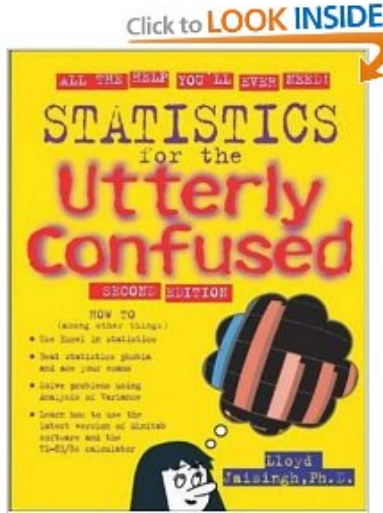
Lay Reactions to Quantitative Statements about the Weight of Forensic Science Evidence

William C. Thompson
University of California, Irvine

Research Issues

- Do jurors understand statistical testimony?
- Which statistics are “best”?
- What weight do jurors give to qualitative statements about the weight of evidence?



Normative Criteria for Evaluation

- Sensitivity to the strength of forensic evidence
- Susceptibility to fallacious interpretations
  - Prosecutor's fallacy/Source probability error
  - Defense attorney's fallacy
- Logical coherence of judgments

Research

Methods

- Participants—Actual jurors or mTurkers
- Evaluate hypothetical cases
- Judgments before and after receiving forensic evidence
- Experimentally varied:
  - Strength of Forensic Evidence
  - Strength of non-Forensic Evidence
  - Presentation Format
  - Type of forensic Evdence (e.g., DNA vs. shoeprint)
  - Dependent measures

Reports

- Thompson, Kaasa & Peterson, J. Empirical Legal Studies (2013)
- Thompson & Newman, Law & Human Behavior (2015)

Sensitivity to Strength of Evidence

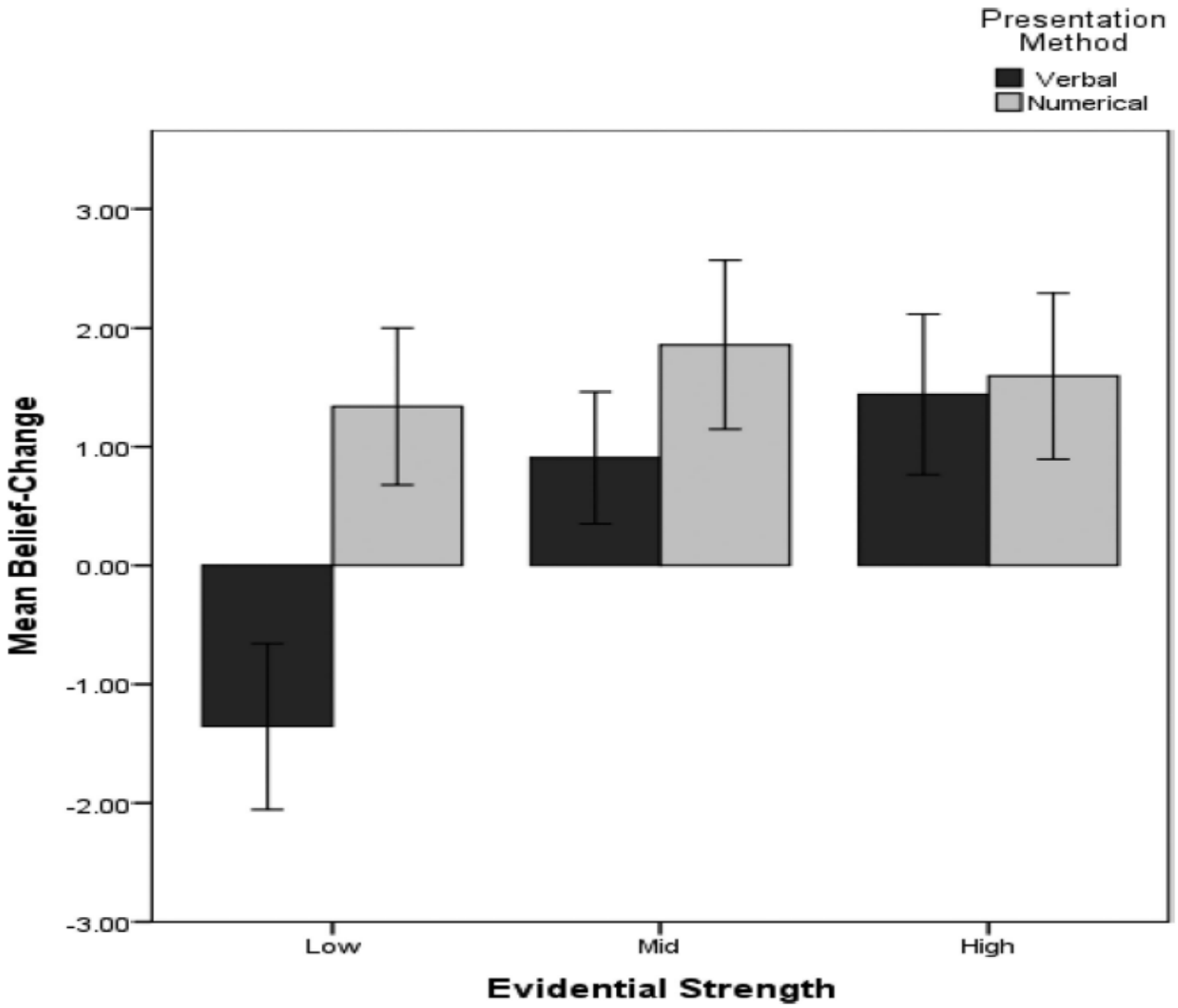
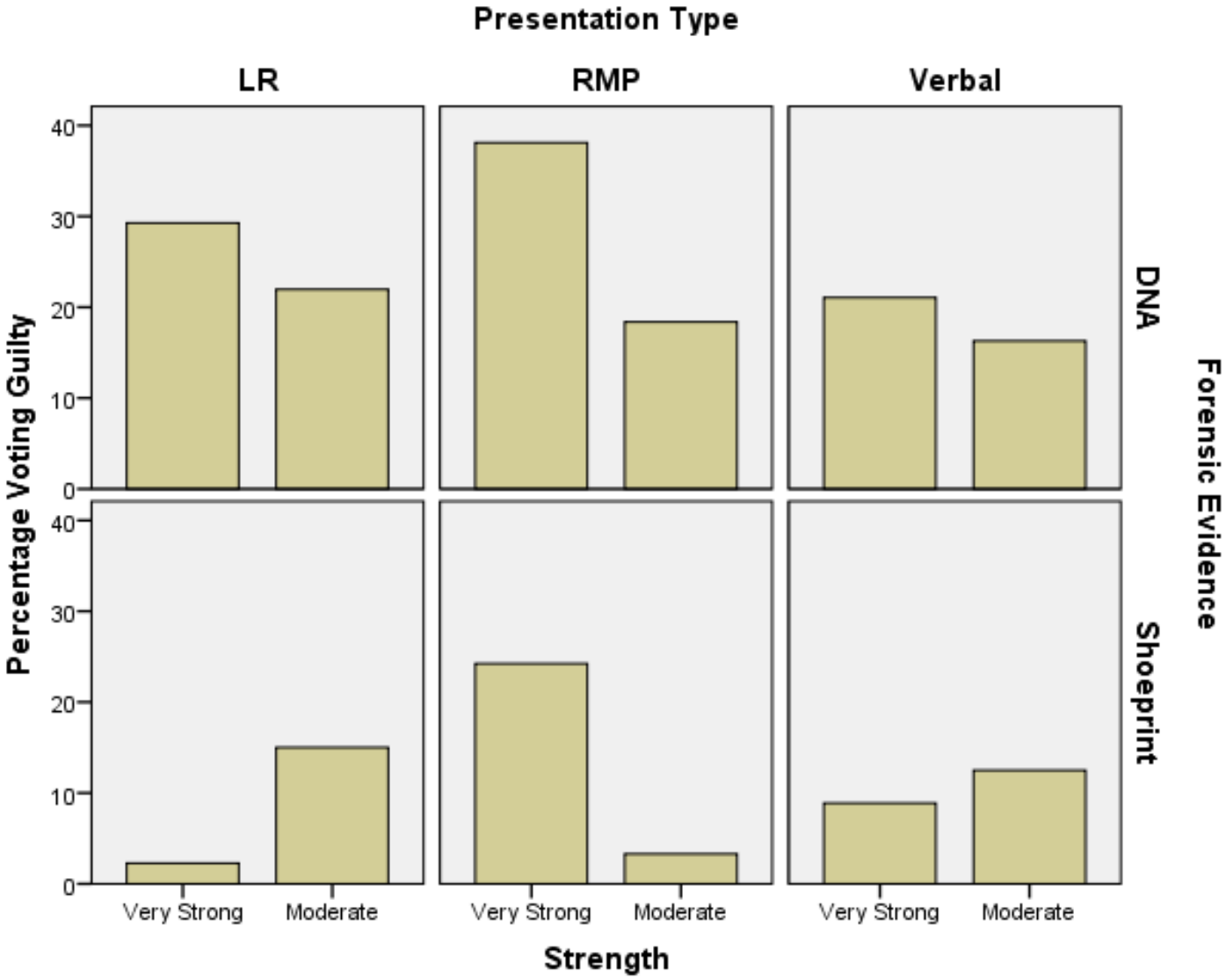


Figure 1. Mean adjusted belief change by presentation method and evidential strength (error bars ± 2 standard errors).

Martire et al. LHB (2013)



Thompson & Newman, *Law & Human Behavior*, 2015

Table 1.

Percentage of subjects who endorsed the source probability error, defense attorney's fallacy, both errors or neither error and conviction rates, log change scores and implicit LRs within each group.

Error Endorsed	Percentage Endorsing Fallacy	Conviction Rate	Log Scale Change Score	Implicit LR
Source Probability Only	35.49% (192)	32.29% (62)	1.93 (3.19)	12.1 (22.18)
Defense Fallacy Only	17.93% (97)	3.09% (3)	1.14 (1.84)	3.09 (9.79)
Both Errors	28.10% (152)	5.26% (8)	1.26 (2.21)	1.4 (.58)
Neither Error	12.20% (66)	15.15% (10)	1.46 (3.25)	4.12 (10.91)

Will people understand likelihood ratios if they are explained by a really good expert?

Crime + Initial Guilt

Forensic Evidence

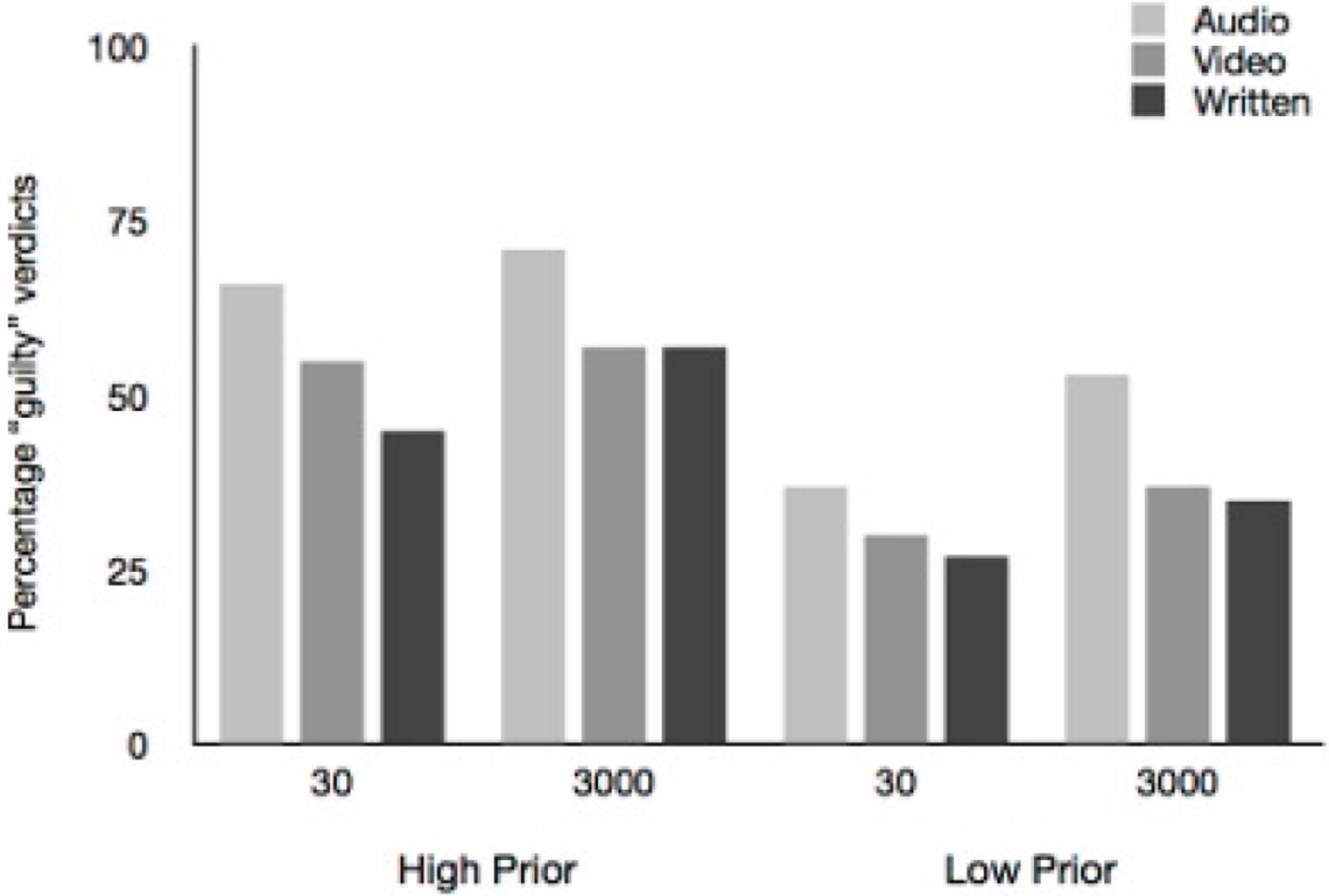
Subsequent Guilt

Crime

High Prior

Low Prior

he police asked an expert on forensic voice comparison to help them determine whether Joseph Anton was the man who called the credit card activation center. The expert they chose was Dr. Geoffrey Stewart Morrison. He is the director of a forensic voice comparison laboratory at the University of New South Wales in Sydney, Australia. His laboratory does research and casework in forensic speech science, which includes forensic voice comparison and disputed utterance analysis. He has submitted reports on forensic voice comparisons to both prosecutors and defense lawyers and had appeared in court as an expert witness on four previous occasions. About half of the voice comparison work he has done in criminal cases was at the behest of the prosecutors and about half at the behest of defense lawyers.

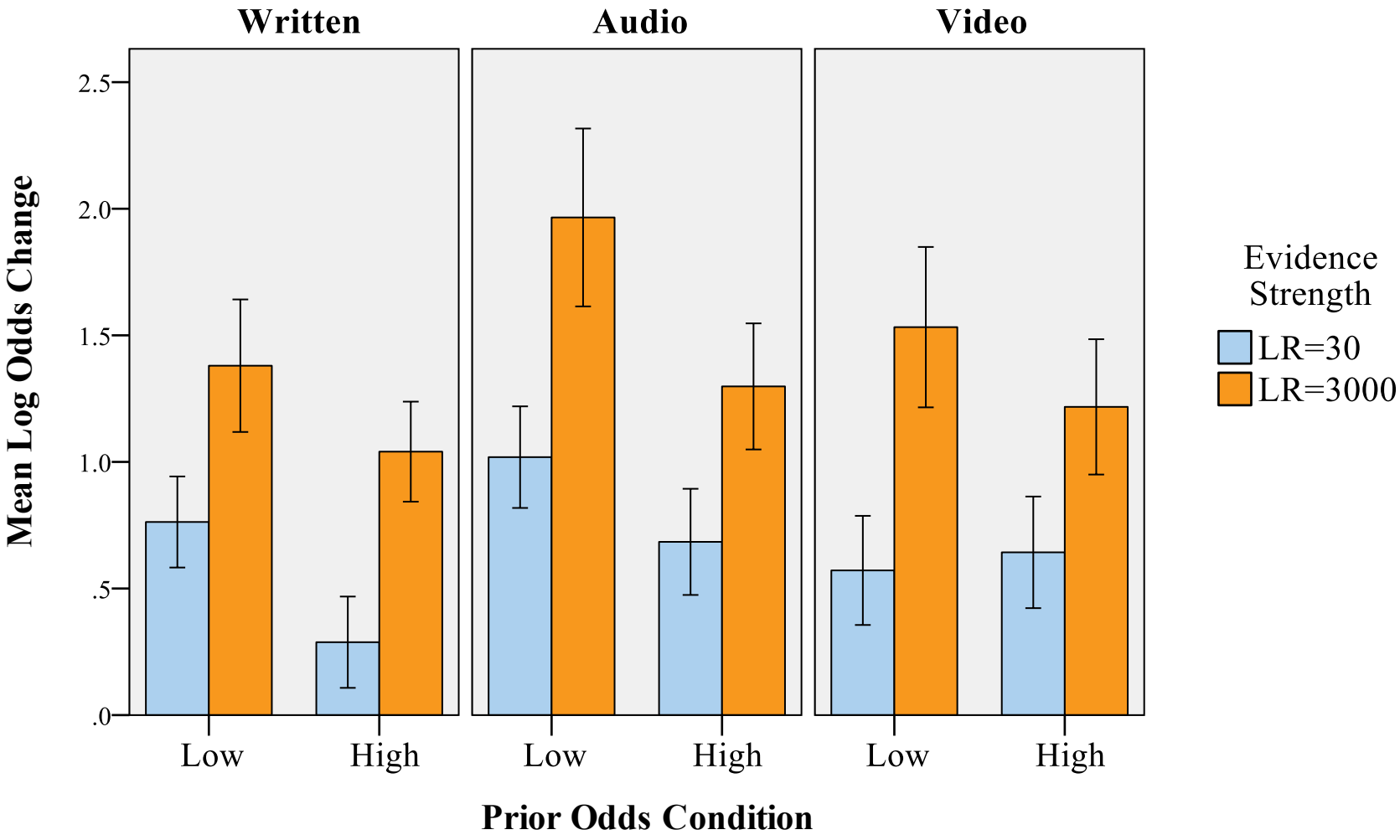

Audio
Video
Written
100
75
50
25
0
Percentage "guilty" verdicts
30
3000
30
3000
High Prior
Low Prior

Perceived Odds of Guilt

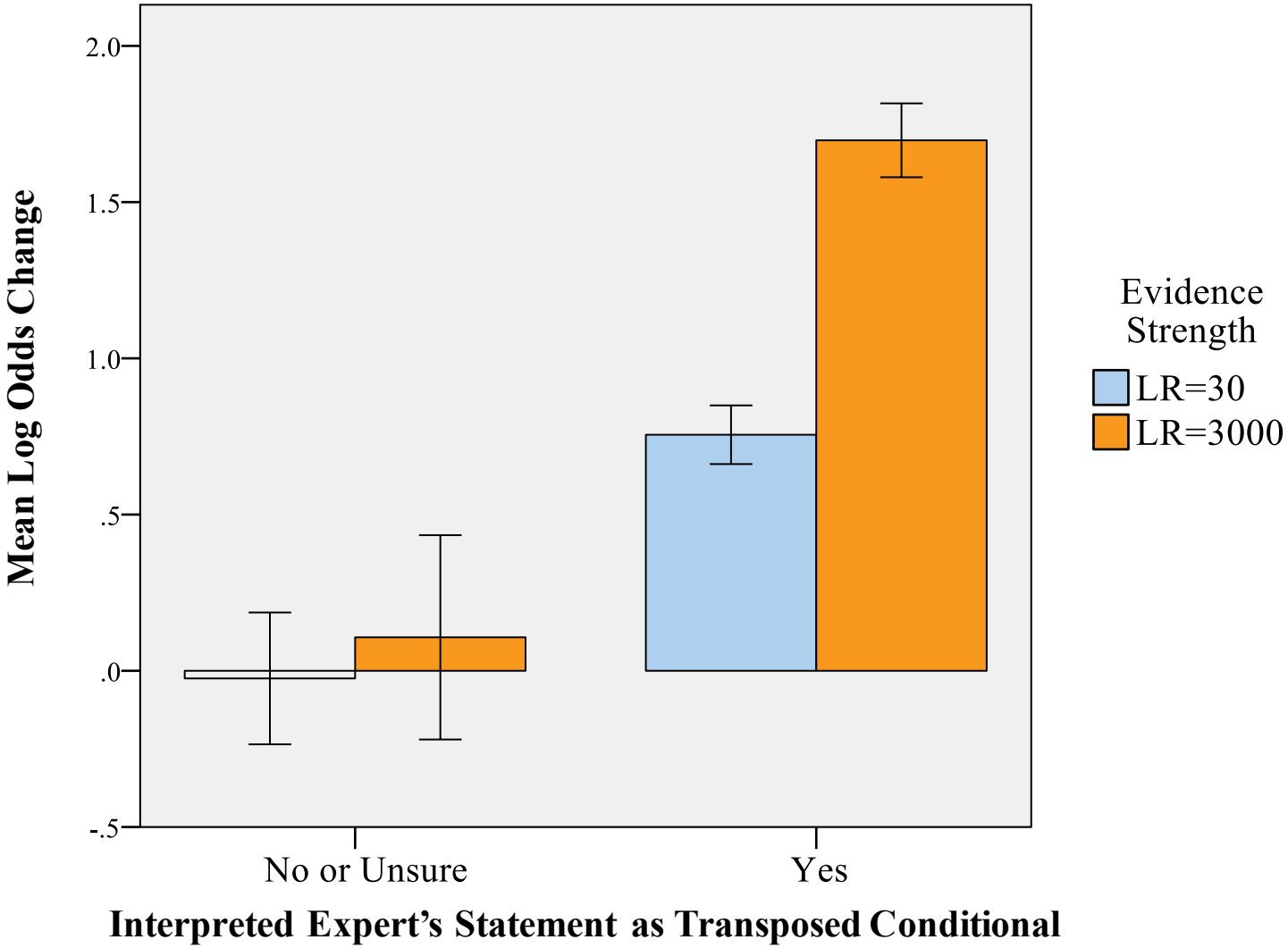
You said it was more likely that the caller was Joseph Anton than someone else. How much more likely?

- Between 1 and 10 times more likely (51%-91% chance it was Anton)
- Between 10 and 99 times more likely (91%-99% chance it was Anton)
- Between 100 and 999 times more likely (99%-99.90% chance it was Anton)
- Between 1000 and 9999 times more likely (99.90%-99.99% chance it was Anton)
- Between 10,000 and 99,999 times more likely (99.99%-99.999% chance it was Anton)
- More than 100,000 times more likely (More than 99.9999% chance it was Anton)

Odds Change by all conditions

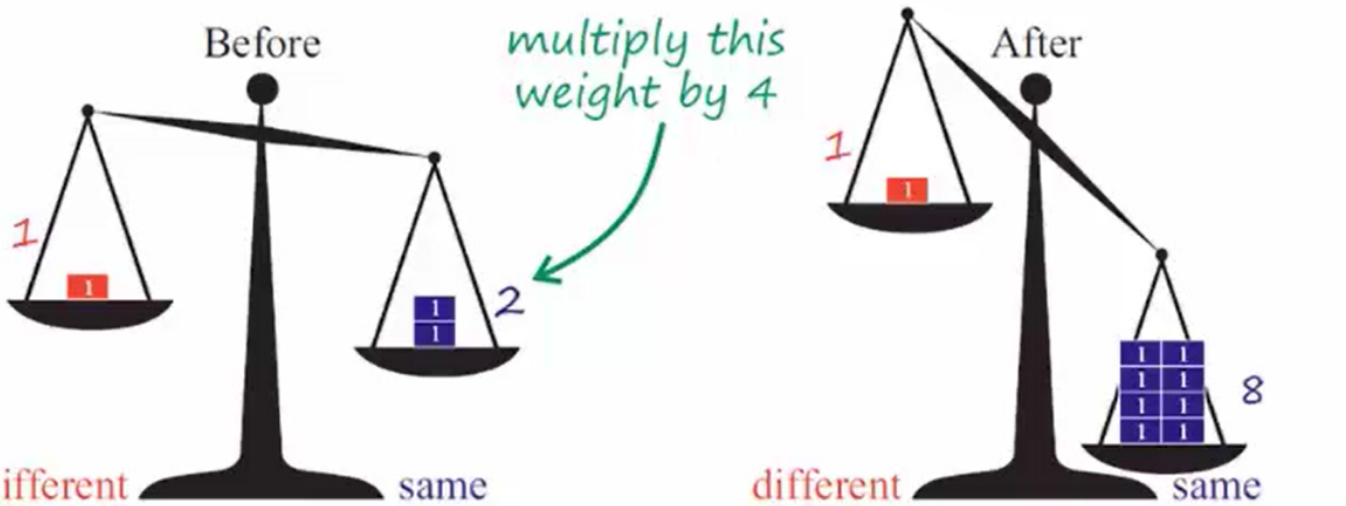


Odds Change by Transposed Conditional



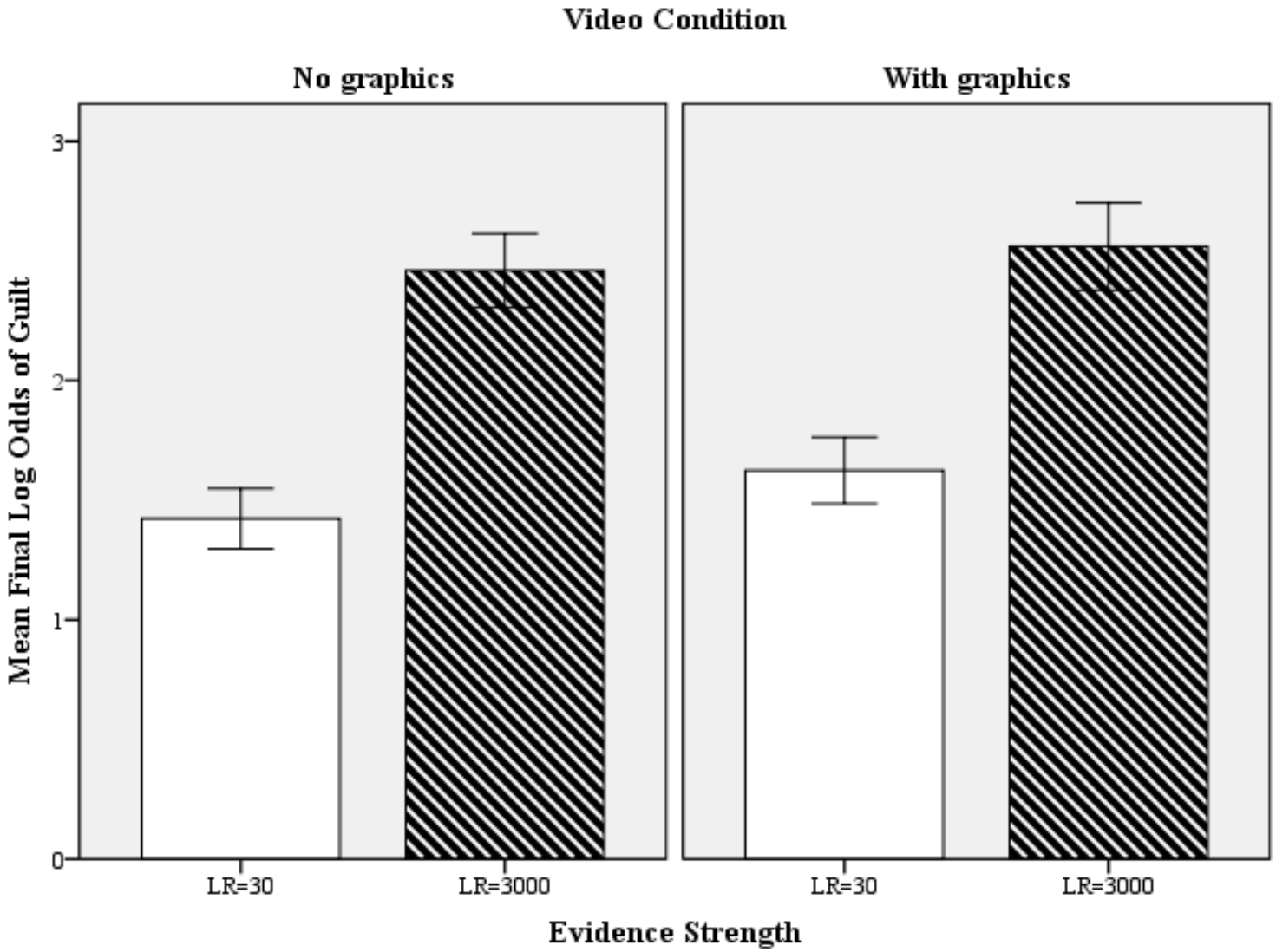
Will graphics help?

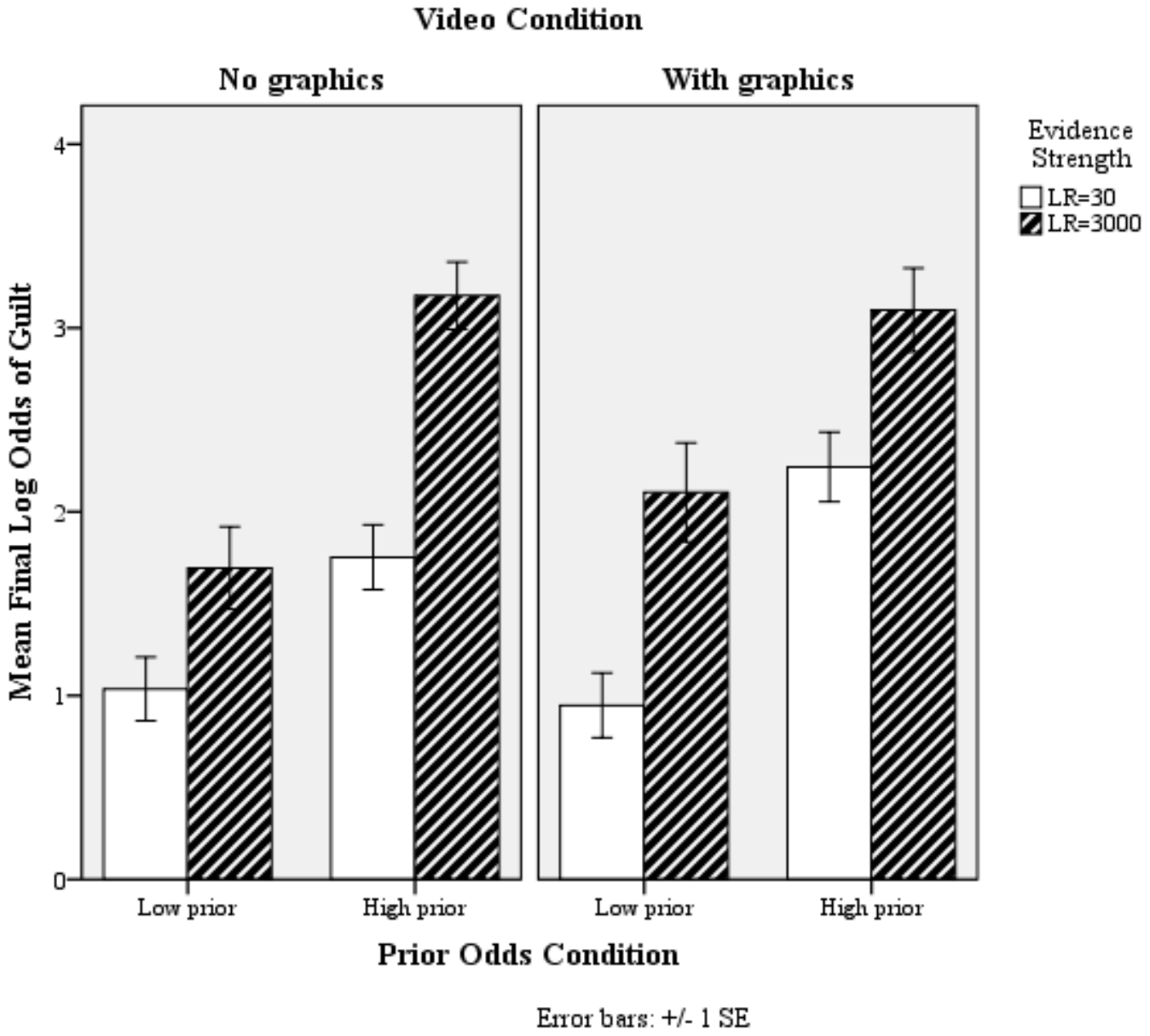
Example: The evidence is 4 time more likely given the same-speaker hypothesis than given the different-speaker hypothesis



if before you believed that the same-speaker hypothesis was 2 times more probable than the different-speaker hypotheses

now you believe that the same-speaker hypothesis is 8 times more probable than the different-speaker hypothesis


Video Condition
No graphics
With graphics
Mean Final Log Odds of Guilt
3
2
1
0
LR=30
LR=3000
LR=30
LR=3000
Evidence Strength
Error bars: +/- 1 SE


Video Condition
No graphics
With graphics
Evidence Strength
LR=30
LR=3000
Mean Final Log Odds of Guilt
0
1
2
3
4
Low prior
High prior
Low prior
High prior
Prior Odds Condition
Error bars: +/- 1 SE

Correct interpretation of LR?

	No Graphics	With Graphics
It is [30/3000] times more likely that the offender was Mr. Anton than some other speaker from the relevant population.	80%	80%
The acoustic properties found on the offender recording are [30/3000] times more probable if the offender was Mr. Anton than if the offender was some other speaker from the relevant population.	**72%**	**80%**
The same-speaker hypothesis is [30/3000] times more likely to be true than the different-speaker hypothesis.	**42%**	**63%**

Perceived strength of qualitative statements

"Given the size and quality of the crime scene print,

- it is [moderately probable; highly probable; practically certain] that the suspect is the person who made the crime scene print" ***(Source probability)***
- these findings provide [weak; moderate; extremely strong] support for the theory that the suspect is the person who made the crime scene print" ***(Weight of evidence)***
- I would expect about one person in [10; 1000; 100,000] to have a fingerprint similar enough to be indistinguishable from it" ***(RMP)***

Perceived Order of Strength

- Practically certain-- 82%
- RMP=1 in 100,000-- 74%
- Extremely Strong Support-- 74%
- Highly probable-- 63%
- RMP=1 in 1000-- 51%
- Moderately probable-- 36%
- Moderate Support-- 36%
- RMP=1 in 10-- 20%
- Weak support-- 13%

Future Studies

- Can we reduce fallacious reasoning and improve sensitivity to relevant variables with:
  - Graphic exhibits?
  - More extensive, realistic testimony?
  - Lawyers' arguments?
- Advantages and disadvantages of statistical characterizations, relative to traditional categorical conclusions
- Can we develop better theoretical explanations for
  - Judgments about the credibility/strength of evidence?
  - Modality effects?

Issues for Discussion

- Numbers or not?
  - Empirical data
  - Subjective estimates
- Form of Conclusion
  - Source probability (posterior probability)
  - Strength of Evidence Statement
    - LR; Verbal Equivalent to LR; other possibilities?
  - Random match/inclusion probability
  - Sensitivity/Specificity
  - Other possibilities?

Are numbers necessary?

2009 National Research Council Report on Forensic Science

- Forensic science reports, and any courtroom testimony stemming from them, must include clear characterizations of the limitations of the analyses, *including associated probabilities* where possible. (p. 186)

Numbers Necessary?

"When you can measure what you are sp
about, and express it in numbers, you know
something about it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts advanced to the stage of science."

– Lord Kelvin (William Thomson)

Disjunction

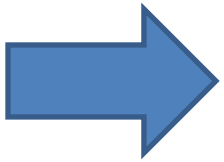
What the expert can say:

What the jury wants to know:

What's the probability it came from the defendant?

Leap of Logic

The items share unusual or rare characteristics



The items have, or probably have, a common source

How rare do the shared characteristics need to be to justify concluding that the items have a common source?

How rare do they need to be to justify the conclusion that the items probably have a common source?

Should we allow experts to opine on source probabilities?

- A scientific expert can never opine on the probability that two items have a common source without
  - making an assumption or taking a position
  - on the strength of the non-scientific evidence in the case.
- Is that something we should allow experts to do?
- If not, what should experts say?

Does testimony about source probabilities usurp the role of the jury?

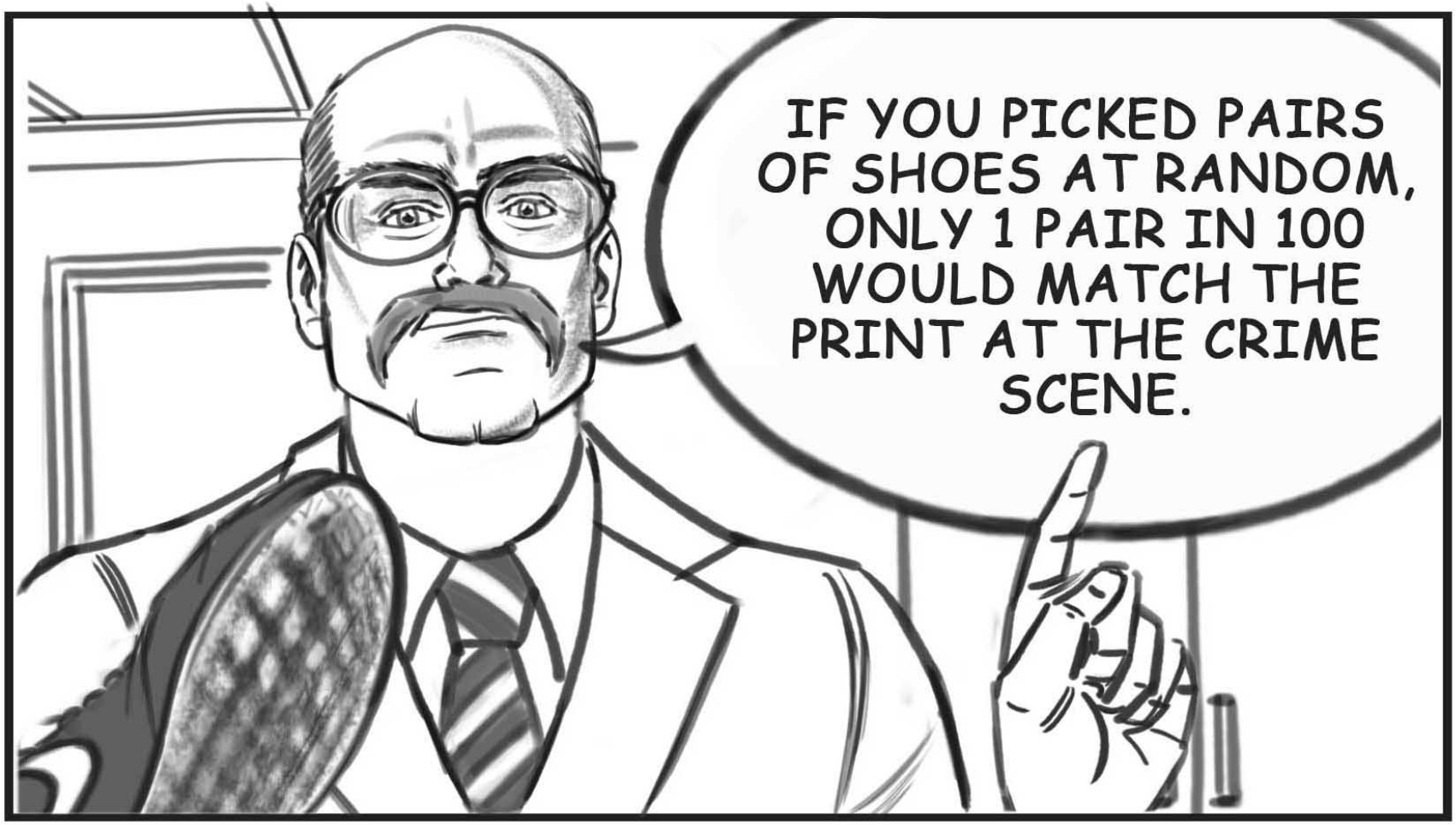
- And what should be the fate of usurpers???

A Random Match Probability

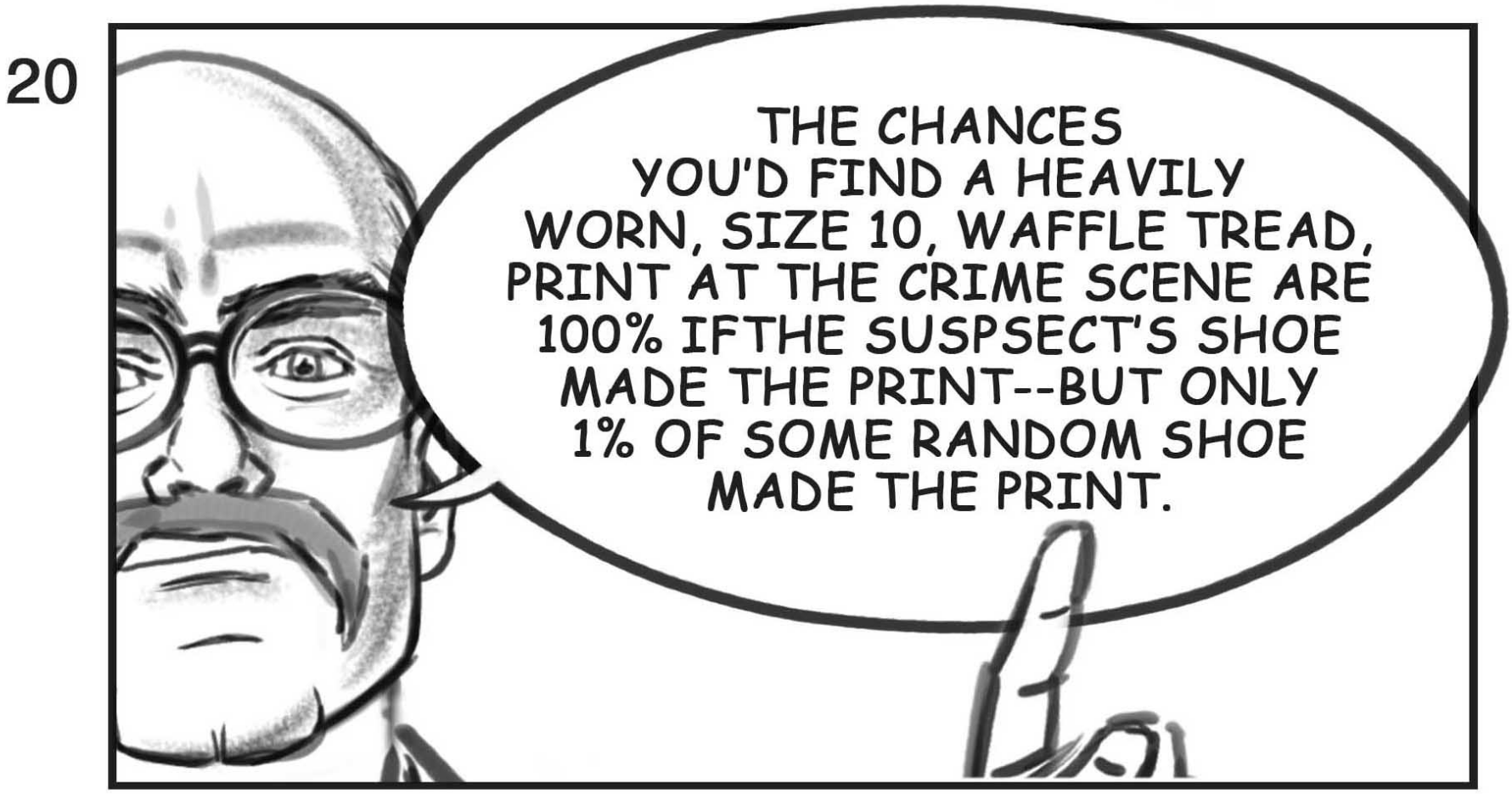
18

A Frequency Estimate

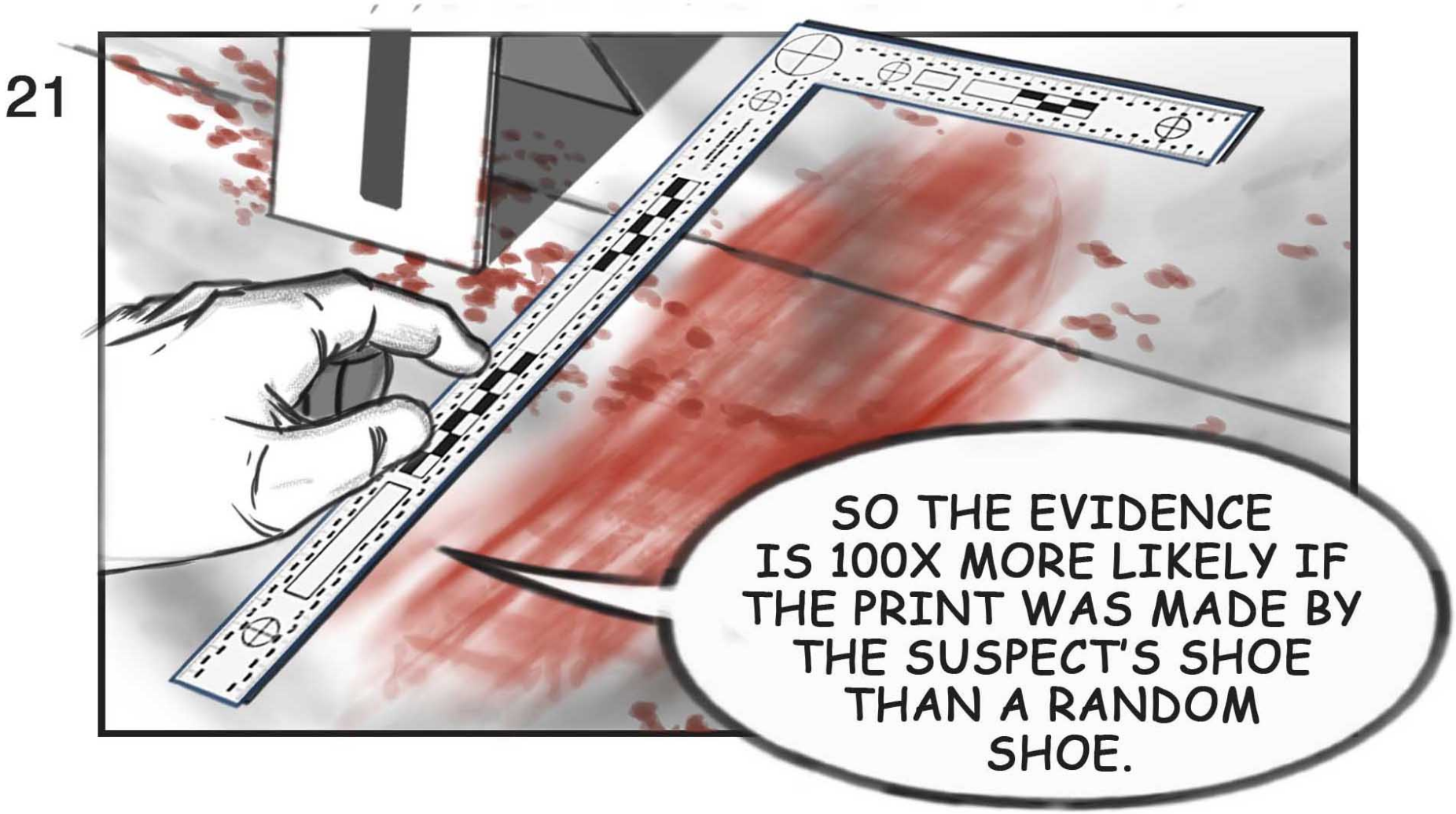
19



Diagnostic Statistics: Sensitivity and Specificity



A Likelihood Ratio



"Verbal Equivalents" to Likelihood Ratios

Standards for Numerical and Verbal Expression of Likelihood Ratios (Association of Forensic Science Providers, 2009)

Recommended likelihood ratio terminology

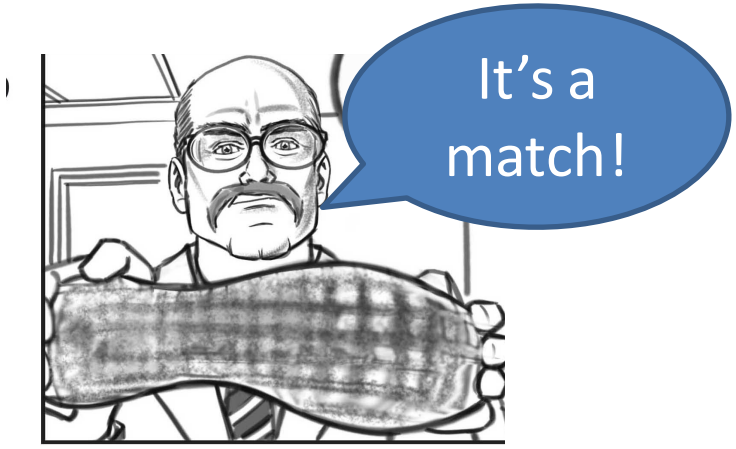
Numerical expression	Verbal expression (support)
> 1–10	Weak or limited
10–100	Moderate
100–1,000	Moderately strong
1,000–10,000	Strong
10,000–1,000,000	Very strong
> 1,000,000	Extremely strong

Issues for Discussion

- Numbers or not?
  - Empirical data
  - Subjective estimates
- Form of Conclusion
  - Source probability (posterior probability)
  - Strength of Evidence Statement
    - LR; Verbal Equivalent to LR; other possibilities?
  - Random match/inclusion probability
  - Sensitivity/Specificity
  - Other possibilities?

How should forensic scientists report their findings?

- Categorical Approach
- Quantitative Approach
  - Empirical
  - Subjective



Traditional Categorical Characterizations

- Individualization/Identification (sometimes to “a reasonable scientific certainty”)
- Match/Inclusion/Consistent with…
- Inconclusive
- Exclusion

Elaborated Categories (SWGDOC)

- Identification (definite conclusion of identity)
- Strong probability (highly probably, very probable)
- Probable
- Indications (evidence to suggest)
- No conclusion (totally inconclusive, indeterminate)
- Indications did not
- Probably did not
- Strong probability did not
- Elimination

SWGTREAD Categories (A reasonable alternative?)

- Lacks sufficient detail
- Exclusion
- Indications of non-association
- Limited association of class characteristics
- Association of class characteristics
- High degree of association
- Identification

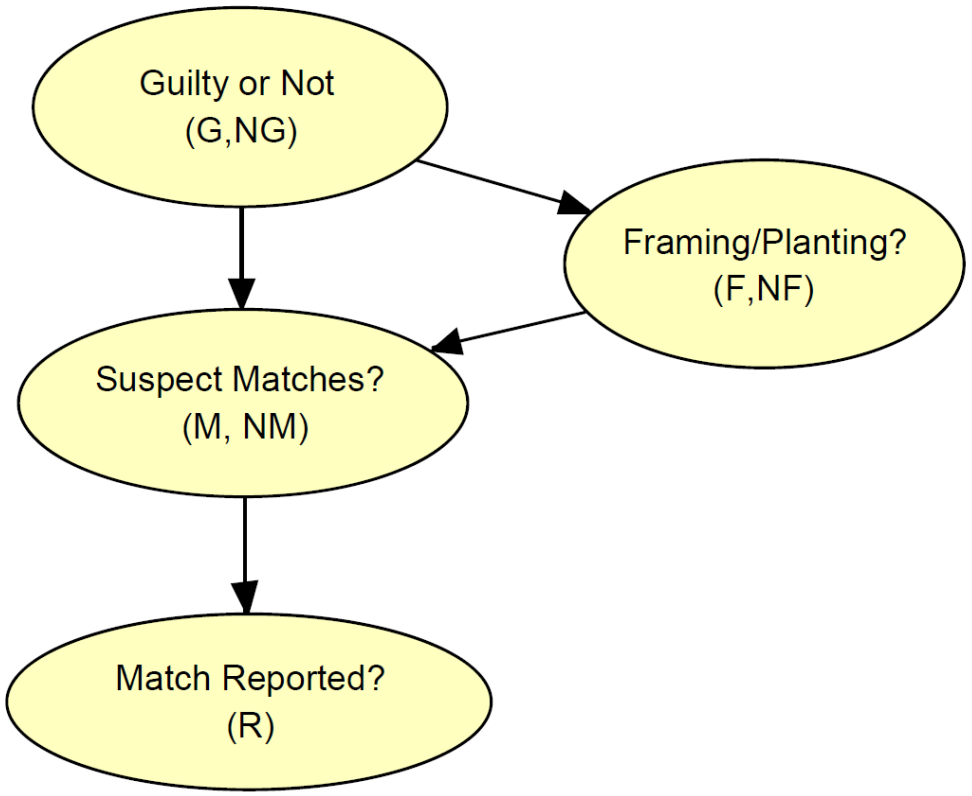
Perceived Benefits of Numbers

- Valid
- Transparent
- Precise

What kind of numbers?

- *Frequencies /Match Probabilities (RMPs)*
  - E.g., "…the shared genetic characteristics would be found in 1 person in 10 million" (in some reference population)
  - E.g., "…the probability of finding these characteristics in a random individual (from a reference population) is 1 in 10 million"
- *Likelihood Ratios*
  - E.g., "The evidence is x times more likely under the proposition that *defendant is the source* than under the proposition that *someone else is the source.*"
- Verbal Equivalents (to Likelihood Ratios)

Figure 5: Bayesian Network Model for Evaluating the Probative Value of the Forensic Evidence Based on Individual Perceptions of the RMP, FRP and FUP



	Guilty (G)	Not Guilty (NG)
Framed (F)	FUP	FUP
Not Framed (NF)	1-FUP	1-FUP

	Match (M)	No Match (NM)
Match Reported (R)	1	FRP
Not Reported	0	1-FRP

	Framed (F)		Not Framed (NF)	
	G	NG	G	NG
Match (M)	1	1	1	RMP
No Match (NM)	0	0	0	1-RMP

- Will jurors understand forensic statistics?