Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Wissenschaftstheorie & Einführung in das wissenschaftliche Arbeiten

Erkenntnistheorie & Experimentalpsychologie:
Theory / Phenomena / Statistical models / Data & Ethics

Dr. Blazej Baczkowski (Błażej Bączkowski)

New hope — Bayesian epistemology

- The old problem: What is it for an observation to provide evidence, or confirm, a scientific theory?
- Proposed solutions:
  - Logical empiricism uses inductive logic to help testing theories. The problem remains unsolved.
  - Karl Popper opposes the idea of confirmation and proposes falsification. The problem remains unsolved.
- Bayesianism — a new theory of confirmation / evidence (you can think of evidence as data / observations).

Bayesianism
— updating the “degree of belief” in light of new evidence

- (1) Uncertainty (or “degree of belief”) is quantified by probability
- (2) The observed data are used to update the *prior* beliefs to become *posterior* beliefs.
  - this update is done according to the Bayes’ theorem
  - evidence confirms a hypothesis if it makes the hypothesis more probable than it would otherwise be
  - (evidence dis-confirms a hypothesis if it makes the hypothesis less probable than it would otherwise be)
- **BAYESIAN INFERENCE IS REALLOCATION OF CREDIBILITY ACROSS POSSIBILITIES**
  (J. Kruschke)

Bayesianism
— updating the “degree of belief” in light of new evidence

Posterior belief
about the hypothesis, i.e.
after having obtained evidence

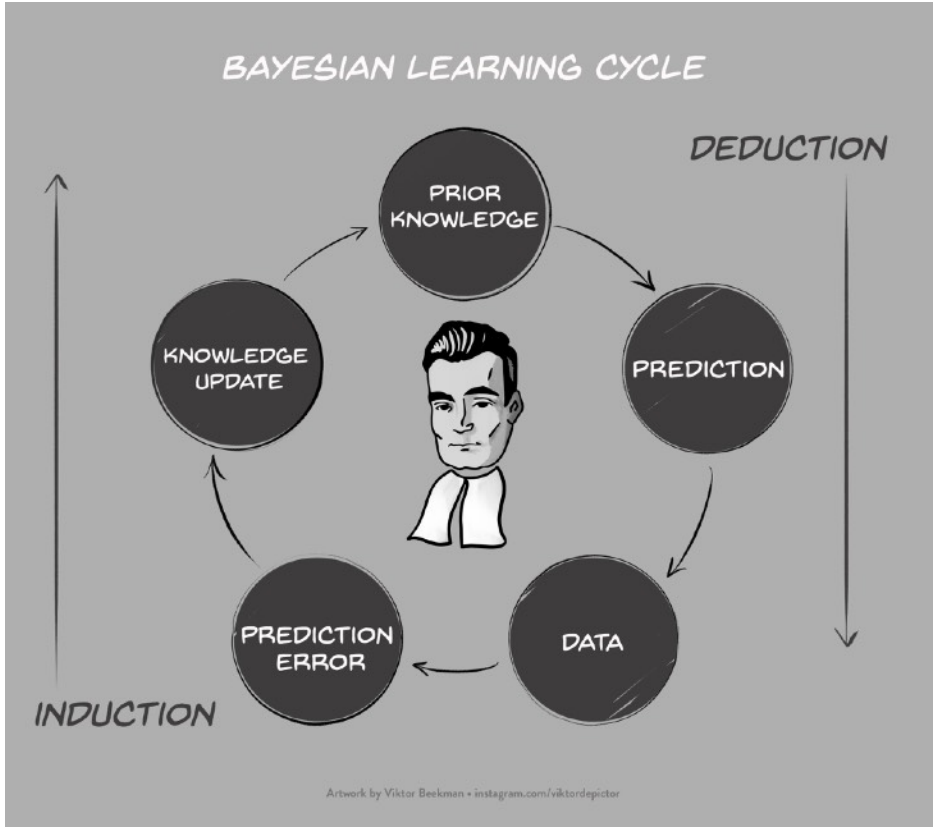
$$p(hypothesis \mid evidence) = p(hypothesis) \times \frac{p(evidence \mid hypothesis)}{p(evidence)}$$

Prior belief about the hypothesis, i.e.,
before any evidence

Predictive updating factor

$$p(evidence) = \sum_{i=1}^{n} p(hypothesis_i) \times p(evidence \mid hypothesis_i)$$

Gradual accumulation of knowledge



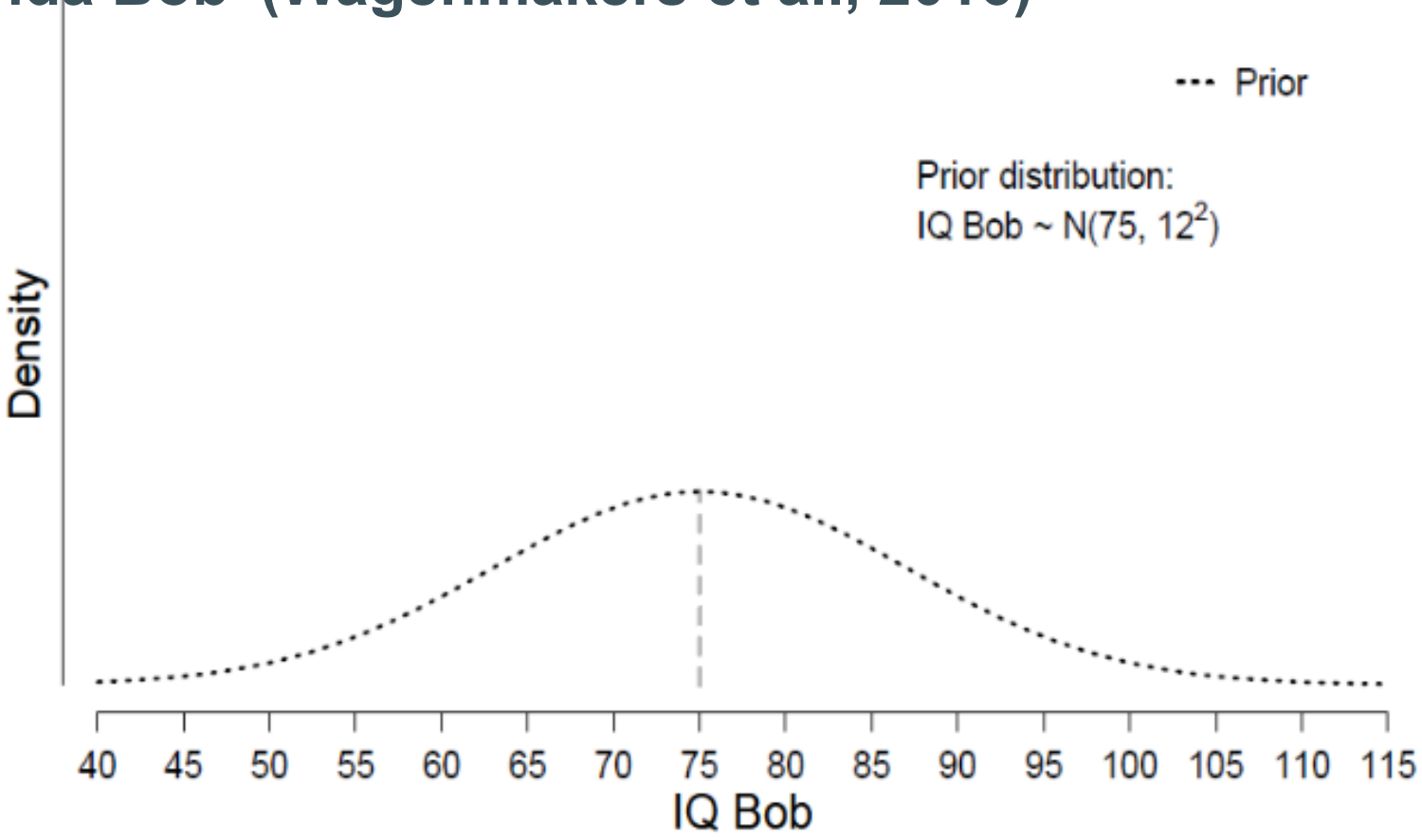
‘Florida Bob’ (Wagenmakers et al., 2016)

- ‘Florida Bob’ killed his wife and may get the death penalty.
- The defence claims that Bob is intellectually disabled: Bob’s IQ is lower than 70.
- Indeed, 20 years earlier, when Bob was incarcerated for a different crime, an IQ test administered upon his entry indicated that he was intellectually disabled. But it is known that such IQ tests tend to underestimate prisoners’ IQs.
- The judge orders three new IQ tests yielding scores of 73, 67, and 79.
- What is the probability that Bob’s IQ is lower than 70?

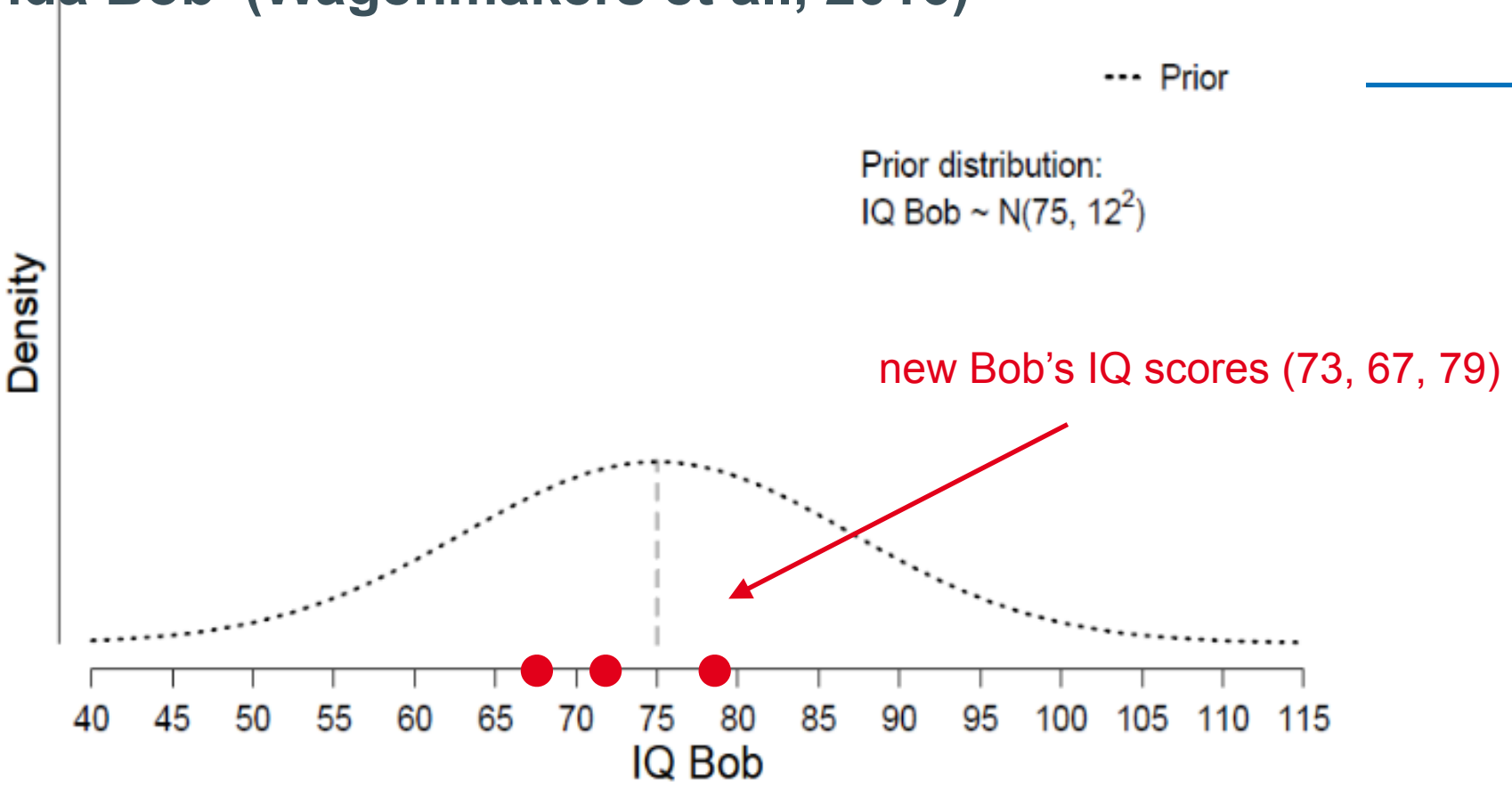
‘Florida Bob’ (Wagenmakers et al., 2016)

- Prior knowledge:
  The literature shows that, as a group, inmates classified as ‘intellectually disabled’ have an IQ of mean = 75 and standard deviation = 12.

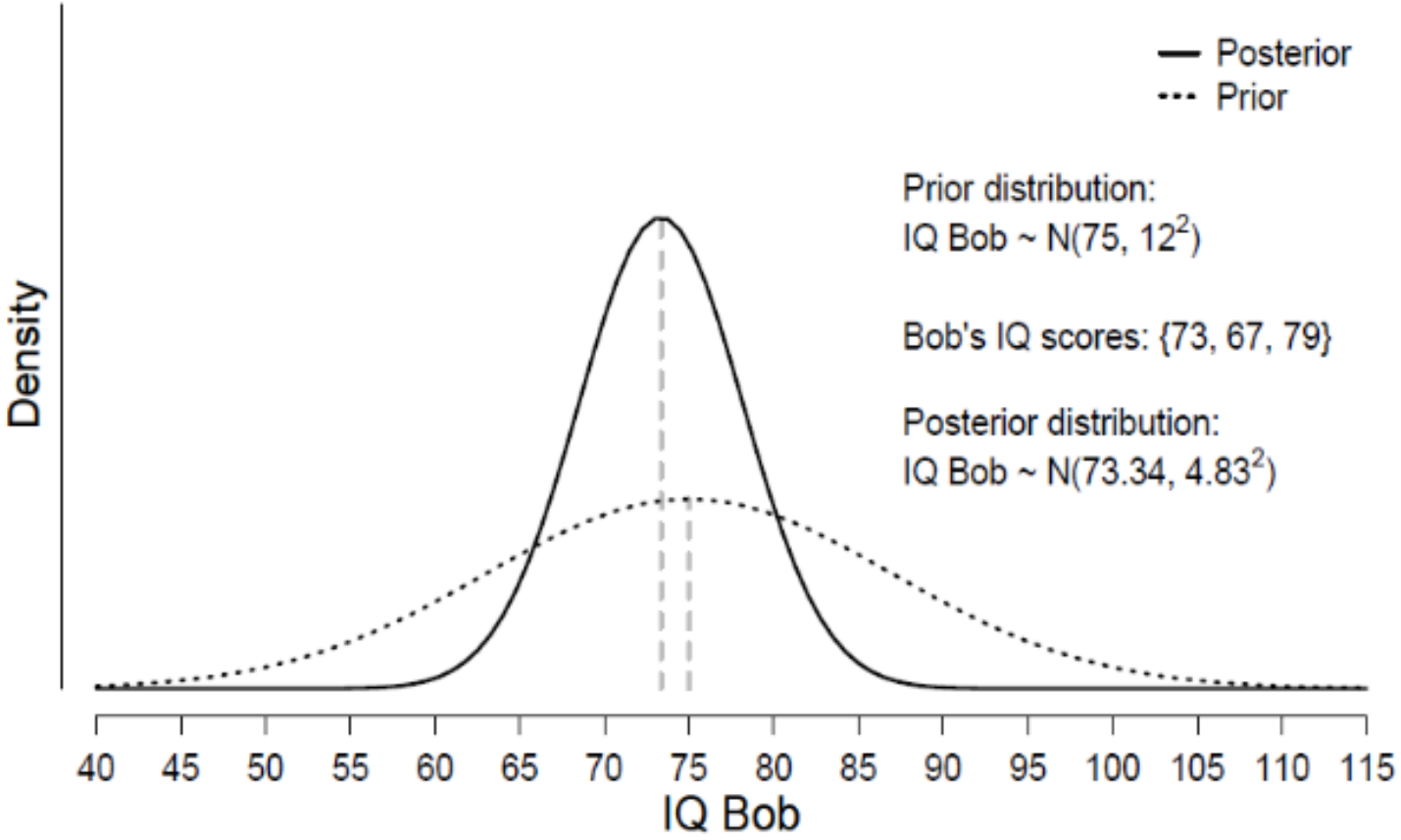
‘Florida Bob’ (Wagenmakers et al., 2016)



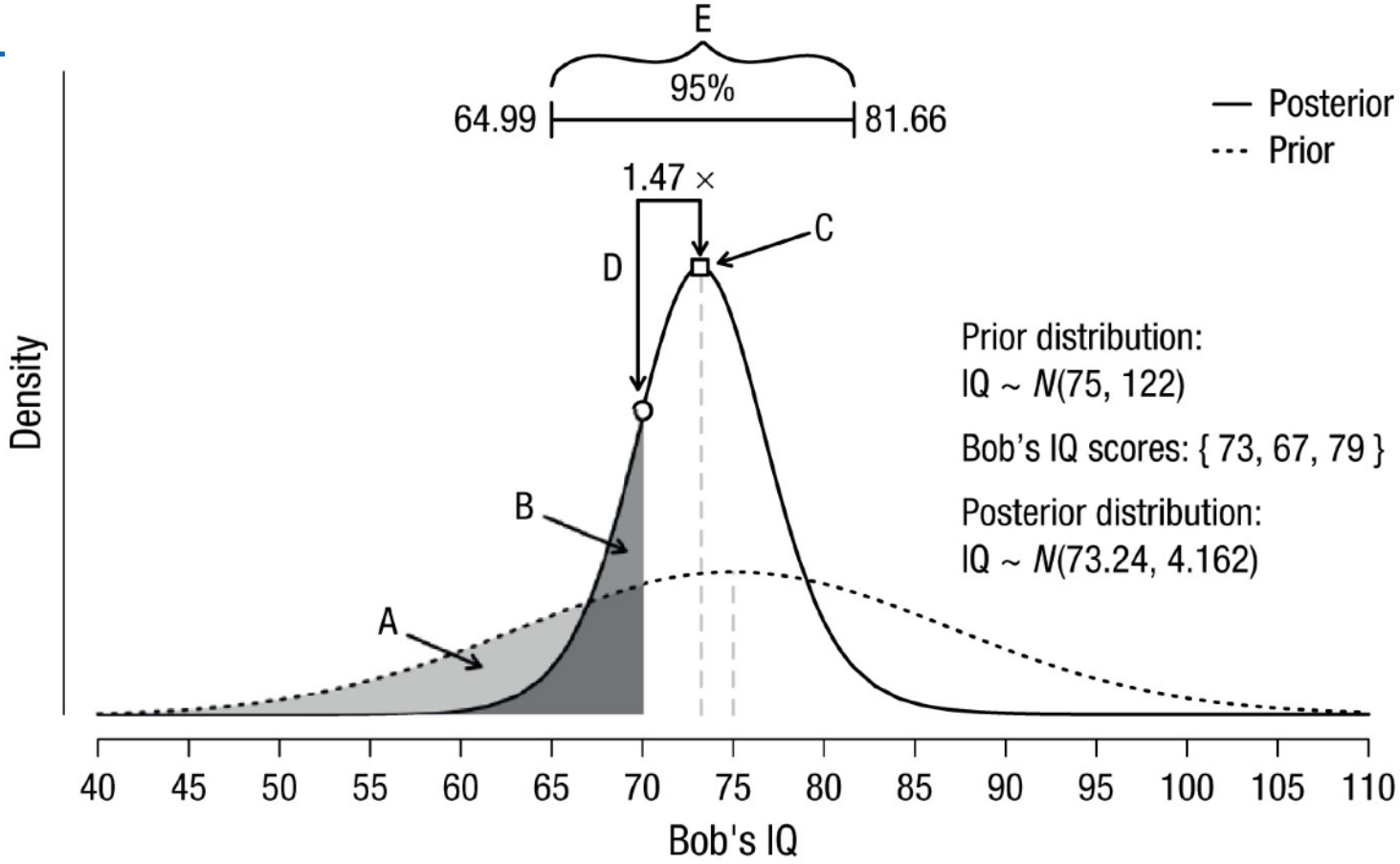
‘Florida Bob’ (Wagenmakers et al., 2016)



‘Florida Bob’ (Wagenmakers et al., 2016)



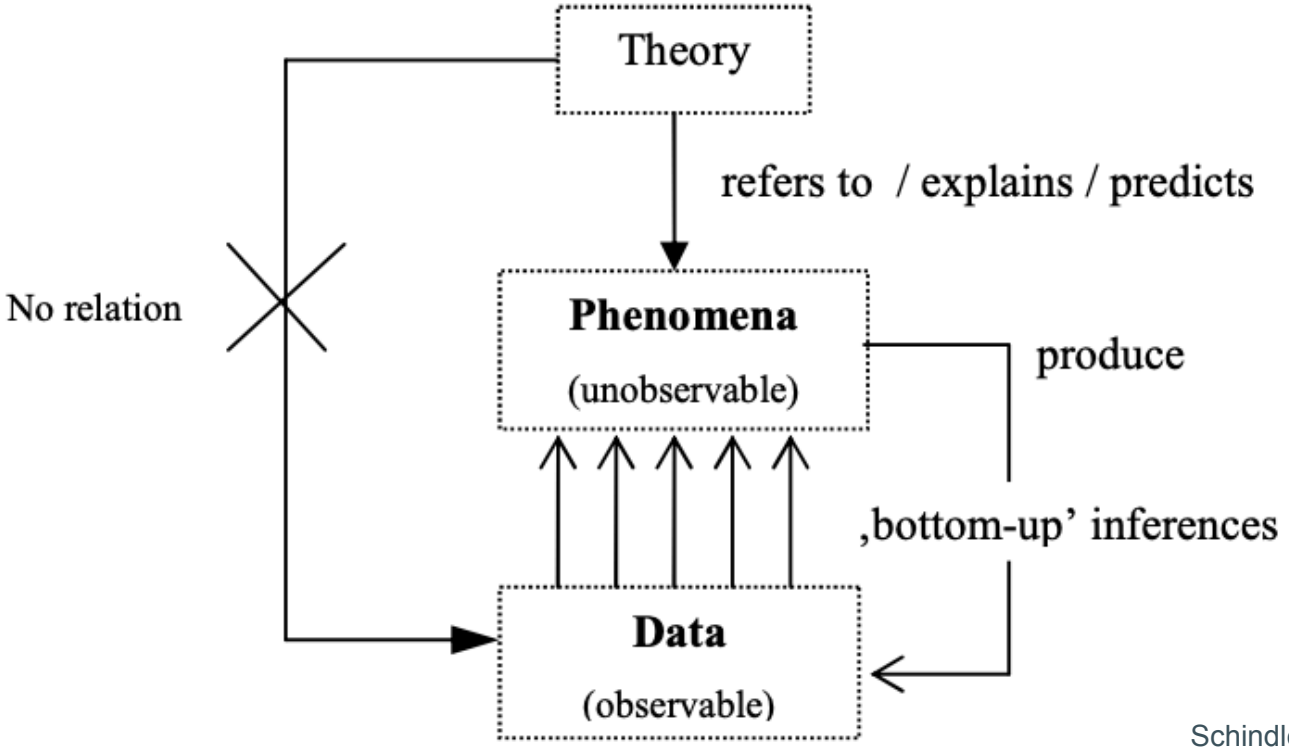
‘Florida Bob’ (Wagenmakers et al., 2016)



‘Florida Bob’ — some comments

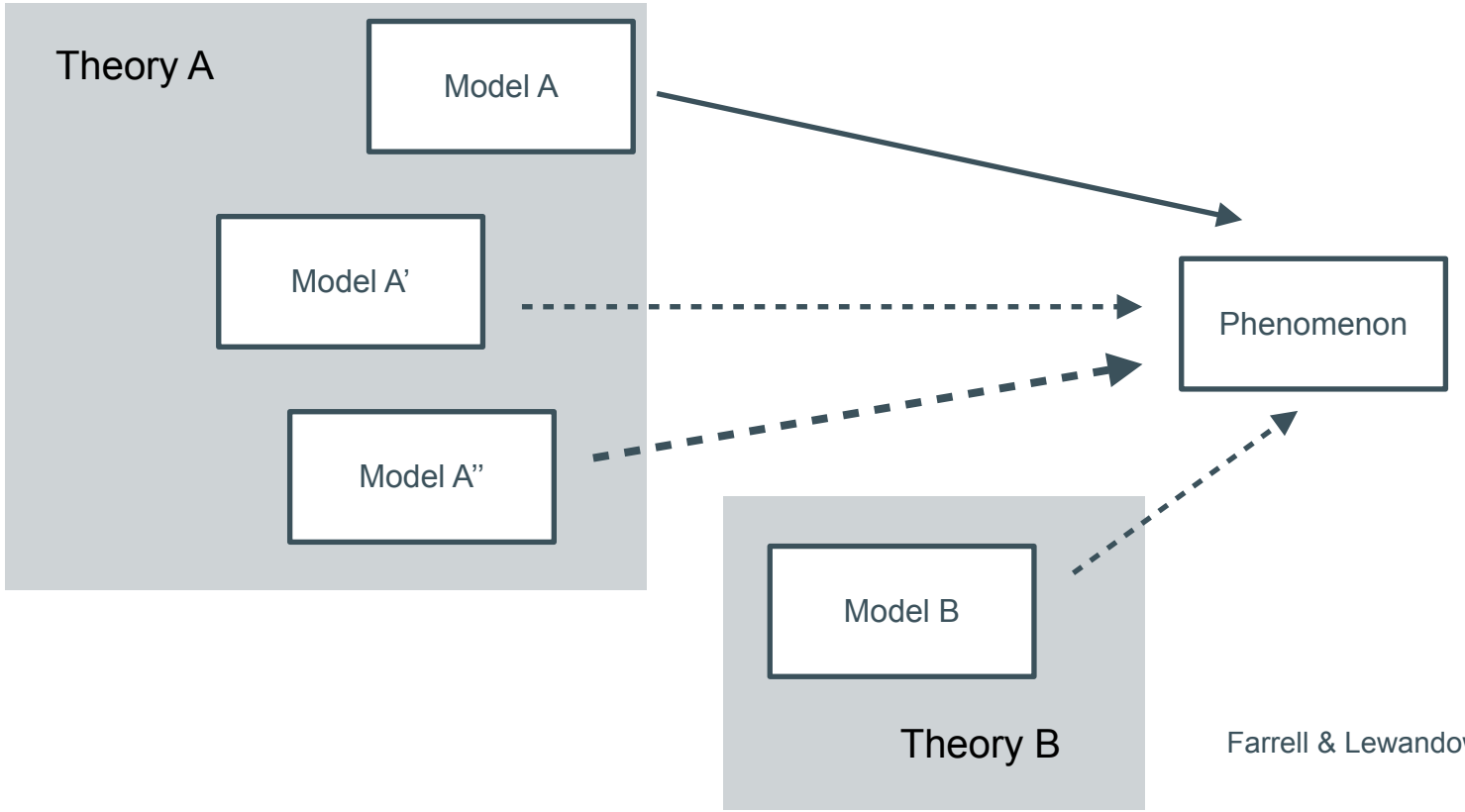
- Bob’s observed IQ scores are determined both by his **latent** (unobserved) *intellectual ability* and by the *reliability* of the IQ test.
- The former requires a ‘theory’ about what intelligence is (cognitive psychology) while the later requires a ‘theory’ about how to measure intelligence (psychometrics)
  - the judge presupposes (a) that intelligence affects behaviour in a principled way (the theory of intelligence explains and predicts behaviour)
  - the judge presupposes (b) that the IQ test measures Bob’s intelligence accurately (the theory of measurement captures latent psychological constructs well)
- But how is theory linked to data???

The relation between theory and observation



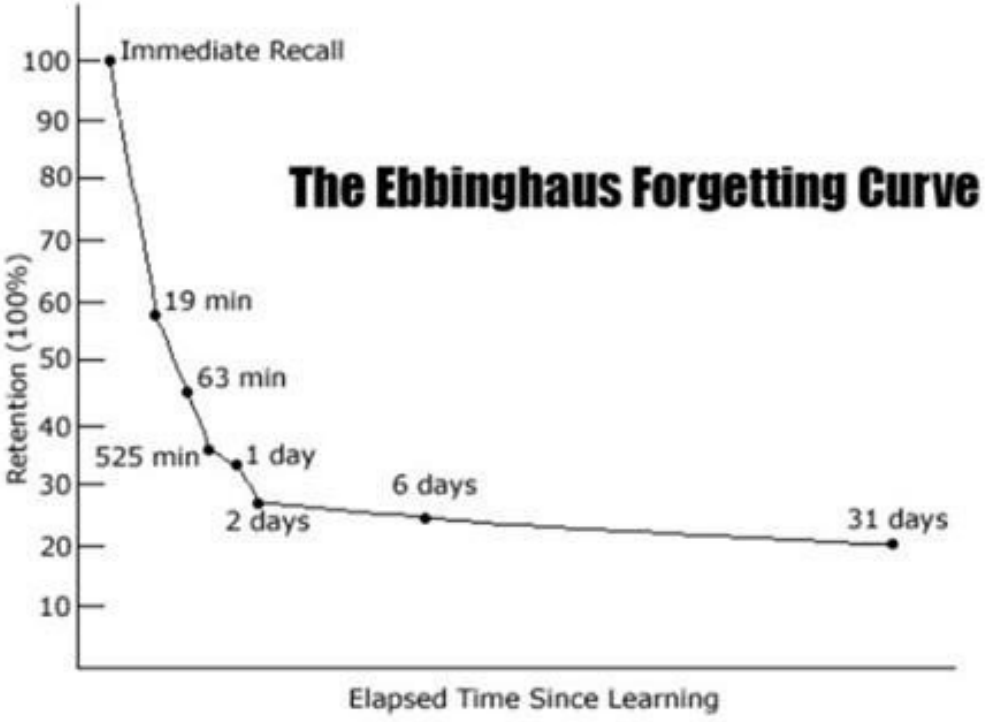
Schindler (2008)

Using statistical models in psychology — models compete to account for phenomena



Farrell & Lewandowsky (2018), modified

Phenomena
— forgetting curve



Memory retention (Lee & Wagenmakers, 2013)

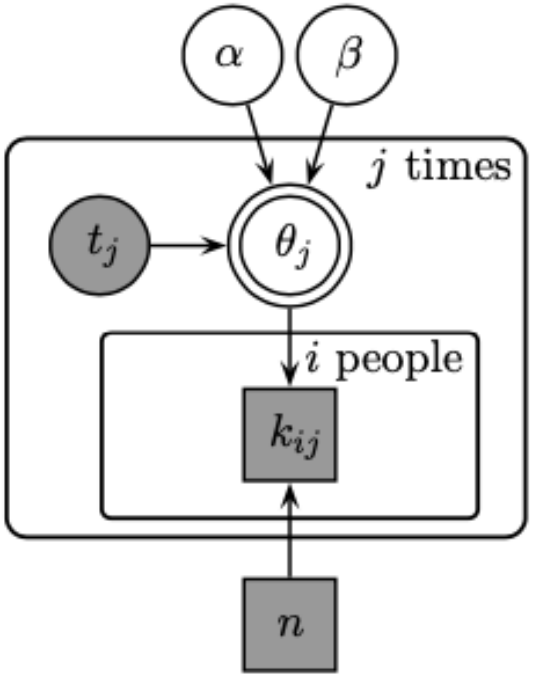
Table 10.1 Memory retention data for 4 subjects and 10 time intervals.

	Time Interval									
Subject	1	2	4	7	12	21	35	59	99	200
1	18	18	16	13	9	6	4	4	4	?
2	17	13	9	6	4	4	4	4	4	?
3	14	10	6	4	4	4	4	4	4	?
4	?	?	?	?	?	?	?	?	?	?

Table 9.1 Notation for nodes used in graphical models

Status of Variable	Type of Variable: Discrete	Type of Variable: Continuous
Observed		
Unobserved		
Stochastic		
Deterministic		

Memory retention (Lee & Wagenmakers, 2013) — graphical model for the exponential decay of memory retention, assuming no individual differences



$$\alpha \sim \text{Beta}(1,1)$$ the rate of decay

$$\beta \sim \text{Beta}(1,1)$$ baseline level of remembering

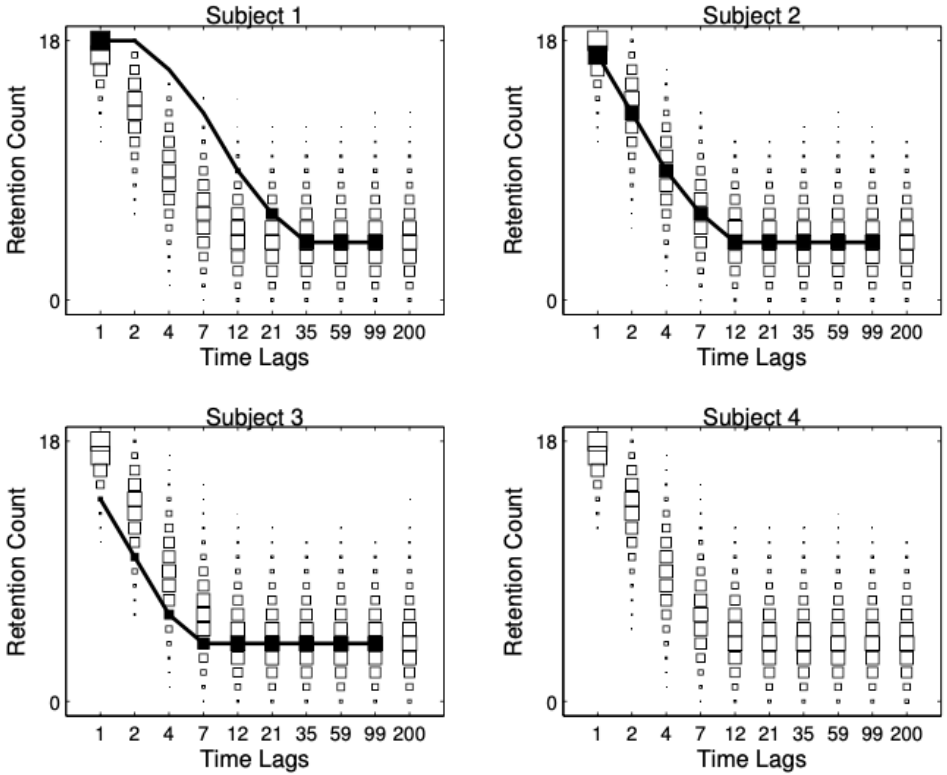
$$\theta_j \leftarrow \min\left(1, \exp\left(-\alpha t_j\right) + \beta\right)$$

$$k_{ij} \sim \text{Binomial}(\theta_j, n)$$

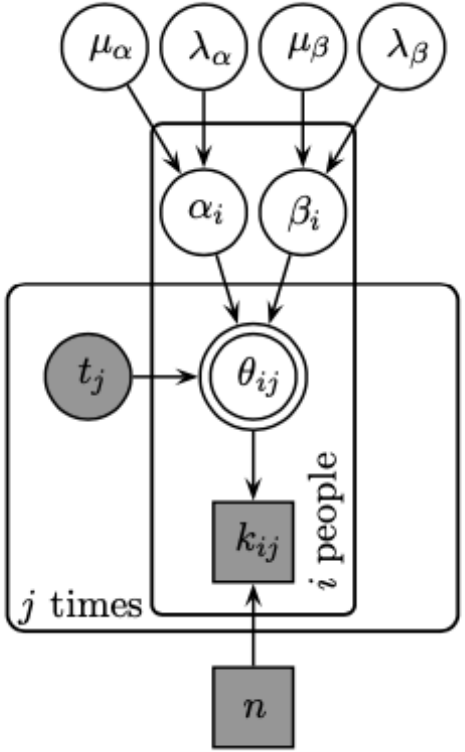
success rate

exponential decay

Memory retention (Lee & Wagenmakers, 2013) — graphical model for the exponential decay of memory retention, assuming no individual differences

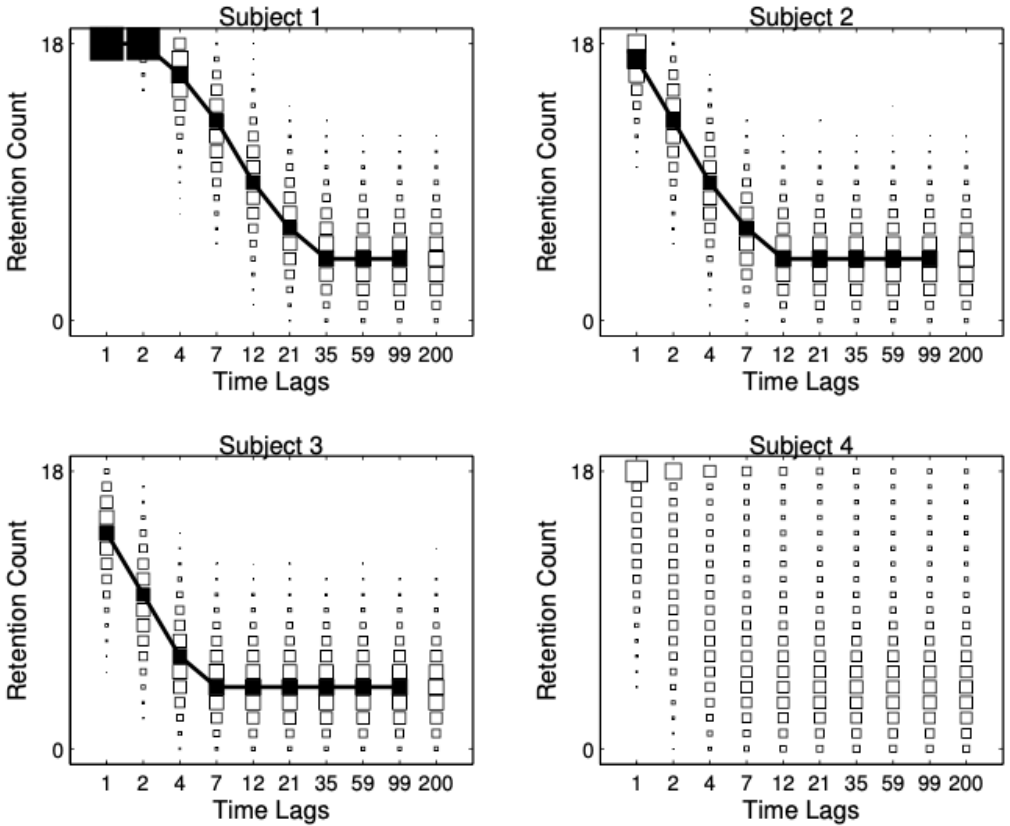


Memory retention (Lee & Wagenmakers, 2013) — graphical model for the exponential decay of memory retention, assuming structured individual differences

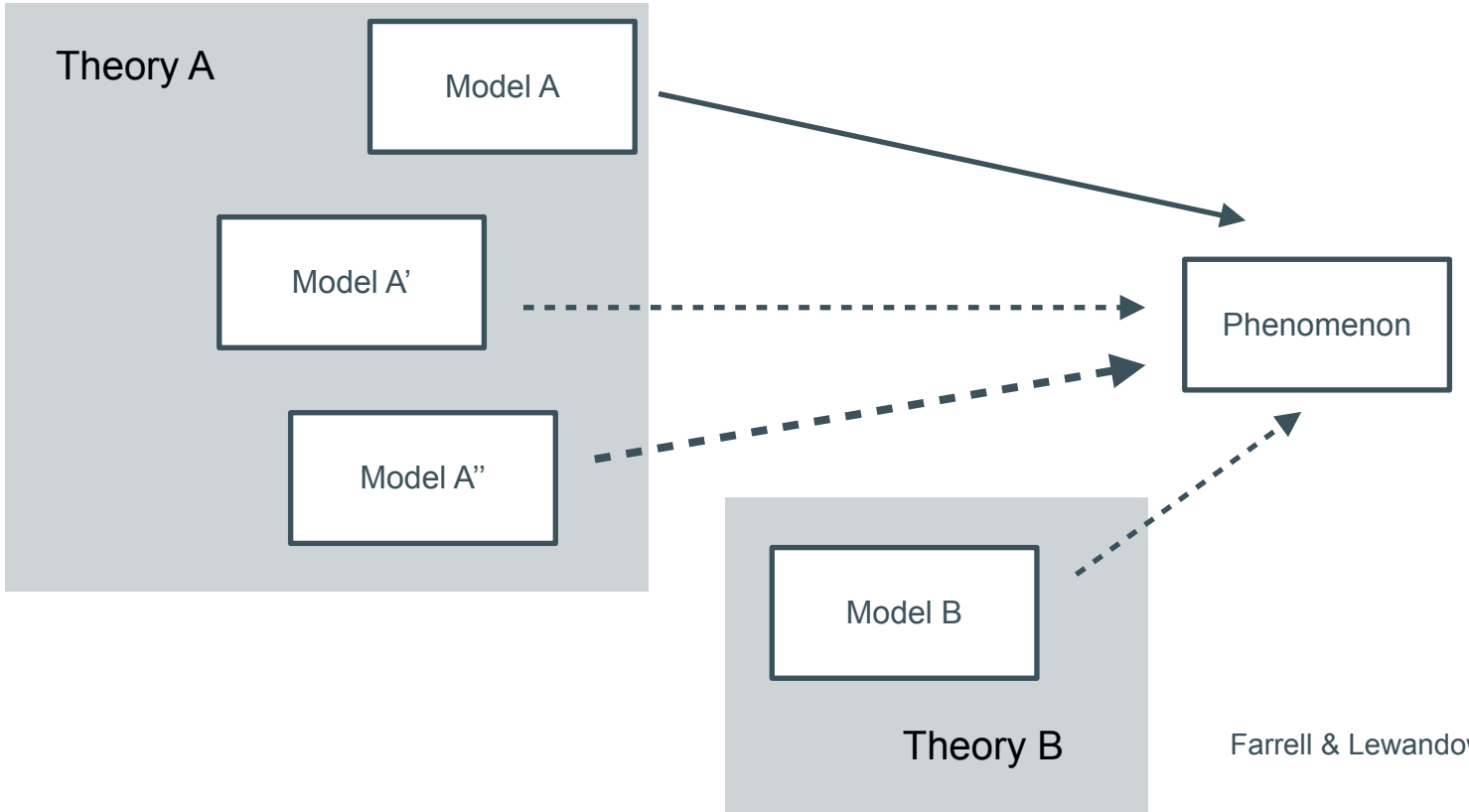


$$\mu_\alpha \sim \text{Beta}(1,1)$$
$$\lambda_\alpha \sim \text{Gamma}(.001,.001)$$
$$\mu_\beta \sim \text{Beta}(1,1)$$
$$\lambda_\beta \sim \text{Gamma}(.001,.001)$$
$$\alpha_i \sim \text{Gaussian}(\mu_\alpha,\lambda_\alpha)_{\mathcal{I}(0,1)}$$
$$\beta_i \sim \text{Gaussian}(\mu_\beta,\lambda_\beta)_{\mathcal{I}(0,1)}$$
$$\theta_{ij} \leftarrow \min\left(1,\exp\left(-\alpha_i t_j\right)+\beta_i\right)$$
$$k_{ij} \sim \text{Binomial}(\theta_{ij},n)$$

Memory retention (Lee & Wagenmakers, 2013) — graphical model for the exponential decay of memory retention, assuming structured individual differences

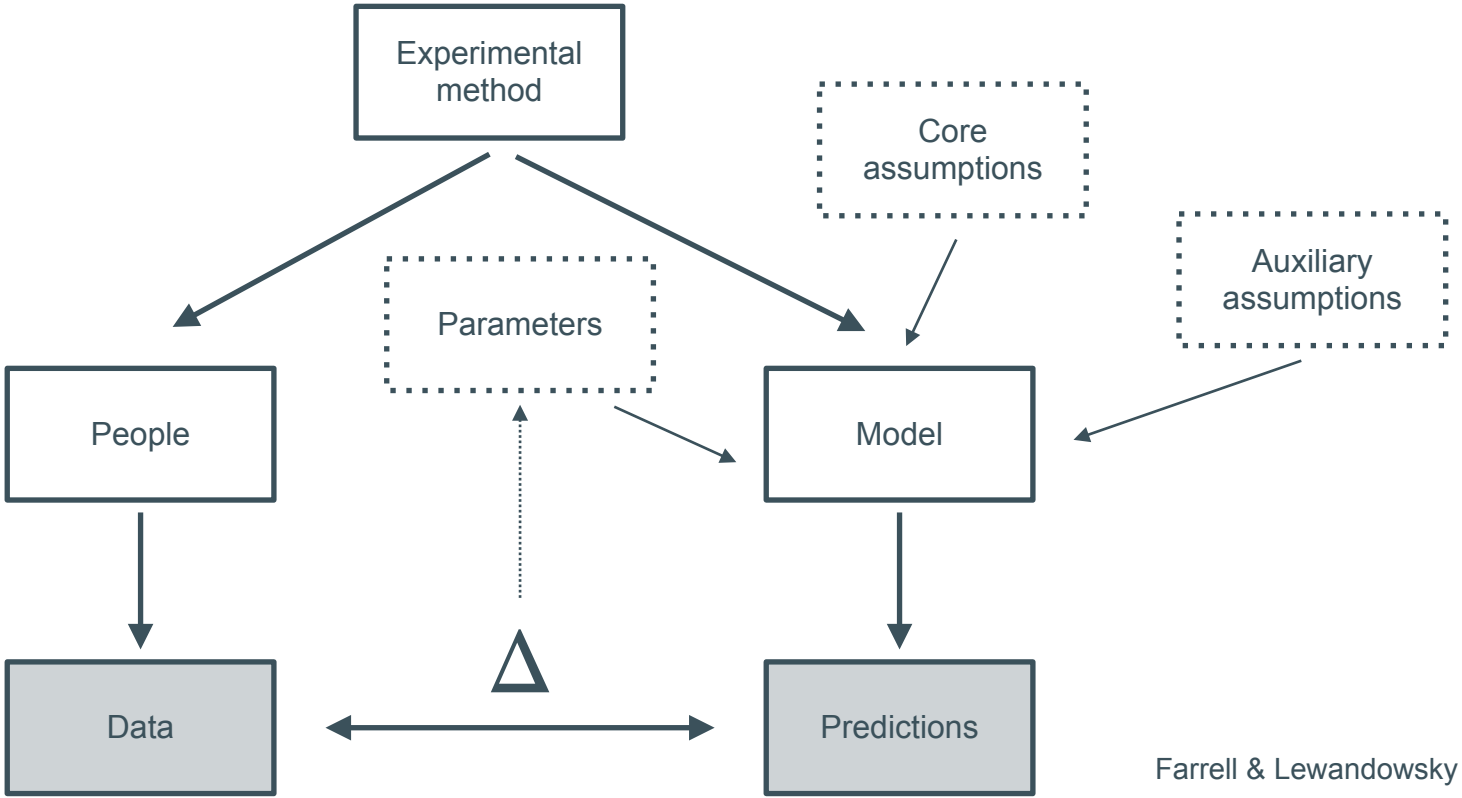


Using models in psychology — models compete to account for phenomena



Farrell & Lewandowsky (2018), modified

Using models in psychology
— model predictions and data are both contingent on experimental method!



Farrell & Lewandowsky (2018)

What are data?

- Data are composed of *variables*.
- A variable is a quantity or quality that varies (it must take on at least two different possible values; otherwise, it is a *constant*).
    - A quality is typically measured by assigning a category label (e.g., “1”=pasta, “2”=pizza, 3=“mango”)
    - A quantity is typically measured by assigning a number (e.g, height, number of students)
        - *binary* numbers are zero or one (true / false; yes / no)
        - *integer* numbers are whole numbers with no fractional or decimal part (e.g., a scale where 1 is “strongly disagree” and 7 is “strongly agree”)
        - *real* numbers are numbers that have a fractional or a decimal part (e.g., weight)
- Values of variables can relate to each other in different ways
    - identity — each value has a unique meaning
    - magnitude — the values reflect different magnitude and have an ordered relationship to one another (larger / smaller)
    - equal intervals — units along the scale of measurement are equal to one another (1 and 2 are equal in magnitude as 18 and 19)
    - absolute zero — the scale has a true meaningful zero point (e.g., speed of zero means absence of the thing being measured)
- These different ways go along with different scales of measurement
    - nominal scale — it satisfies the identity but the numbers are simply labels (e.g., 1=“kid”, 2=“teenager”, 3=“adult”)
    - ordinal scale — satisfies the criteria of identity and magnitude such that the values can be ordered in therms of their magnitude (e.g., level of pain on a scale from 0 to 10). The ordering gives us information about relative magnitude, but the differences between values are not necessarily equal in magnitude.
    - interval scale — An interval scale has all of the features of an ordinal scale, but in addition the intervals between units on the measurement scale can be treated as equal (e.g., temperature in Celsius)
    - ratio scale — A ratio scale variable has all four of the features outlined above. The difference between a ratio scale variable and an interval scale variable is that the ratio scale variable has a true zero point (e.g., weight)

Poldrack (2019)

What are data?

The scale determines what kind of mathematical operations we can apply to the data.

Table 2.2: Different scales of measurement admit different types of numeric operations

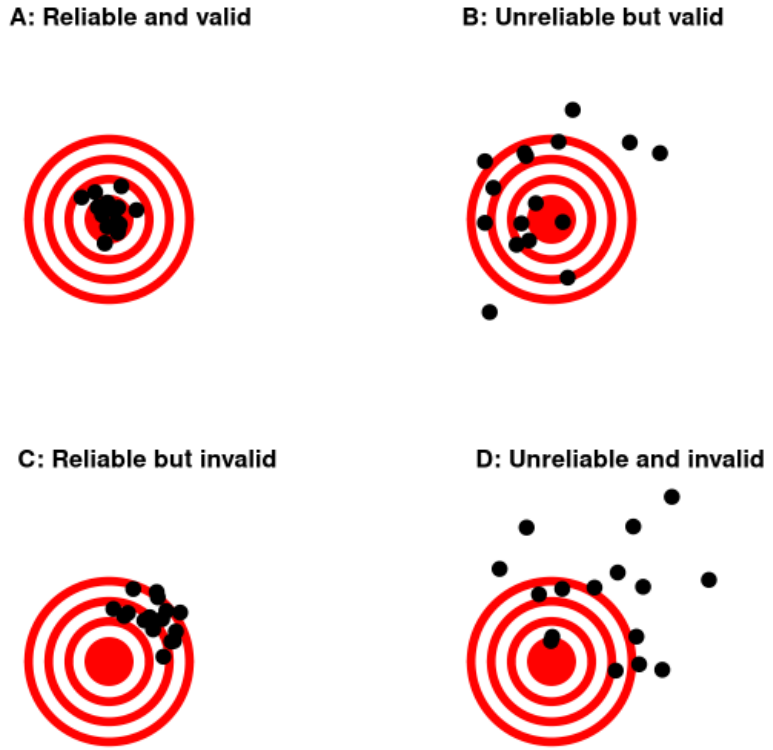
	Equal/not equal	**>/<**	**+/-**	**Multiply/divide**
Nominal	OK			
Ordinal	OK	OK		
Interval	OK	OK	OK	
Ratio	OK	OK	OK	OK

What makes a good measurement?

- Psychology deals with an unobservable theoretical concept (e.g., intelligence), which is usually referred to as a *construct*.
- To “measure” a construct we need an *operational definition* — a way of how a construct is to be measured (e.g., “intelligence” -> IQ test, “fear” -> physiological response, such as heart rate, to a threat)
- Reliability (Reliabilität) — the consistency of measurements (a measure is said to have a high reliability if it produces similar results under consistent conditions)
- Validity (Validität) — measuring the construct we actually want to measure; the extent to which the scores from a measure represent the construct they are intended to.
  - Example: “face validity” — the extent to which a measurement appears “on its face” to measure the construct of interest (consider self-report questionnaires)

What makes a good measurement?

Reliability refers to the consistency of location of shots, and validity refers to the accuracy of the shots with respect to the center of the bullseye.

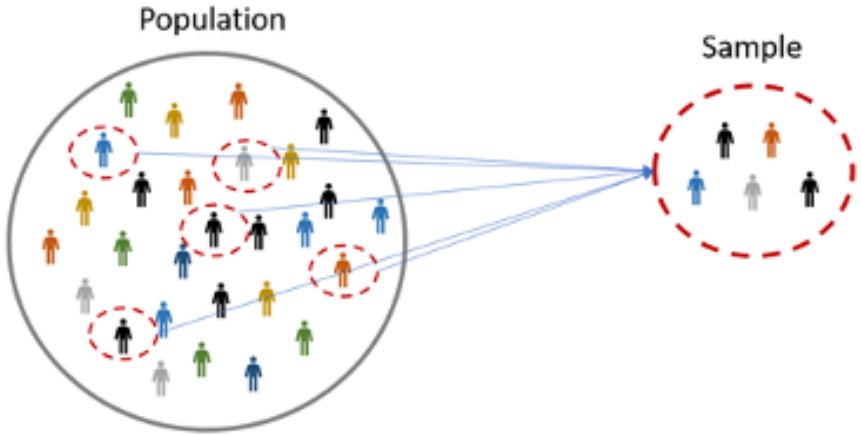


Poldrack (2019)

Who to measure?

We are interested to draw conclusions about **population** but can test only its subset, i.e., a **sample (Stichprobe)**.

- Group research
  - Many individuals from a single group
  - Many individuals from multiple groups
- Single-subject research
  - small-n designs
    - to demonstrate systematic and functional relationships between constructs as manifested at the individual participant
  - case study
    - to generate new research questions and hypothesis



Study design (Forschungsdesign)

- Dependent variables (Abhängige Variable) — their values depend on the values of other variables in the scope of a research question (e.g., reaction time when it depends on age); its variation is being investigated; it's the outcome that we aim to explain.
- Independent variables (Unabhängige Variable) — their values do not depend on the values of other variables in the scope of a research question (e.g., age, sex); we use it to explain the dependent variable.
- Confounding variables (Störvariable) — any variable other than the one being studied that influences, or "confounds", the relationship between the independent and dependent variables (e.g., rate of ice cream is associated with number of sunburns because of the temperature; temperature is a confounding variable; hot temperatures cause people to both eat more ice cream and spend more time outdoors under the sun, resulting in more sunburns.).

Study design (Forschungsdesign)

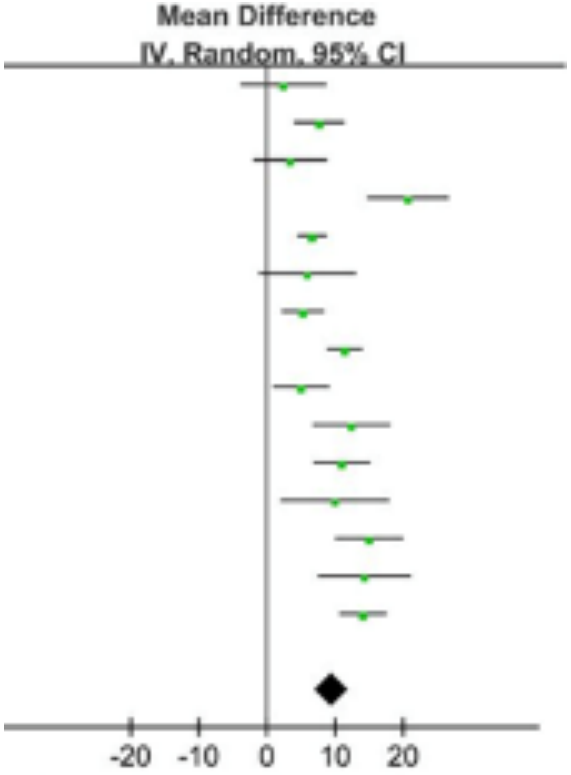
- Correlational (korrelative Studien) — aims to determine the strength of associations between variables without influencing them, e.g., the relation between depression and neuroticism scores
  - no causal relationship
- Experimental (experimentelle Studien) — aims to determine the cause and effect
  - randomised (laboratory) experiments (randomisierte Experimente) — subjects are randomly assigned to experimental conditions; e.g., a subject is randomly assigned to either a drug or placebo condition
  - quasi-experimental (Quasi-Experimente) — non-random method is used to assign subjects to experimental conditions, e.g., one school implements a new teaching program while the other does not approve it, and the children's performance is compared between schools

Randomization (Randomisierung)

- In a **between-subjects design** (Zwischensubjekt), individuals receive only one of the possible levels of an experimental treatment.
- In a **within-subjects design** (Innersubjekt), every individual receives each of the experimental treatments, and their responses to each treatment are measured.
  - **Counterbalancing** (randomizing or reversing the order of treatments among subjects) is often used in within-subjects designs to ensure that the order of treatment application doesn't influence the results of the experiment.

Meta-analysis (Metaanalyse)

- Meta-analysis combines (and synthesises) the results of multiple studies that address the same question in order to assess a pooled estimate closest to the true but unknown estimate.



Research Ethics (Forschungsethik) — obedience experiment by Milgram (1963)

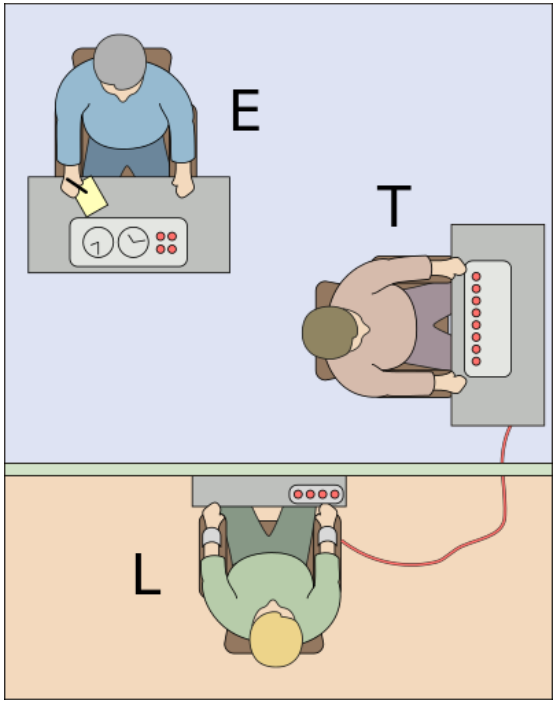
Journal of Abnormal and Social Psychology
1963, Vol. 67, No. 4, 371–378

BEHAVIORAL STUDY OF OBEDIENCE [1]

STANLEY MILGRAM [2]

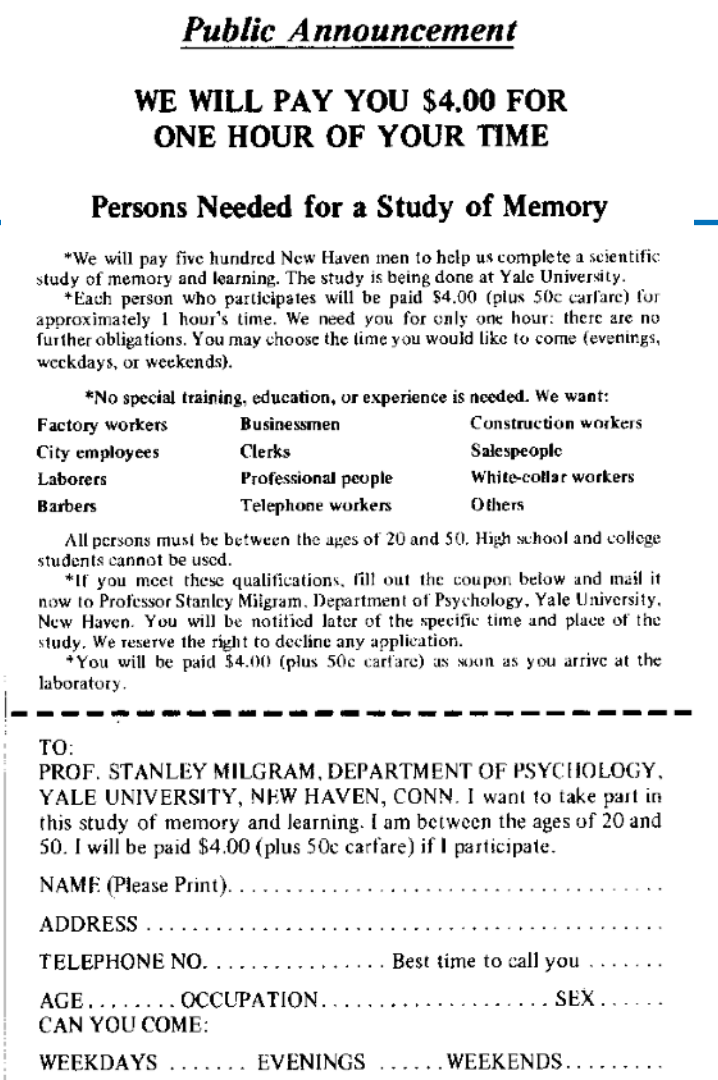
Yale University

This article describes a procedure for the study of destructive obedience in the laboratory. It consists of ordering a naive S to administer increasingly more severe punishment to a victim in the context of a learning experiment. Punishment is administered by means of a shock generator with 30 graded switches ranging from Slight Shock to Danger: Severe Shock. The victim is a confederate of the E. The primary dependent variable is the maximum shock the S is willing to administer before he refuses to continue further. 26 Ss obeyed the experimental commands fully, and administered the highest shock on the generator. 14 Ss broke off the experiment at some point after the victim protested and refused to provide further answers. The procedure created extreme levels of nervous tension in some Ss. Profuse sweating, trembling, and stuttering were typical expressions of this emotional disturbance. One unexpected sign of tension—yet to be explained—was the regular occurrence of nervous laughter, which in some Ss developed into uncontrollable seizures. The variety of interesting behavioral dynamics observed in the experiment, the reality of the situation for the S, and the possibility of parametric variation within the framework of the procedure, point to the fruitfulness of further study.



Research Ethics — obedience experiment by Milgram (1963)

- Ethical issues arise in psychological research (human beings and non-human animals)
- Weighing risks against benefits (is it worth it?):
  - what is “lost” if a study is not performed
  - what is the risk if a study is performed
- Respecting autonomy — the right to make own decisions free from coercion
  - **informed consent (Einwilligungserklärung)** — agreement to participate in a study after having being informed about everything that might reasonably be expected to affect the decision to participate
  - the right to withdraw at any time



Public Announcement

WE WILL PAY YOU $4.00 FOR ONE HOUR OF YOUR TIME

Persons Needed for a Study of Memory

*We will pay five hundred New Haven men to help us complete a scientific study of memory and learning. The study is being done at Yale University.

*Each person who participates will be paid $4.00 (plus 50c carfare) for approximately 1 hour's time. We need you for only one hour: there are no further obligations. You may choose the time you would like to come (evenings, weekdays, or weekends).

*No special training, education, or experience is needed. We want:

Factory workers	Businessmen	Construction workers
City employees	Clerks	Salespeople
Laborers	Professional people	White-collar workers
Barbers	Telephone workers	Others

All persons must be between the ages of 20 and 50. High school and college students cannot be used.

*If you meet these qualifications, fill out the coupon below and mail it now to Professor Stanley Milgram, Department of Psychology, Yale University, New Haven. You will be notified later of the specific time and place of the study. We reserve the right to decline any application.

*You will be paid $4.00 (plus 50c carfare) as soon as you arrive at the laboratory.

TO:
PROF. STANLEY MILGRAM, DEPARTMENT OF PSYCHOLOGY, YALE UNIVERSITY, NEW HAVEN, CONN. I want to take part in this study of memory and learning. I am between the ages of 20 and 50. I will be paid $4.00 (plus 50c carfare) if I participate.

NAME (Please Print). . . . . .

ADDRESS . . . . . .

TELEPHONE NO. . . . . Best time to call you . . . . . .

AGE . . . . . OCCUPATION . . . . . SEX . . . . . .
CAN YOU COME:

WEEKDAYS . . . . . EVENINGS . . . . . WEEKENDS . . . . . .

Research ethics

- Respecting privacy (the right to decide what information is shared with others)
  - Confidentiality and anonymity
- Deception (only if justified)
  - e.g., misinforming about the purpose of a study, presenting false feedback about performance
  - why? knowing about the true purpose of the study may affect the behaviour under investigation
- Debriefing (minimising any harm that might have occurred)
  - informing participants as soon as possible of the actual purpose of the study, revealing any deception, and correcting any other misconceptions they might have as a result of participating.
- local ethics committee (https://www.dgps.de/die-dgps/kommissionen/)
- APA ethics code (https://www.apa.org/ethics/code)
- Declaration of Helsinki

Schlüsselwörter

- die Bayesianische Erkenntnistheorie / bayesian epistemology (core idea)
- Operationale Definition (Operationalisierung) / operational definition (operationalisation)
- Abhängige Variable, Unabhängige Variable, Störvariable / dependent variables, independent variables, confounding variables
- Population, Stichprobe / population, sample
- Reliabilität, Validität (Messungen) / reliability, validity (measurements)
- Forschungsdesign: / study designs:
    - korrelative Studien / correlational studies
    - experimentelle Studie / experimental studies
    - randomisierte Experimente / randomised (laboratory) experiments
    - Quasi-Experimente / quasi-experimental studies
    - Zwischensubjekt-design / between-subjects design
    - Innersubjekt-design / within-subjects design
    - counterbalancing
    - Randomisierung / randomization
- Metaanalyse / Meta-analysis
- Einwilligungserklärung / informed consent
- Täuschung / deception
- Aufklärung / debriefing

Ergänzende Literatur / Quellenmaterial

- Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review*, *97*(3), 303–352.
- Schindler, S. (2008). How to Discern a Physical Effect from Background Noise: The Discovery of Weak Neutral Currents.
- Farrell, S., & Lewandowsky, S. (2018). *Computational Modeling of Cognition and Behavior*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781316272503
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course.* Cambridge University Press. https://doi.org/10.1017/CBO9781139087759
- Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian Benefits for the Pragmatic Researcher. Current Directions in Psychological Science, 25(3), 169–176. https://doi.org/10.1177/0963721416643289
- Kruschke, J. (2015). Doing Bayesian Data Analysis (Second Edition). Boston: Academic Press.
- https://plato.stanford.edu/entries/epistemology-bayesian/
- Poldrack. R. A. (2019). Statistical Thinking for the 21st Century. https://statsthinking21.org/
- Smith, P.L., Little, D.R. (2018). Small is beautiful: In defense of the small-*N* design. *Psychon Bull Rev* **25**, 2083–2101.
- https://www.youtube.com/watch?v=KqCKZwh5WY8 (Conditional probability)
- https://www.youtube.com/watch?v=ynoBpUJq72M (Kane B “Data / Phenomena)