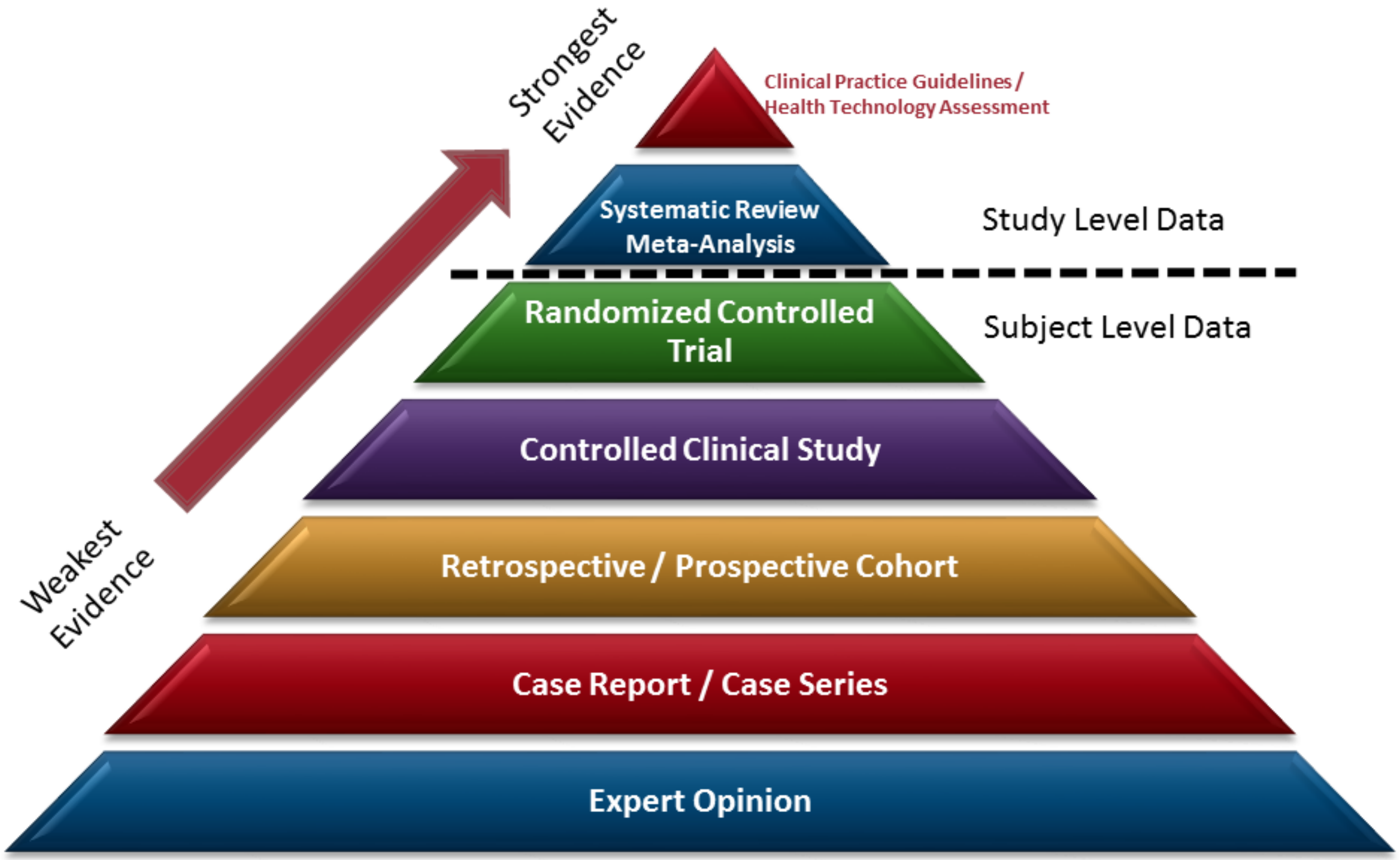
Evidence-based decision making
The power of algorithms

Rui Mata, FS 2023

Version: March 6th, 2023



https://en.wikipedia.org/wiki/Levels_of_evidence

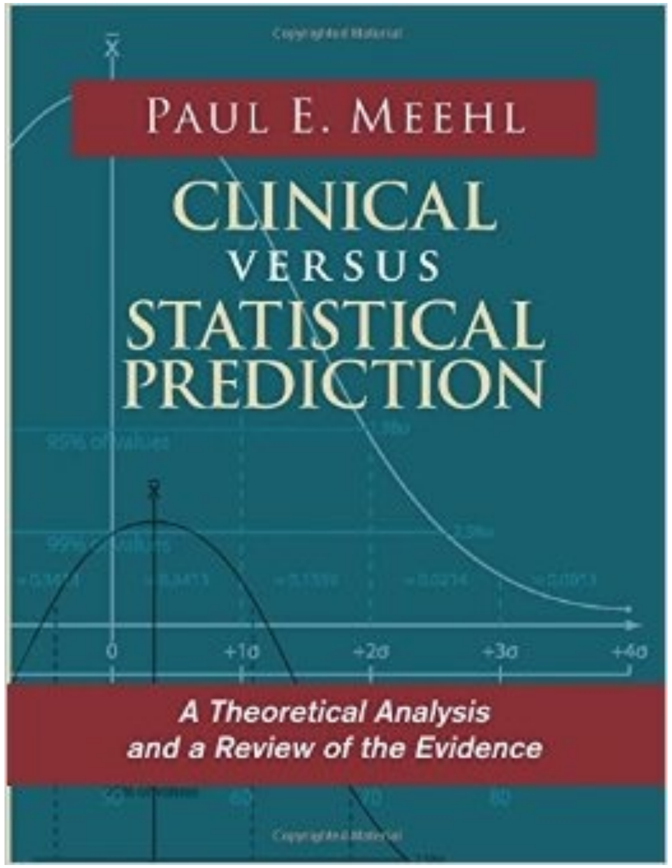
Learning Objectives for Today

- distinguish clinical and actuarial judgment
- list potential advantages and limitations of clinical and actuarial judgment
- understand the lens model and learn associated terminology
- be aware of the sizeable increase in predictive accuracy of actuarial relative to clinical judgment

Clinical vs. Statistical Prediction: "A disturbing little book"

Paul Meehl (1920-2003)

- clinician
- psychodynamic orientation
- familiar with projective tests



1954

Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Echo Point Books & Media.

Clinical vs. Statistical Prediction: ‘‘A disturbing little book’’

Meehl looked at ca. 20 studies (depending on inclusion criteria) In all but one case, predictions made by actuarial means were equal to or better than clinical methods

> “…it is clear that the dogmatic, complacent assertion sometimes heard from clinicians that ‘naturally’ clinical prediction, being based on ‘real understanding’ is superior, is simply not justified by the facts to date”.

Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Echo Point Books & Media.

Clinical vs. Statistical Prediction

Clinical Versus Actuarial Judgment

ROBYN M. DAWES, DAVID FAUST, PAUL E. MEEHL

Professionals are frequently consulted to diagnose and predict human behavior; optimal treatment and planning often hinge on the consultant's judgmental accuracy. The consultant may rely on one of two contrasting approaches to decision-making—the clinical and actuarial methods. Research comparing these two approaches shows the actuarial method to be superior. Factors underlying the greater accuracy of actuarial methods, sources of resistance to the scientific findings, and the benefits of increased reliance on actuarial approaches are discussed.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668–1674.

Potential Advantages of Clinical & Statistical Prediction

Clinical	Actuarial
Ability to use theory to form judgments	Immunity from fatigue and other limitations (forgetfulness, over-confidence)
Ability to use rare events	Consistency and proper weighting (variables are weighted the same way every time, according to their importance)
Ability to detect complex predictive cues	Feedback & base-rates 'built-in' to the system (clinicians rarely get immediate feedback and have imperfect memory)
Ability to re-weight cues as a function of changing circumstances	Not overly sensitive to optimal weightings (simple linear weightings often do well)

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. ***Science, 243***(4899), 1668–1674.

Clinical Prediction: Experience helps (but not much)

Effects of experience on judgment accuracy in clinical judgment ($d \approx .15$)

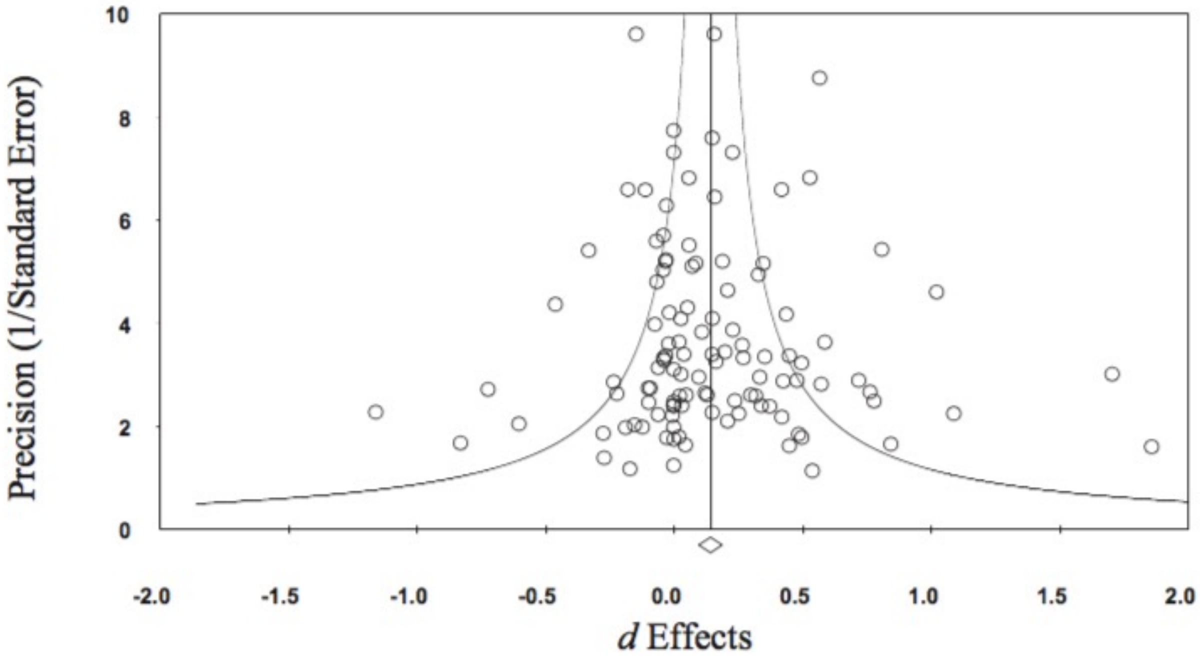


Figure 2. Funnel plot of precision by *d* effects with removal of Garcia (1993), d = 3.08.

discussion: heterogeneity across studies in defining “experience” is problematic; overall, effect of experience is small; crucially, moderator analysis that consider extreme groups (e.g., experts/novices) don’t find significant differences!

Spengler, P. M., & Pilipis, L. A. (2015). A comprehensive meta-reanalysis of the robustness of the experience-accuracy effect in clinical judgment. ***Journal of Counseling Psychology, 62***(3), 360–378. http://doi.org/10.1037/cou0000065

Clinical Prediction: Experts are not perfectly calibrated

Correlation between confidence and judgment accuracy in clinical judgment ($r \approx .15$)

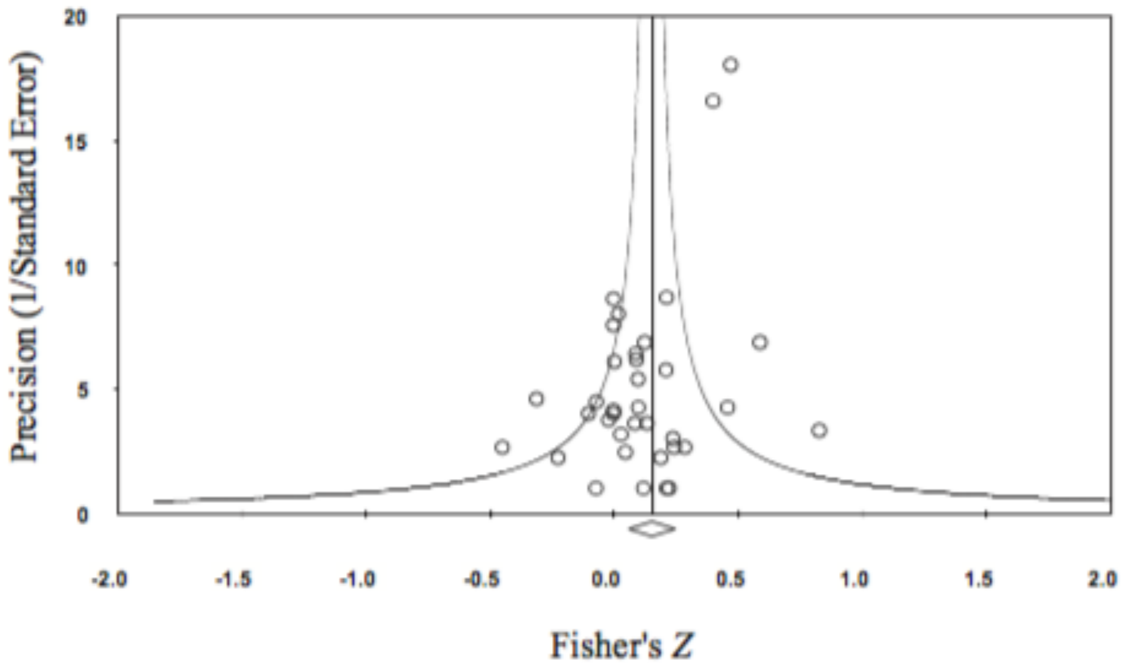


Figure 3. Funnel plot of precision by Fisher's Z.

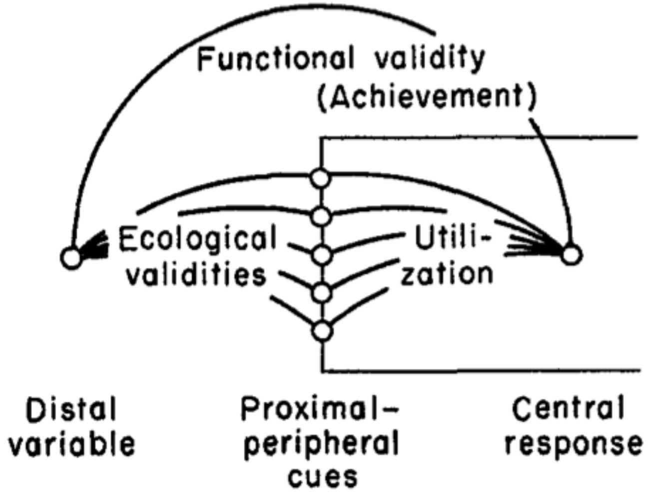
"(…) an r of .15 reflects that confidence accounts for 2% of variance in judgment accuracy (r^2 = .0225), which by any standard seems inconsequential. If counseling and other psychologists do in reality have the ability to appropriately gauge the accuracy of their own judgments, one would expect the aggregated effect size to be much larger"

Miller, D. J., Spengler, E. S., & Spengler, P. M. (2015). A meta-analysis of confidence and judgment accuracy in clinical decision making. ***Journal of Counseling Psychology, 62***(4), 553–567. http://doi.org/10.1037/cou0000105

The Lens Model and Policy Capturing

Egon Brunswick
(1903-1955)

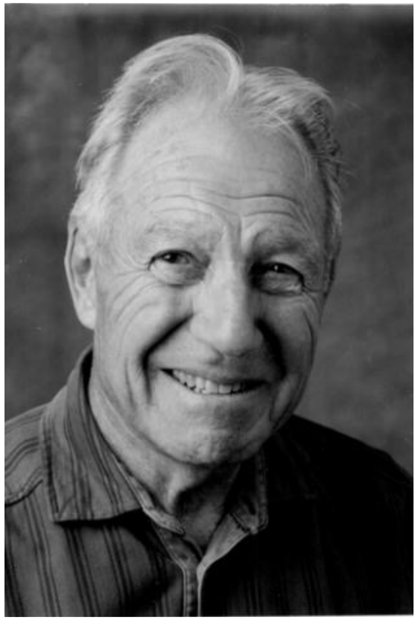
Brunswick proposed the **lens model** which describes the link between characteristics of the world and individual's perception of these characteristics.



Egon Brunswik (born in Budapest, studied in Vienna, later emigrated to USA) argued that psychology should give as much attention to the properties of the organism's environment as it does to the organism itself. He asserted that the environment with which the organism comes into contact is an uncertain, probabilistic one, however lawful it may be in terms of physical principles. Adaptation to a probabilistic world requires that the organism learn to employ probabilistic uncertain evidence (proximal cues) about the world (the distal object). His work has influenced psychology of perception (cf. Roger Shepard) and judgment and decision making (cf. Ken Hammond). His focus on the environment also led him to argue for the need to use representative designs in psychology (i.e., naturalistic sampling of stimuli)

Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. http://doi.org/10.1037/h0046845

The Lens Model and Policy Capturing



Kenneth R. Hammond
(1917-2015)

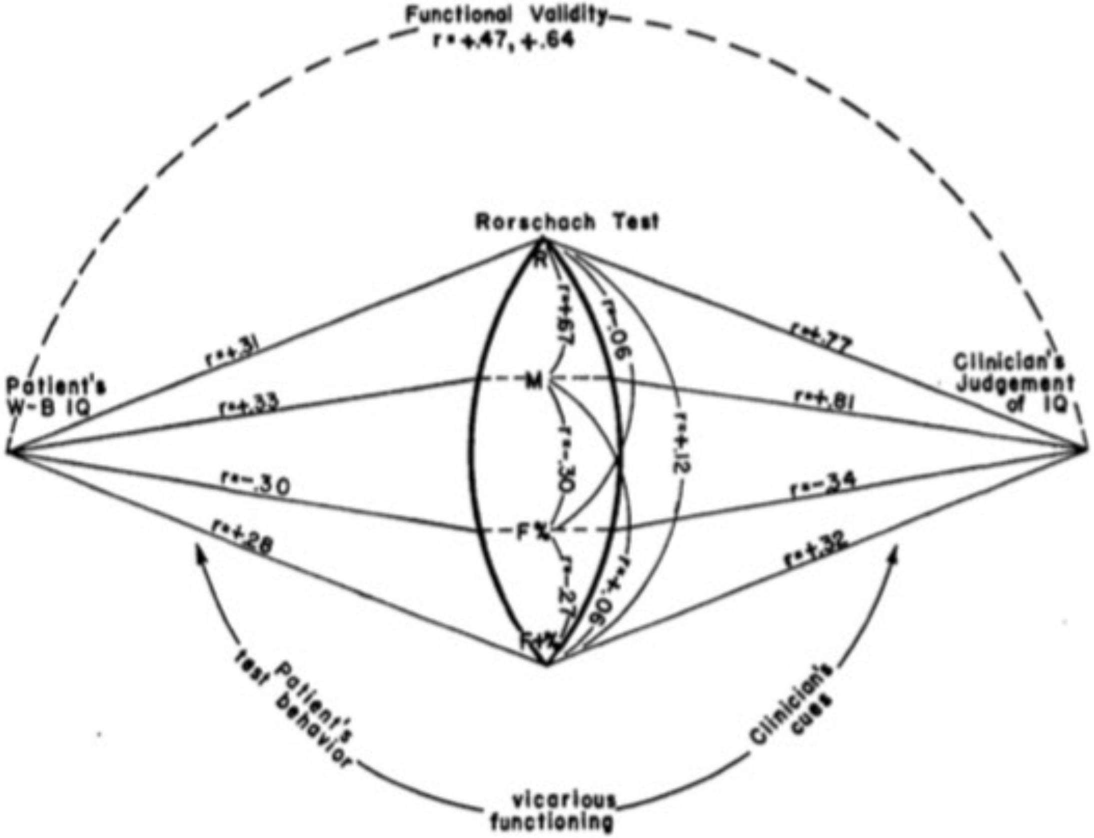


FIG. 1. Functional validity and mediating factors in clinicians' judgments of IQ from the Rorschach test.

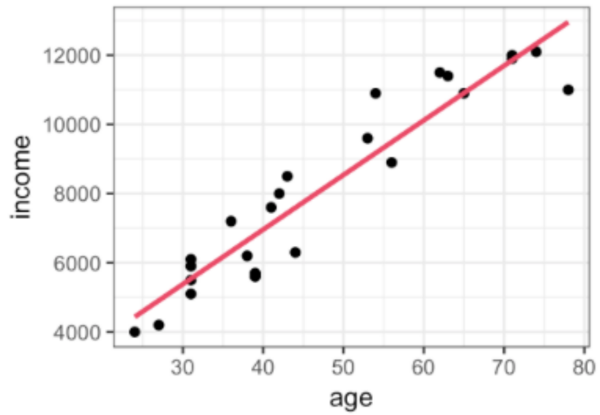
Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62*(4), 255–262.

Dhami, M.K, & Mumpower, J.L. (2018). Kenneth R. Hammond's contributions to the study of judgment and decision making. (2018). *Judgment and Decision Making, 13*(1), 1–22.

Simple Linear Regression

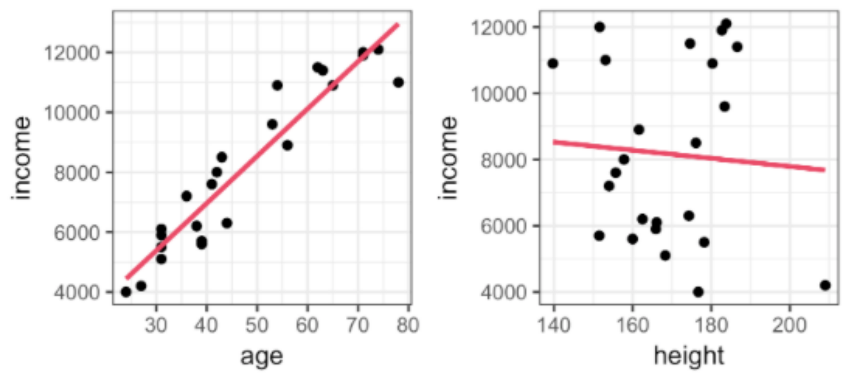
Definition: Simple linear regression is a linear model with one predictor x, and where the error term ϵ is Normally distributed.

$$y = \beta_0 + \beta_1 x + \epsilon$$



Multiple Linear Regression

Definition: Multiple linear regression is a linear model with many predictors $x_1, x_2, \ldots. x_n$, and where the error term ϵ is Normally distributed.



$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

Parameter	Description	In words
β_0	Intercept	When all x values are 0, what is the predicted value for y?
β_1, β_2, ...	Coefficient for x_1, x_2, ...	For every increase of 1 in coefficient for x_1, x_2, ... how does y change?

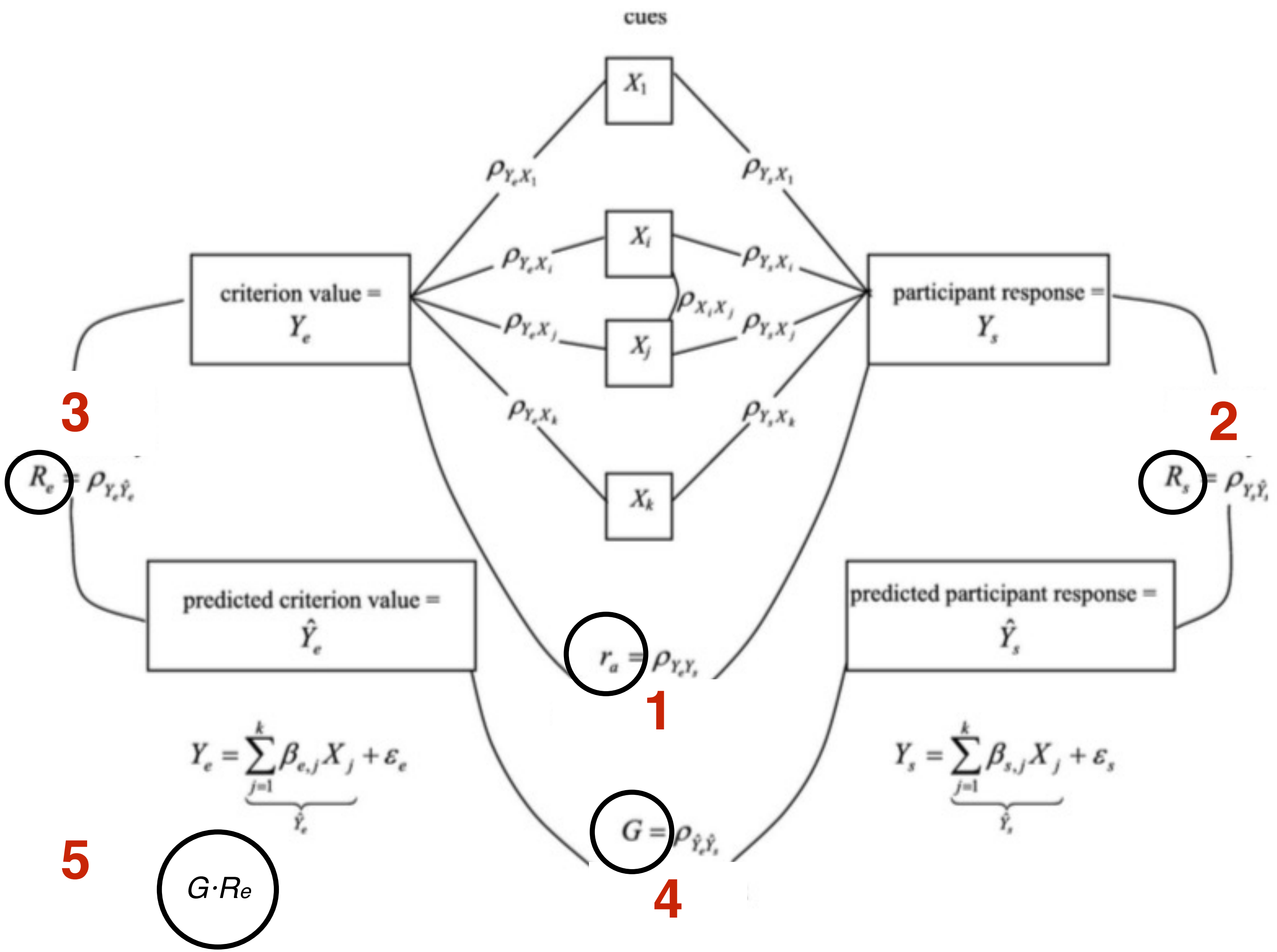
Formula

$$income = 1628 + 147 \times age - 4.1 \times height + \epsilon$$

Coefficients

$$\beta_0 = 1628, \beta_{age} = 147, \beta_{weight} = -4.1$$

The Lens Model and Policy Capturing



Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*(3), 404–426. http://doi.org/10.1037/0033-2909.134.3.404

Clinical Prediction: Experts are inconsistent

"Proper linear models are those in which predictor variables are given weights in such a way that the resulting linear composite optimally predicts some criterion of interest; examples of proper linear models are standard regression (…). Research summarized in Paul Meehl's book on clinical versus statistical prediction—and a plethora of research stimulated in part by that book—all indicates that when a numerical criterion variable (e.g., graduate grade point average) is to be predicted from numerical predictor variables, proper linear models outperform clinical intuition. Improper linear models are those in which the weights of the predictor variables are obtained by some nonoptimal method; for example, they may be obtained on the basis of intuition, derived from simulating a clinical judge's predictions, or set to be equal. This article presents evidence that even such improper linear models are superior to clinical intuition when predicting a numerical criterion from numerical predictors."

Paramorphic and improper models beat the experts!

Dawes, R. (1979). The robust beauty of improper linear models in decision making. ***American Psychologist, 34***(7), 571–582.

Clinical Prediction: Experts are inconsistent

Paramorphic and simple models beat the experts!

equal weighting model: models in which weights are 1 but sign is consistent with those from linear model

optimal linear model: linear regression of attributes on criterion

random model: models in which weights were randomly chosen except for sign (which is obtained from linear model)

bootstrapping model: build a paramorphic model of the judge's judgements by linking attributes to the judge's estimated criterion

Correlations Between Predictions and Criterion Values

Example	Average validity of judge	Average validity of judge model	Average validity of random model	Validity of equal weighting model	Validity of optimal linear model
Prediction of neurosis vs. psychosis	.28	.31	.30	.34	.46
Illinois students' predictions of GPA	.33	.50	.51	.60	.69
Oregon students' predictions of GPA	.37	.43	.51	.60	.69
Prediction of later faculty ratings at Oregon	.19	.25	.39	.48	.54
Yntema & Torgerson's (1961) experiment	.84	.89	.84	.97	.97

Note. GPA = grade point average.

Dawes, R. (1979). The robust beauty of improper linear models in decision making. ***American Psychologist, 34***(7), 571–582.

Clinical Prediction: Experts are <u>inconsistent</u>

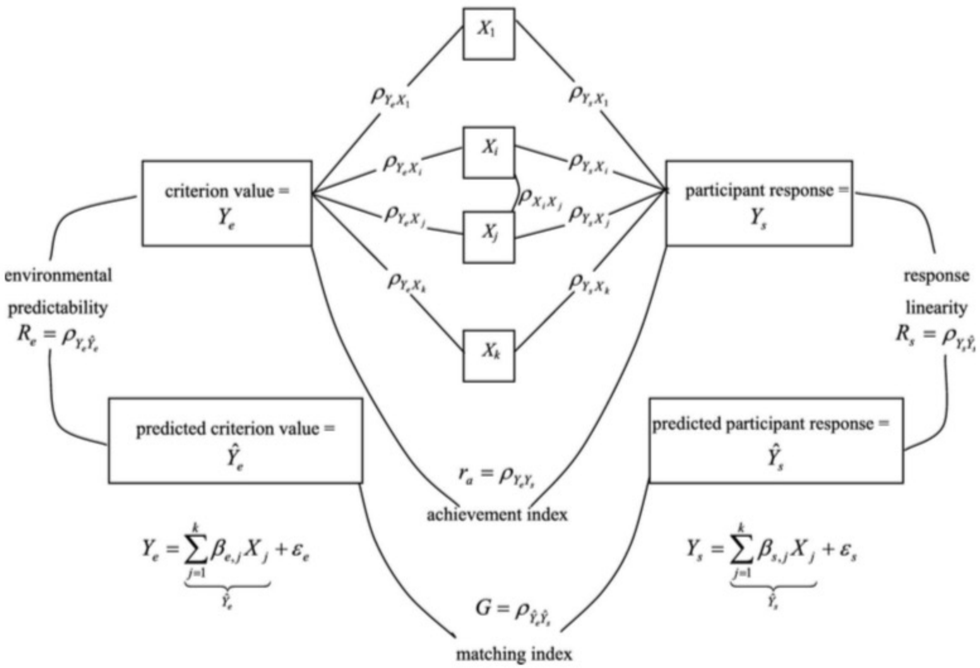
Surely we all know that the human brain is poor at weighting and computing. When you check out at a supermarket, you don't eyeball the heap of purchases and say to the clerk, "Well it looks to me as if it's about $17.00 worth; what do you think?" The clerk adds it up. There are no strong arguments . . . from empirical studies . . . for believing that human beings can assign optimal weights in equations subjectively or that they apply their own weights consistently.

It might be objected that this analogy, offered not probatively but pedagogically, presupposes an additive model that a proponent of configural judgment will not accept. Suppose instead that the supermarket pricing rule were, "Whenever both beef and fresh vegetables are involved, multiply the logarithm of 0.78 of the meat price by the square root of twice the vegetable price"; would the clerk and customer eyeball that any better? Worse, almost certainly. When human judges perform poorly at estimating and applying the parameters of a simple or component mathematical function, they should not be expected to do better when required to weight a complex composite of these variables.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668–1674.

The Lens Model

Karelaia and Hogarth conducted a meta-analysis of 86 field and experimental studies (249 between-subject conditions) that estimated lens model parameters



Index	Label	Mean *r*	k
r_a	achievement	.56	249
G	matching	.80	236
R_e	env predict.	.81	246
R_s	consistency	.80	237
$G \cdot R_e$	bootstrapping	.65	236
$G \cdot R_e - r_a$	bootstrapping - achievement	.10	236

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*(3), 404–426. http://doi.org/10.1037/0033-2909.134.3.404

Clinical vs. Statistical Prediction: Meta-Analyses

Domain	Improvement	Reference
"Human health and behaviour" (e.g., psychology, medicine, forensics, finance)		Grove, W.M., Zald, D.H., Hallberg, A.M., Lebow, B., Snitz, E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. ***Psychological Assessment, 12***, 19–30.
"Counseling psychology" (e.g., diagnosis, prognostic in therapy)		Ægisdóttir, S., White, M. J., & Spengler, P. M. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. ***The Counseling Psychologist, 34***, 341-382.
"Employee selection and academic admission"		Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. ***Journal of Applied Psychology, 98***(6), 1060–1072.

Clinical vs. Statistical Prediction: Meta-Analyses

Table 1
Studies Contributing to the Meta-Analysis

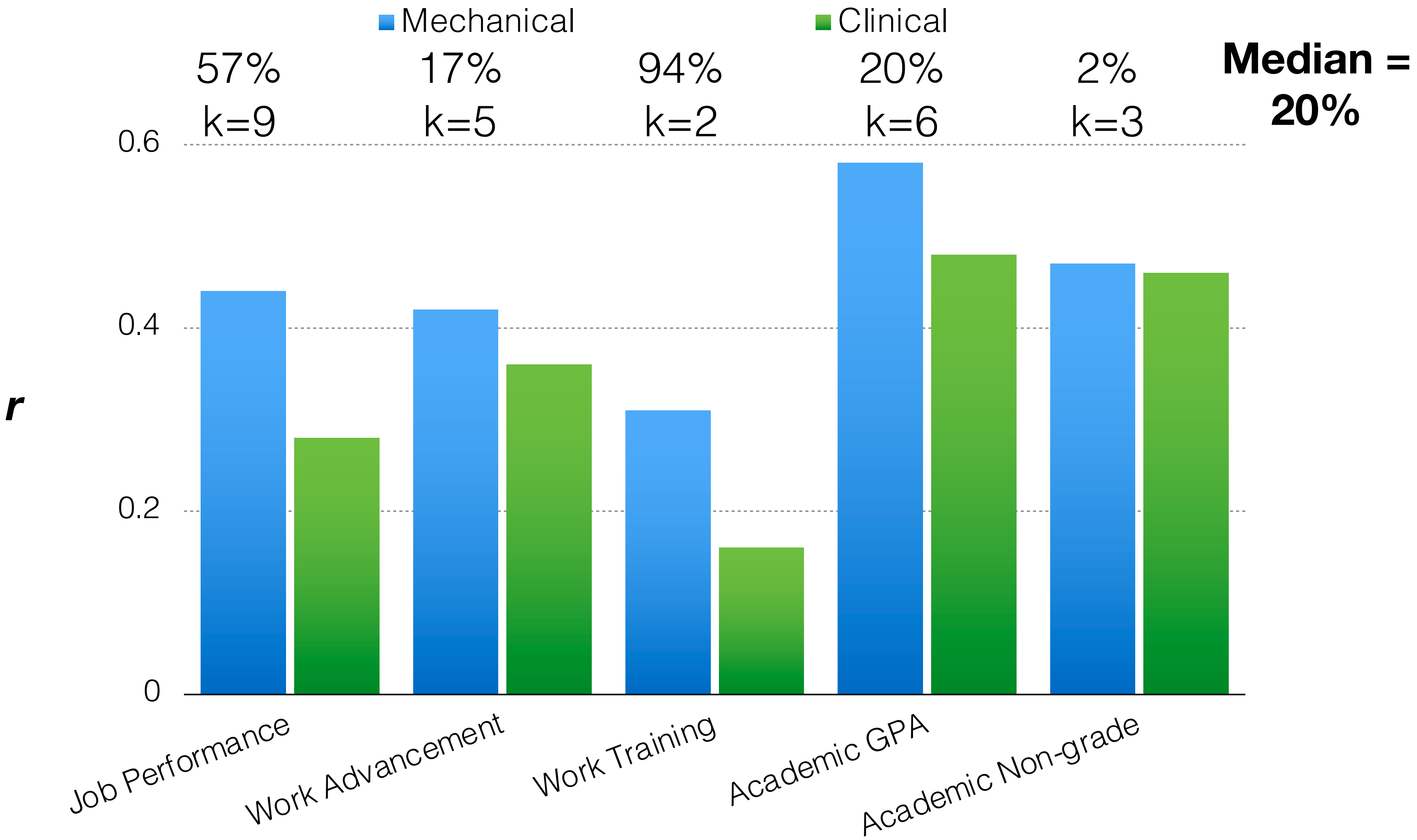
Analyses in which included	Authors (Year)	Criteria	Predictors	$Type_{mech}$	r_{mech}	N_{mech}	$Type_{clin}$	r_{clin}	N_{clin}
1. Acad.–GPA	Sarbin (1943)	Acad. Ach.	Achievement^{1,2}; cognitive ability^{1,2}; vocational interest^{1,2}; personality^{1,2}; records^{1,2}; interviews^{1,2}	MR	0.70	89	Judgment of clinical counselors	0.69	89
2. Acad.–GPA	Sarbin (1943)	Acad. Ach.	Achievement^{1,2}; cognitive ability^{1,2}; vocational interest^{1,2}; personality^{1,2}; records^{1,2}; interviews^{1,2}	MR	0.45	73	Judgment of clinical counselors	0.35	73
3. Acad.–GPA	Stuit (Ed.) (1947)	Acad. Ach.	Cognitive ability^{1,2} and interviews (with predictor scores available to interviewer)^{2}	C & r	0.50	3,246 (73)^{3}	Interview	0.41	3,246 (89)^{3}
4. Acad.–Non-Grade	Truesdell & Bath (1957)	Acad. persistence	Achievement^{1,2}; vocational interests (inventory)^{2}; vocational interests (subscales)^{1}; personality (inventories)^{2}; personality (subscales)^{1}	DF	0.50	314	Average of validities of academic staff judgments	0.42	100

table continues…

Note. $Type_{mech}$ = type of mechanical data combination; r_{mech} = observed correlation for the mechanical data combination method; N_{mech} = number of persons for whom a mechanical data combination method was used to make a prediction; $Type_{clin}$ = type of clinical data combination; r_{clin} = observed correlation for the clinical data combination method; and N_{clin} = number of persons for whom a clinical data combination method was used to make a prediction. For Analyses column, Acad. = academic; GPA = grade point average. For "Criteria" column, Ach. = achievement. For "Predictors" column, superscript 1 = used in mechanical data combination, and superscript 2 = available to clinician for clinical data combination. For "$Type_{mech}$" column, MR = multiple regression; r = correlation; DF = discriminant function; C = compositing; C-B = bootstrapped compositing; P = pooling; S = summation; A = averaging; UW = unit-weighting; DW = differential weighting; and BW = bootstrapped weighting. For "N_{mech}" and "N_{clin}" columns, superscript 3 = *N* used for meta-analysis appears in parentheses and was the median of the *N*s of the other studies in the analysis for which an *N* was known. For the studies for which the median was used as the *N* in the meta-analysis, either (a) the original source materials for the study could not be located, but we knew the effect size and other pertinent information except for the *N*, or (b) the actual *N* was so large that if it were used in the meta-analyses other than for estimating study-specific sampling error, then it would mathematically overwhelm the results.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. ***Journal of Applied Psychology, 98***(6), 1060–1072. doi:10.1037/a0034156

Clinical vs. Statistical Prediction: Meta-Analyses



Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. ***Journal of Applied Psychology, 98***(6), 1060–1072. doi:10.1037/a0034156

Summary

- **Clinical vs. actuarial judgment:** clinical judgment as the integration of data in the head; actuarial judgment as the integration of data through an algorithm; problem of integration NOT data (although that could be an issue too)…
- **Why humans fail**: increased experience may not be strongly related to improved performance (lack of immediate/appropriate feedback?), (over)confidence, incorrect or inconsistent weighting
- **Lens model and policy capturing:** Lens model as a general depiction of data integration; supplies framework and terminology to help assess the relative benefits of clinical and actuarial judgment; formalisation of judgment process using an algorithm (regression model)
- **Empirical evidence**: Meta-analyses of field studies in several domains (e.g., academic, mental health) suggest that actuarial judgment can outperform clinical judgment by ca. 20% or more

Policy capturing in practice

https://evidencebaseddm.formr.org