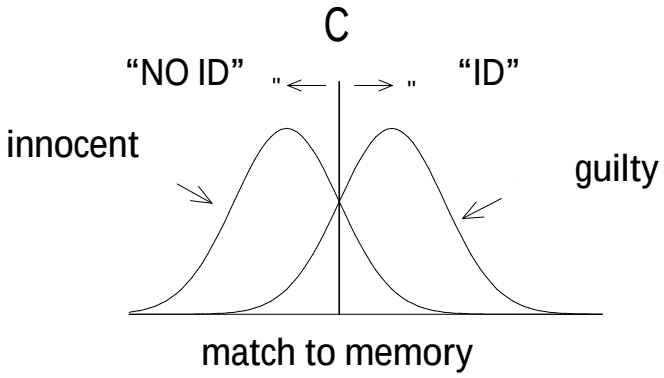


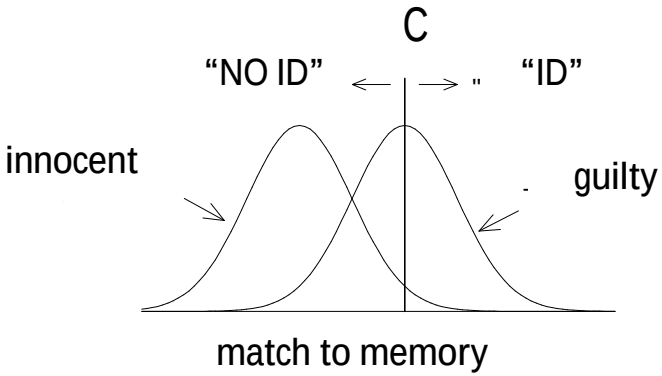
Signal Detection Theories of Memory



adapted from Wixted & Mickes (in press)



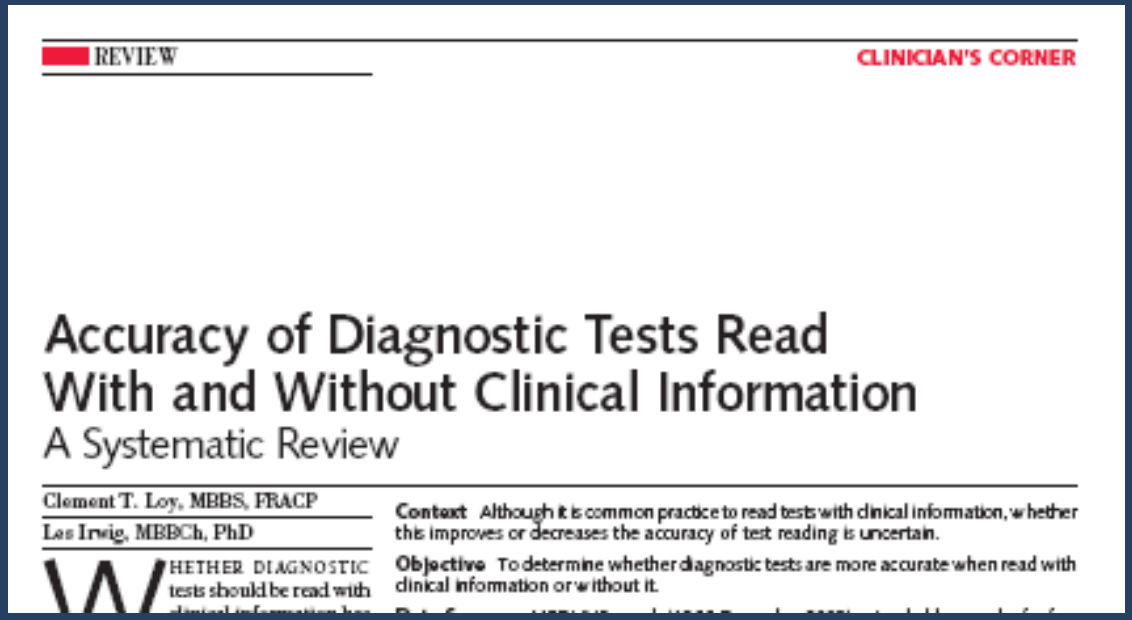




adapted from Wixted and Mickes (in press)

Another trade-off?

- Background information can increase diagnostic accuracy (Loy & Irwig, 2004).



REVIEW

CLINICIAN'S CORNER

Accuracy of Diagnostic Tests Read With and Without Clinical Information

A Systematic Review

Clement T. Loy, MBBS, FRACP
Les Irwig, MBBCh, PhD

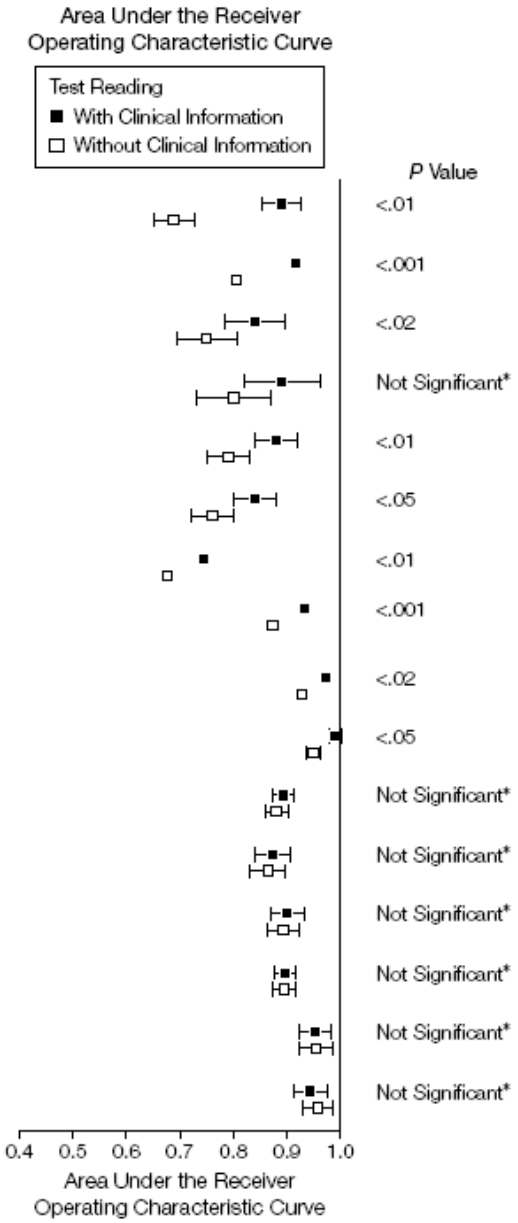
Context Although it is common practice to read tests with clinical information, whether this improves or decreases the accuracy of test reading is uncertain.

Objective To determine whether diagnostic tests are more accurate when read with clinical information or without it.

WHETHER DIAGNOSTIC tests should be read with

Figure 1. Areas Under the Receiver Operating Characteristic Curve for Diagnostic Tests Read With or Without Clinical Information

Study	Diagnostic Test	Subset of Disease or Readers	Clinical Information	*P* Value
Berbaum et al,[29] 1988	Chest Radiographs	Nonnodular Lung Disease	Constructed	<.01
Berbaum et al,[31] 1989	Bone Radiographs	Orthopedic Surgeons	Constructed	<.001
Song et al,[38] 1992	Chest, Abdomen, and Bone Radiographs	Not Applicable	Actual	<.02
Berbaum et al,[29] 1988	Chest Radiographs	Nodular Lung Disease	Constructed	Not Significant*
Berbaum et al,[28] 1986	Chest Radiographs	Not Applicable	Constructed	<.01
Raab et al,[37] 2000	Bronchial Brush Cytology	Not Applicable	Actual	<.05
Berbaum et al,[32] 1994	Pediatric Chest and Abdominal Radiographs	Not Applicable	Constructed	<.01
Berbaum et al,[30] 1988 Berbaum et al,[31] 1989	Bone Radiographs	Radiologists	Constructed	<.001
Ehara and Katsuragawa et al,[35] 1999	Wrist Radiographs in Children	Not Applicable	Constructed	<.02
McNeil et al,[36] 1983	Head Computed Tomography	Not Applicable	Actual	<.05
Cooperstein et al,[33] 1990	Digitized Chest Images	Interstitial Lung Disease	Actual	Not Significant*
Cooperstein et al,[33] 1990	Digitized Chest Images	Nodular Lung Disease	Actual	Not Significant*
Good et al,[34] 1990	Chest Radiographs	Nodular Lung Disease	Actual	Not Significant*
Good et al,[34] 1990	Chest Radiographs	Interstitial Lung Disease	Actual	Not Significant*
Good et al,[34] 1990	Chest Radiographs	Pneumothorax	Actual	Not Significant*
Cooperstein et al,[33] 1990	Digitized Chest Images	Pneumothorax	Actual	Not Significant*



- Legal scholars
  - Findlay (2004)
  - Garrett (2008)
- Policy-makers
  - Wisconsin Attorney General (2006)
- Textbooks
  - Myers (2002)
- Popular media
  - Gawande (2001, New Yorker)
  - Fenster (2012, New Haven Register)
  - Hart (2012, Houston Chronicle)

- Legal scholars
  - Findlay (2004)
  - Garrett (2008)
- Policy-makers
  - Wisconsin Attorney General (2006)
- Textbooks
  - Myers (2002)
- Popular media
  - Gawande (2001, New Yorker)
  - Fenster (2012, New Haven Register)
  - Hart (2012, Houston Chronicle)

THE BEHAVIORAL AND BRAIN SCIENCES (1978), 3, 377–415
Printed in the United States of America

Interpersonal expectancy effects: the first 345 studies

Robert Rosenthal
Department of Psychology and Social Relations, Harvard University, Cambridge, Mass. 02138

Donald B. Rubin
Department of Statistics, Harvard University, Cambridge, Mass. 02138
(On leave from Educational Testing Service, Princeton, New Jersey)