



Instructor Guide to Heuristics and Belief Formation

Laboratory implementation, theoretical foundations,
experimental evidence, and variants

Behavioral and Experimental Economics Research Lab
Department of Economics



Revised edition | October 2026

<https://beer-lab.org/>

How to use this guide

This guide supports the selection, implementation, and discussion of activities on heuristics and belief formation.

The basic edition separates representativeness, availability, and anchoring; the core combines all three.

The extended edition provides stand-alone modules on small samples, gambler's fallacy, Bayesian updating, conjunction, confirmation, confidence, and hindsight.

Author

This guide was prepared by Stanislaw Maldonado, research professor in the Department of Economics at Tecnológico de Monterrey, Monterrey Campus. For questions or suggestions about its content, please contact the author at stanislao.maldonado@tec.mx.

For the instructor

All activities are individual. Do not reveal normative answers or other treatments before completion. A heuristic does not imply general irrationality: it may reduce cognitive cost and work well in some environments while creating predictable errors in others.

Contents

How to use this guide	i
1 Quick reference	1
1.1 Access through the updated laboratory website	1
1.2 Learning objectives	5
2 Theoretical foundation	7
2.1 Heuristics under uncertainty	7
2.2 Representativeness and base rates	7
2.3 Availability and anchoring	7
2.4 Samples and sequences	7
2.5 Bayes and information format	7
2.6 Conjunction and confirmation	8
2.7 Calibration and hindsight	8
3 Access and preparation	9
4 Participant sequence	11
5 Instructor report	25
5.1 Basic edition	32
5.2 Representativeness	32
5.3 Availability	32
5.4 Anchoring	32
5.5 Core	32
5.6 Extended edition	32
5.7 Small samples	32
5.8 Gambler's fallacy	32
5.9 Bayesian updating	33
5.10 Conjunction and confirmation	33
5.11 Calibration	33
5.12 Hindsight	33
5.13 Complete diagnostic	33
5.14 Configuration catalog	34
5.15 Data, limitations, and troubleshooting	34
6 Expanded teaching manual for nonspecialist instructors	35
6.1 Conceptual map before opening the session	35
6.1.1 Unit of analysis	35
6.1.2 Benchmark prediction	35
6.1.3 Alternative mechanisms	36
6.2 Worked example and explanation	36

6.3	Reading the instructor report	36
6.4	Step-by-step debrief script	37
6.5	Comparable class plans	37
6.5.1	A 50-minute class	37
6.5.2	An 80-minute class	38
6.5.3	A 110-minute class	38
6.6	Exercises with answer guidance	38
6.6.1	1. Reconstructing the prediction	38
6.6.2	2. Auditing a percentage	38
6.6.3	3. Reading heterogeneity	38
6.6.4	4. Separating explanations	38
6.6.5	5. Interpreting no difference	39
6.6.6	6. Transferring carefully	39
6.7	Data, inference, and learning assessment	39
6.7.1	A minimum results table	39
6.7.2	Comparing without overstatement	39
6.7.3	Assessing understanding	39
6.7.4	Closing protocol	40
6.8	Operational glossary and contingencies	40
6.9	Frequently asked questions and model answers	40
6.10	Diagnostic cases for interpretation practice	41
6.10.1	Agreement with the benchmark	41
6.10.2	Two groups with opposite behavior	41
6.10.3	A large difference with few units	42
6.10.4	A null difference	42
6.11	Explanatory sequence for a complete class	42
6.11.1	Step 1: pose the problem without labels	42
6.11.2	Step 2: reconstruct rules and incentives	42
6.11.3	Step 3: derive the benchmark	42
6.11.4	Step 4: record predictions	43
6.11.5	Step 5: reveal in layers	43
6.11.6	Step 6: work through one sequence	43
6.11.7	Step 7: compare institutions	43
6.11.8	Step 8: transfer and close	43
6.11.9	A short script to read aloud	43
6.12	Assessment bank with solutions	43
6.13	Advanced seminar: from evidence to mechanism	44
6.14	Quality control before teaching	45
6.14.1	Full rehearsal	45
6.14.2	Fifteen control questions	45
6.14.3	A micro-rehearsal of the debrief	46
6.14.4	After class	46

List of Figures

1.1	Current institutional website. Select CLASSROOM in the top navigation menu to open the teaching experiments. Captured on October 7, 2026.	1
1.2	Current catalog: search by title, topic, or JEL code. Then locate this family's card and select language and edition. Captured on October 7, 2026.	2
1.3	Demo contains only the configurations of this game family, in both languages and editions.	3
1.4	Rooms contains the four rooms for this family: basic and extended editions in Spanish and English.	4
1.5	Session-creation form for the basic English room. Check the selected configuration and participant requirement before creating the session.	4
3.1	Public entry page for the extended room.	9
3.2	Links for a six-participant session.	10
4.1	Institutional welcome and laboratory rules.	11
4.2	Current instructions for the basic edition: initial resources, decision rules, and experimental units.	12
4.3	Current comprehension check with an incorrect answer, its specific explanation, and retained responses. Continuation 1/2.	13
4.5	Base rate with or without a nondiagnostic profile.	14
4.6	Statistics with or without an already-counted vivid case.	15
4.7	Arbitrary anchor and incentivized estimate.	16
4.8	Variability in hospitals of different sizes.	17
4.9	Prediction after a mixed sequence or streak.	18
4.10	The same problem in probabilities or natural frequencies.	19
4.11	Single event versus conjunction.	20
4.12	Confirming cases versus informative comparison.	21
4.13	Choice and confidence from 50 to 100 percent.	22
4.14	Retrospective probability with outcome unknown or known.	23
4.15	Explicit closing and accuracy balance when applicable.	24
5.1	Locked report with pending responses.	25
5.2	Report ready for controlled disclosure.	25
5.3	Aggregate report after disclosure. Continuation 1/3.	26
5.6	Estimates, choices, and accuracy by treatment.	29
5.7	Calibration, histograms, and arm contrasts.	30
5.8	Contrasts and ECU-balance distribution.	31

Quick reference

1.1 Access through the updated laboratory website

Start at <https://beer-lab.org/en/>. Select **CLASSROOM** in the top navigation menu; it opens <https://teaching.beer-lab.org/en/>. In the **Classroom experiments** catalog, use the search box to locate the activity by name. You can also filter by **Topic** or **JEL code**. Clear any filters left from an earlier search before looking for another game.

Each family's card distinguishes **Español** and **English** access links. The **Show extended version** checkbox selects that edition before following a link. Catalog language and session language are separate choices: verify both. Students use the room specified by the instructor; the instructor prepares and monitors the session through this family's own administrative pages.



Figure 1.1: Current institutional website. Select CLASSROOM in the top navigation menu to open the teaching experiments. Captured on October 7, 2026.



Figure 1.2: Current catalog: search by title, topic, or JEL code. Then locate this family's card and select language and edition. Captured on October 7, 2026.

Checking resources, comprehension, and participant requirements

Before opening the room, check the selected configuration and the number of students present. The session-creation form states the group size and minimum required by that configuration. An individual activity accepts any positive participant count; an interactive activity must satisfy its stated requirement: complete groups when size is fixed, or a minimum participant count when size is flexible. Requirements may change across treatments, for example when adding a third party or changing an institution's size. If you edit a grouping parameter, check the requirement for the edited configuration, rather than relying only on the demo's suggested count.

In the instructions, ensure that students can identify what each role receives, what they can decide, and how the outcome is calculated. An initial account, a distributable budget, and final earnings need not be the same amount. Refer to the resources panel and the example for the selected treatment. Do not carry the basic edition's endowment into a treatment that changes accounts or introduces additional positions.

ECU stands for *Experimental Currency Units*: the activity's unit of account. An outcome of 100 ECU does not automatically equal 100 pesos or 100 dollars. Conversion depends on the session configuration. Before announcing a payment, confirm with the laboratory whether money is paid, the conversion rate, and which decisions count for payment. In hypothetical modules, describe amounts as outcomes in the stated situation.

When a comprehension check is included, an incorrect response receives an explanation beside that question, identifying the submitted answer and the correct answer. Other answers remain in the form. Ask the student to return to the example, locate the relevant rule, and redo the calculation before correcting the response. Avoid revealing answers to the entire class for an individual error, and do not interpret a comprehension mistake as an economic choice in the game.

creencias — representatividad	Heurísticas y formación de creencias / Heuristics and Belief Formation — BEER Lab Seleccione una variante de este juego. Select a version of this game.
creencias — disponibilidad	
creencias — anclaje	
creencias — núcleo	
nd Beliefs — representativeness	
nd Beliefs — availability	
nd Beliefs — anchoring	
nd Beliefs — core	
creencias extendidas — muestras pequeñas	
creencias extendidas — falacia del jugador	
creencias extendidas — actualización bayesiana	
creencias extendidas — conjunción	
creencias extendidas — sesgo de confirmación	
creencias extendidas — calibración	
creencias extendidas — sesgo retrospectivo	
creencias extendidas — diagnóstico completo	
euristics and Beliefs — small samples	
euristics and Beliefs — gambler's fallacy	
euristics and Beliefs — Bayesian updating	
euristics and Beliefs — conjunction	
euristics and Beliefs — confirmation bias	
euristics and Beliefs — calibration	
euristics and Beliefs — hindsight bias	
euristics and Beliefs — complete diagnostic	

Figure 1.3: Demo contains only the configurations of this game family, in both languages and editions.

Rooms

Current rooms:

[ES] Heurísticas y formación de creencias
[EN] Heuristics and Belief Formation
[ES] Heurísticas y creencias: extendido
[EN] Heuristics and Beliefs: extended

Figure 1.4: Rooms contains the four rooms for this family: basic and extended editions in Spanish and English.

Room: [EN] Heuristics and Belief Formation

Create a new session

Session Config

[EN] Heuristics and Beliefs — representativeness ▾

Number of participants

Individual decisions: any positive participant count is allowed.

Create

[Configure session](#) ⚙️

App sequence

heuristics_beliefs_game (10 rounds)

Individual classroom activities on representativeness, availability, anchoring, small samples, gambler's fallacy, Bayesian updating, conjunction, confirmation, confidence calibration, and hindsight bias.

Figure 1.5: Session-creation form for the basic English room. Check the selected configuration and participant requirement before creating the session.

Downloading a session with several rounds

For the instructor

Open **Data** within the relevant session and use that session's CSV or CSV-for-Excel download link. The general export has one row per participant and columns for the different rounds. Ten rounds do not necessarily mean ten rows per person. When the family provides its own long-format export, observations may instead be organized by participant and round or another unit defined by the design.

Save the original file together with the configuration name, session code, and active-round count. Check that columns for the last active round are present and that the data belong to the selected session. Empty cells can represent inactive rounds, a pending decision, or a field that does not apply to a role; do not automatically recode them as zeros. To compare learning, preserve round and group identifiers alongside participant identifiers. Share only anonymous aggregates in class, and record session incidents before constructing derived tables.

Purpose	Observe how salient information, descriptions, anchors, sequences, and formats alter probability judgments.
Participants	Any number; six or more provide balanced arms and useful denominators.
Basic edition	Representativeness, availability, anchoring, or core.
Extended edition	Small samples, gambler's fallacy, Bayes, conjunction, confirmation, calibration, hindsight, or complete diagnostic.
Interaction	Individual; balanced assignment to two treatments.
Payoff	Only anchoring generates ECU: $100 - 2 estimate - 28 $, bounded below at zero.
Students	https://heuristics.beer-lab.org/room/heuristics_beliefs_en
Instructor	https://heuristics.beer-lab.org/demo
Local test	http://localhost:8018/demo with GAME_FAMILY=heuristics.
Experimental unit	Individual decisions; each person completes their own tasks.
Decision	Answer probability judgments, estimates, and vignettes in the selected module.
Topics	heuristics; belief formation; judgment biases; Bayesian updating; behavioral economics.
JEL	C91; D83; D91.

Attention

The heuristics.beer-lab.org domain is a proposed route. It requires DNS, an isolated process, and a separate PostgreSQL database. The local version was verified on port 8018.

1.2 Learning objectives

- distinguish a heuristic from the relevant normative rule;
- explain representativeness, availability, and insufficient adjustment from an anchor;

- apply base rates, sampling variability, independence, and the conjunction rule;
- compute a Bayesian posterior and compare probability with natural-frequency formats;
- identify confirmatory search and design an informative test;
- compare confidence with accuracy and recognize overconfidence or underconfidence;
- explain why outcome knowledge changes apparent predictability;
- interpret distributions and denominators without diagnosing individuals.

Theoretical foundation

2.1 Heuristics under uncertainty

[Tversky and Kahneman \(1974\)](#) organize three major families: representativeness, availability, and anchoring. They economize on attention and calculation but can separate intuition from probability rules. These classroom tasks are pedagogical adaptations rather than exact replications.

2.2 Representativeness and base rates

Representativeness asks how much a case resembles a category. [Kahneman and Tversky \(1973\)](#) show that descriptions can displace base rates even with little predictive value. Our profile is explicitly equally frequent across occupations, so the normative answer remains 30%.

2.3 Availability and anchoring

Availability uses ease of retrieval as a cue to frequency ([Tversky and Kahneman, 1973](#)). The module holds 20 delays in 1,000 fixed and makes one already-counted case vivid. Anchoring displays an arbitrary 10 or 65 before the same estimate. Payment rewards proximity to 28%; the anchor is not informative.

2.4 Samples and sequences

The law of small numbers expects small samples to resemble the population too closely ([Tversky and Kahneman, 1971](#)). Daily proportions are more variable at the small hospital, so it crosses 55% and 60% more often. Gambler's fallacy treats an independent streak as needing reversal. A fair coin remains 50% after every sequence; economic models connect this belief to mistaken sampling intuitions ([Rabin, 2002](#)).

2.5 Bayes and information format

With 1% prevalence, 80% sensitivity, and 10% false positives, $P(C \mid +) = 0.008 / (0.008 + 0.099) \approx 7.5\%$. [Gigerenzer and Hoffrage \(1995\)](#) show that natural frequencies can facilitate Bayesian reasoning. Both arms contain identical information.

2.6 Conjunction and confirmation

Extension requires $P(A \cap B) \leq P(A)$. A representative description can make the conjunction feel more plausible ([Tversky and Kahneman, 1983](#)). The confirmation task compares reviewing successes alone with examining favorable and unfavorable evidence, adapting the hypothesis-testing tradition of [Wason \(1960\)](#).

2.7 Calibration and hindsight

Confidence is calibrated when judgments reported at p percent are correct roughly p percent of the time. [Moore and Healy \(2008\)](#) distinguish overestimation, overplacement, and overprecision; this task measures only confidence versus aggregate accuracy. One item per person supports group-level interpretation only. Outcome knowledge can raise retrospective probability and apparent foreseeability ([Fischhoff, 1975](#)); the hindsight module holds prior signals constant and reveals success to only one arm.

Theoretical foundation

Do not label every difference a bias. Identify the normative rule, what information remains fixed, what changes, and whether the treatment truly changes information content.

Access and preparation

Start at <https://beer-lab.org/>, open <https://teaching.beer-lab.org/>, and select the family after deployment.

Basic Spanish: https://heuristics.beer-lab.org/room/heurísticas_creencias_es

Basic English: https://heuristics.beer-lab.org/room/heuristics_beliefs_en

Extended Spanish: https://heuristics.beer-lab.org/room/heurísticas_creencias_extendido

Extended English: https://heuristics.beer-lab.org/room/heuristics_beliefs_extended

Instructors: <https://heuristics.beer-lab.org/demo>

The isolated instance lists 24 configurations and four rooms.

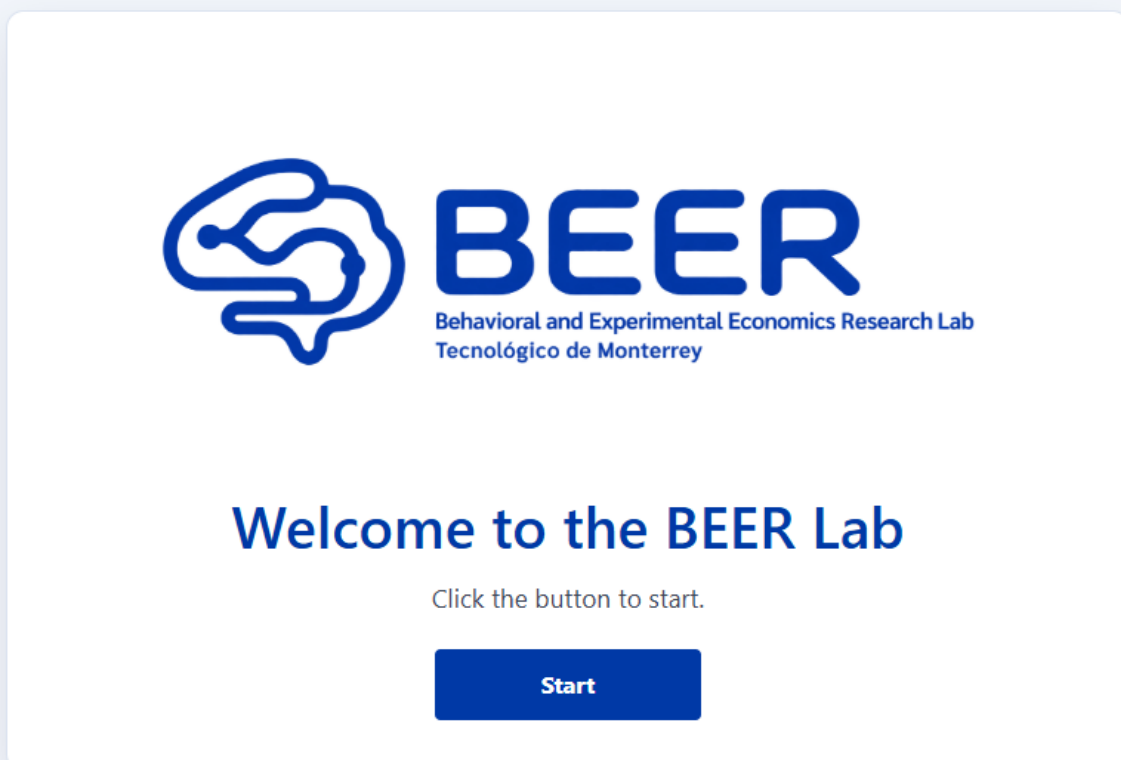


Figure 3.1: Public entry page for the extended room.

oTree

DemoSessionsRoomsDataServer Check

[EN] Extended Heuristics and Beliefs — complete diagnostic: session 'jx1st4jb' (demo)

New

Links

Monitor

Data

Payments

Report

Description

Below are temporary links for testing and demonstration. To launch a real study, either create persistent links by setting up a [room](#), or create a session through the [sessions page](#).

Choose one of the following options to play the session:

Combined views

Split-screen view

Launch

Grid view

Launch

Session-wide testing link

Open the below link in up to 6 browser tabs.

<http://127.0.0.1:8018/join/bohufuvi>

Single-use links

Open each link in its own browser tab.

P1	http://127.0.0.1:8018/initializeParticipant/faf7lamz
P2	http://127.0.0.1:8018/initializeParticipant/t526hbm5
P3	http://127.0.0.1:8018/initializeParticipant/t8tvaxco
P4	http://127.0.0.1:8018/initializeParticipant/bf591vtj
P5	http://127.0.0.1:8018/initializeParticipant/fb942pos
P6	http://127.0.0.1:8018/initializeParticipant/6dqpetou

Figure 3.2: Links for a six-participant session.

Participant sequence



Welcome

The **Behavioral and Experimental Economics Research Lab** of the Department of Economics at Tecnológico de Monterrey is a space designed to learn about economic decision making through controlled activities.

In this activity, you will examine how people form beliefs and make judgments under uncertainty. Read every situation carefully and respond independently.

General laboratory rules

- Remain quiet throughout the session.
- Do not talk or communicate with other participants.
- Make your decisions individually and privately.
- Follow the instructions shown on screen.
- Do not refresh, close, or leave this page.
- If you have a question or technical problem, raise your hand and wait for assistance.

☐ I have read these rules and agree to follow them.

Start

Figure 4.1: Institutional welcome and laboratory rules.

The interface separates highlighted information in blue from normative data in amber.



Instructions

1 decision(s)

Resources and unit of account

ECU stands for Experimental Currency Units. It is the unit of account for this activity.

This session does not convert ECU into real money.

Basic version · Representativeness and base rates

Each situation asks for a probability, estimate, or choice. Some versions present information differently; use only the information shown on your screen.

1

Situations and treatments

The system may assign different versions of the same situation. Do not infer or compare other participants' treatments.

2

Private judgments

Respond according to your own judgment. The activity is individual and does not require waiting for anyone else.

3

Decisions and payments

Most situations are hypothetical. Only an estimate marked in green can generate an ECU balance.

Continue

Con tecnología de [oTree](#)

Figure 4.2: Current instructions for the basic edition: initial resources, decision rules, and experimental units.



Review the marked questions. Each shows your answer and an explanation; your other answers are retained.

Comprehension questions

Resources and unit of account

ECU stands for Experimental Currency Units. It is the unit of account for this activity.
This session does not convert ECU into real money.

Representativeness and base rates

Answer correctly before proceeding to the decisions.

How should you respond?

- ☐ Individually and privately
- ☒ By comparing answers with other participants

Your answer: By comparing answers with other participants. The correct answer is Individually and privately.
Review the instructions: answer privately, use your own judgment, and remember that only the green decision can generate ECU.

Which decision can generate an ECU balance?

- ☒ Only the one marked in green
- ☐ All situations

Which probability should you report?

- ☒ The one that reflects my own judgment using the information shown
- ☐ The answer I think most people will choose

Check and continue

Con tecnología de [oTree](#)

Figure 4.4: Current comprehension check with an incorrect answer, its specific explanation, and retained responses. Continuation 2/2.

Representativeness and base rate

Decision 1 of 10

Representativeness

Representativeness

One person was randomly selected from a group containing 30 engineers and 70 lawyers. The person is described as methodical and fond of solving technical problems. A prior study found this description equally frequent among engineers and lawyers.

HIGHLIGHTED INFORMATION

Salient profile

Methodical and technical

SITUATION DATA

Base rate

30% engineers

What is the probability that the person is an engineer?

This situation is hypothetical. Respond as you would if the judgment were real.

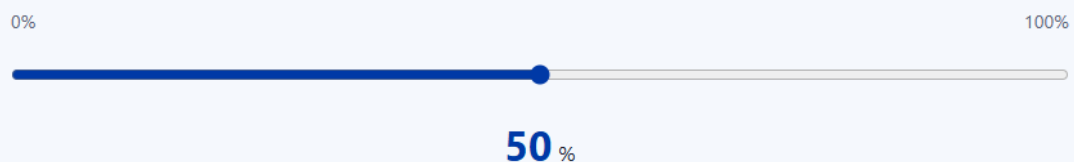
**Confirm response**Con tecnología de [oTree](#)

Figure 4.5: Base rate with or without a nondiagnostic profile.

Availability and frequency

Decision 2 of 10

Availability

Availability

A company recorded 20 delayed deliveries among 1,000 comparable deliveries. One of those 20 was a five-day delay that stopped a plant; that case is already included in the total.

HIGHLIGHTED INFORMATION

Vivid case

Five-day delay**SITUATION DATA**

Statistical record

20 of 1,000

What probability do you assign to a delay?

This situation is hypothetical. Respond as you would if the judgment were real.

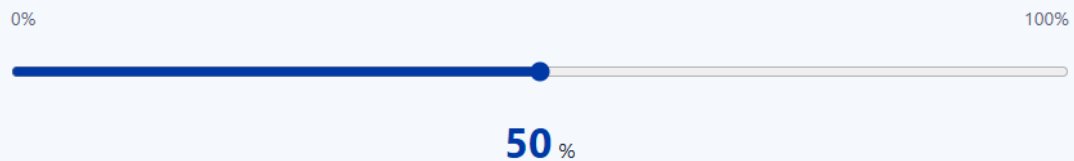
**Confirm response**Con tecnología de [oTree](#)

Figure 4.6: Statistics with or without an already-counted vivid case.

Numerical anchoring

Decision 3 of 10

Anchoring and adjustment

Anchoring and adjustment

A wheel producing arbitrary numbers displayed 10. Estimate what percentage of United Nations member states are African countries.

HIGHLIGHTED INFORMATION

Arbitrary number

10

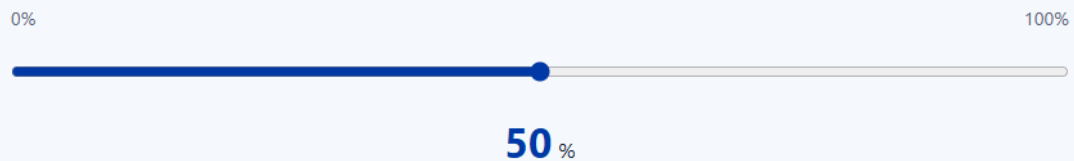
SITUATION DATA

Course reference answer

28%

What is your estimate?

This estimate is marked in green. Its accuracy can determine a balance from 0 to 100 ECU.



Confirm response

Con tecnología de [oTree](#)

Figure 4.7: Arbitrary anchor and incentivized estimate.

Variability in small samples

Decision 4 of 10

Law of small numbers

Law of small numbers

Two hospitals record births for a year. The smaller hospital has 15 births per day and the larger has 45. On average, about 50% of babies are boys.

HIGHLIGHTED INFORMATION

Daily threshold

More than 55% boys**SITUATION DATA**

Sample size

15 versus 45 births

Which hospital will have more days above the threshold?

This situation is hypothetical. Respond as you would if the judgment were real.

Option A**Smaller hospital****Option B****Larger hospital****Option C****About the same****Confirm response**Con tecnología de [oTree](#)

Figure 4.8: Variability in hospitals of different sizes.

Prediction after a sequence

Decision 5 of 10

Gambler's fallacy

Gambler's fallacy

A fair coin produced heads, heads, heads, heads, heads. It will be tossed once more.

HIGHLIGHTED INFORMATION

Recent sequence

H-H-H-H-H**SITUATION DATA**

Process

Fair coin

Which outcome do you consider more likely?

This situation is hypothetical. Respond as you would if the judgment were real.

Option A**Heads**

- Objective probability: 50%.

Option B**Tails**

- Objective probability: 50%.

Confirm responseCon tecnología de [oTree](#)

Figure 4.9: Prediction after a mixed sequence or streak.

Updating after a test

Decision 6 of 10

Bayesian reasoning

Bayesian reasoning

A condition affects 1% of the population. A test is positive for 80% of people with the condition and 10% of people without it. A randomly selected person tested positive.

HIGHLIGHTED INFORMATION

Format

Probabilities

SITUATION DATA

Data

1%, 80%, and 10%

What is the probability that the person has the condition?

This situation is hypothetical. Respond as you would if the judgment were real.

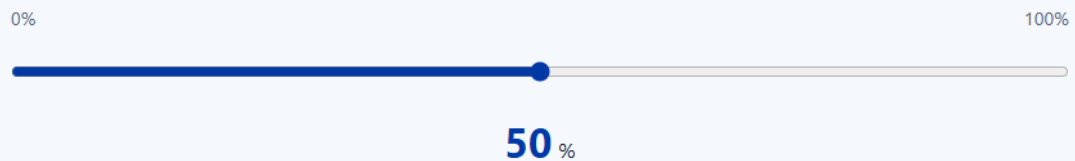
**Confirm response**Con tecnología de [oTree](#)

Figure 4.10: The same problem in probabilities or natural frequencies.

Probability of a conjunction

Decision 7 of 10

Conjunction fallacy

Conjunction fallacy

Laura is 31, studied philosophy, and was active in social-justice and environmental campaigns at university.

HIGHLIGHTED INFORMATION

Description

Representative of activism

SITUATION DATA

Probability rule

A conjunction cannot be more probable than either component

Which option is more probable?

This situation is hypothetical. Respond as you would if the judgment were real.

Option A

**Laura works at a bank.**

Option B

**Laura works at a bank and is active in an environmental organization.****Confirm response**Con tecnología de [oTree](#)

Figure 4.11: Single event versus conjunction.

How to test a hypothesis

Decision 8 of 10

Confirmation bias

Confirmation bias

A company believes its voluntary training program raises productivity. The stated objective is to find evidence confirming that idea.

HIGHLIGHTED INFORMATION

Objective

Confirm the hypothesis

SITUATION DATA

Resources

Only one analysis can be selected

Which analysis would be more informative?

This situation is hypothetical. Respond as you would if the judgment were real.

Option A



Review cases of workers who took the program and improved.

- Looks for cases consistent with the hypothesis.

Option B



Compare participants and nonparticipants, including improvements and declines.

- Can reveal both favorable and unfavorable evidence.

Confirm responseCon tecnología de [oTree](#)

Figure 4.12: Confirming cases versus informative comparison.

Answer and confidence

Decision 9 of 10

Calibration

Calibration

Answer the question and then state the probability that your answer is correct.

HIGHLIGHTED INFORMATION

Difficulty

Difficult question**SITUATION DATA**

Question

Geographic comparison

Which city is farther north?

This situation is hypothetical. Respond as you would if the judgment were real.

Option A**Rome****Option B****New York City**

What probability do you assign to your answer being correct?

50%

100%

**75 %****Confirm response**

Figure 4.13: Choice and confidence from 50 to 100 percent.

Retrospective evaluation

Decision 10 of 10

Hindsight bias

Hindsight bias

Before launching a new product, a team observed moderate demand, costs within budget, and a positive but small pilot. The product later succeeded.

HIGHLIGHTED INFORMATION

Later information

The product succeeded

SITUATION DATA

Prior information

The same mixed signals

Using only what was available then, what probability of success would you assign?

This situation is hypothetical. Respond as you would if the judgment were real.

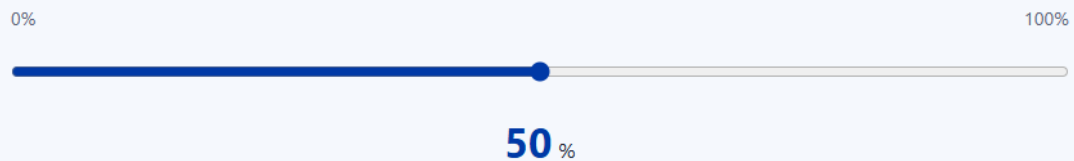
**Confirm response**Con tecnología de [oTree](#)

Figure 4.14: Retrospective probability with outcome unknown or known.



Activity completed

Your responses were recorded successfully.

ECU estimate outcome

Selected decision

Numerical anchoring

Your estimate

38%

Reference answer

28%

Balance

80 ECU

Remain quiet and follow the instructor's directions.

Con tecnología de [oTree](#)

Figure 4.15: Explicit closing and accuracy balance when applicable.

Instructor report

Aggregates stay hidden until everyone finishes.

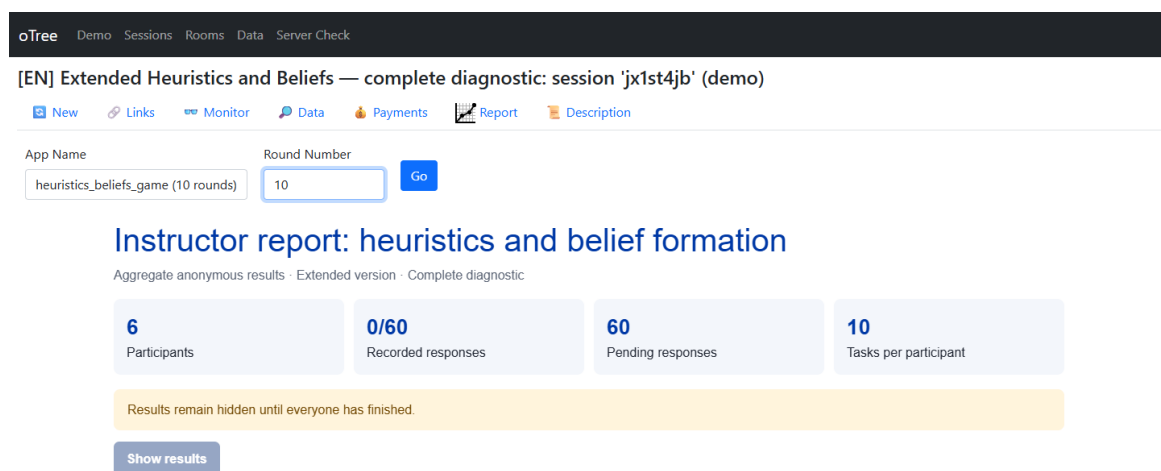


Figure 5.1: Locked report with pending responses.

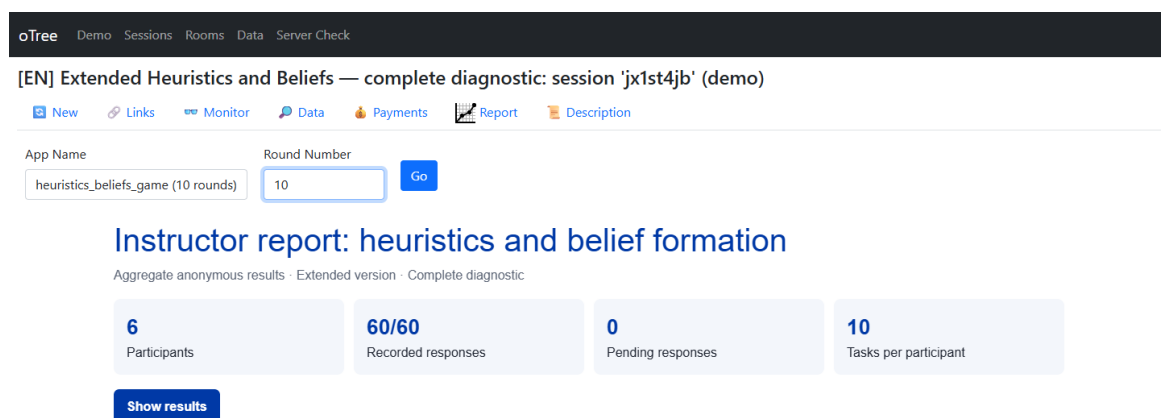


Figure 5.2: Report ready for controlled disclosure.

Feedback script

1. Write the normative rule before revealing results.
2. Elicit directional predictions, not only whether arms differ.

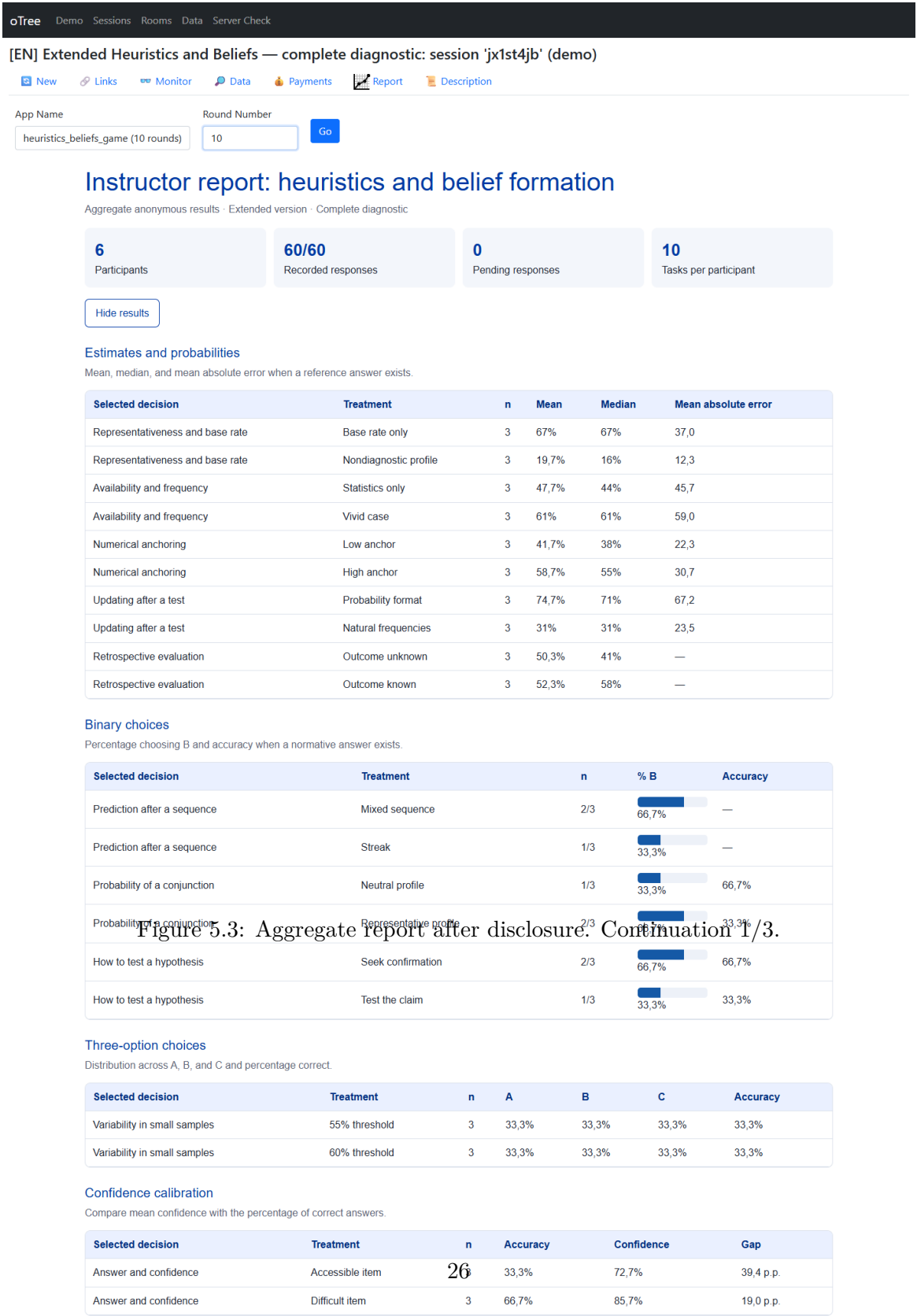


Figure 5.3: Aggregate report after disclosure. Continuation 1/3.

Availability and frequency	Statistics only	3	47,7%	44%	45,7
Availability and frequency	Vivid case	3	61%	61%	59,0
Numerical anchoring	Low anchor	3	41,7%	38%	22,3
Numerical anchoring	High anchor	3	58,7%	55%	30,7
Updating after a test	Probability format	3	74,7%	71%	67,2
Updating after a test	Natural frequencies	3	31%	31%	23,5
Retrospective evaluation	Outcome unknown	3	50,3%	41%	—
Retrospective evaluation	Outcome known	3	52,3%	58%	—

Binary choices

Percentage choosing B and accuracy when a normative answer exists.

Selected decision	Treatment	n	% B	Accuracy
Prediction after a sequence	Mixed sequence	2/3	66,7%	—
Prediction after a sequence	Streak	1/3	33,3%	—
Probability of a conjunction	Neutral profile	1/3	33,3%	66,7%
Probability of a conjunction	Representative profile	2/3	66,7%	33,3%
How to test a hypothesis	Seek confirmation	2/3	66,7%	66,7%
How to test a hypothesis	Test the claim	1/3	33,3%	33,3%

Three-option choices

Distribution across A, B, and C and percentage correct.

Selected decision	Treatment	n	A	B	C	Accuracy
Variability in small samples	55% threshold	3	33,3%	33,3%	33,3%	33,3%
Variability in small samples	60% threshold	3	33,3%	33,3%	33,3%	33,3%

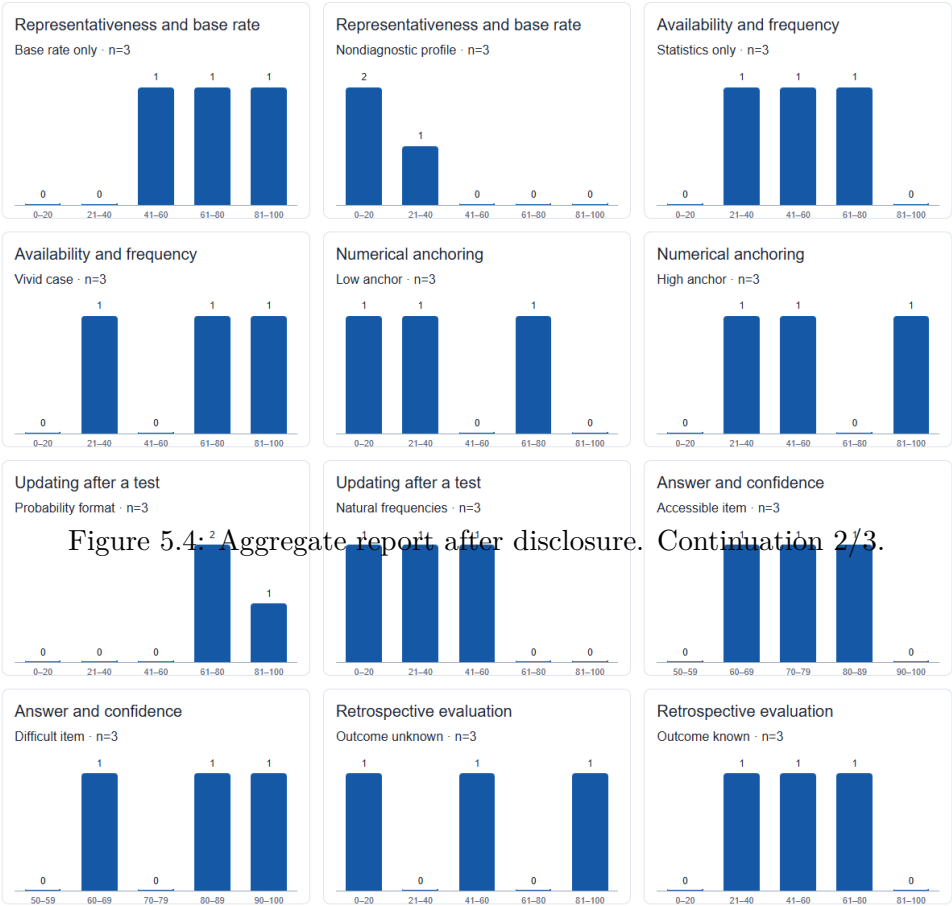
Confidence calibration

Compare mean confidence with the percentage of correct answers.

Selected decision	Treatment	n	Accuracy	Confidence	Gap
Answer and confidence	Accessible item	3	33,3%	72,7%	39,4 p.p.
Answer and confidence	Difficult item	3	66,7%	85,7%	19,0 p.p.

Distributions

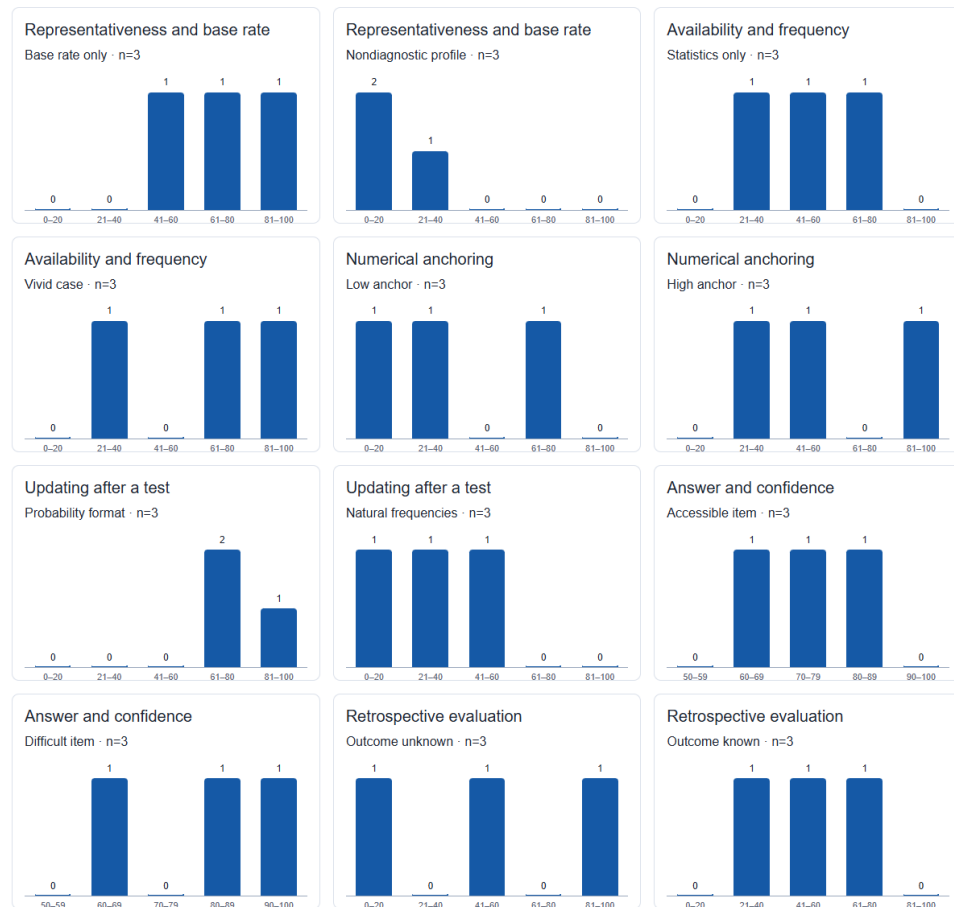
Histograms by task and treatment.



Descriptive contrasts

Differences between treatments; they do not replace statistical inference.

Selected decision	Contrast	Difference
Representativeness and base rate	Nondiagnostic profile – Base rate only	-47,3 points
Availability and frequency	Vivid case – Statistics only	13,3 points
Numerical anchoring	Low anchor – High anchor	-17,0 points



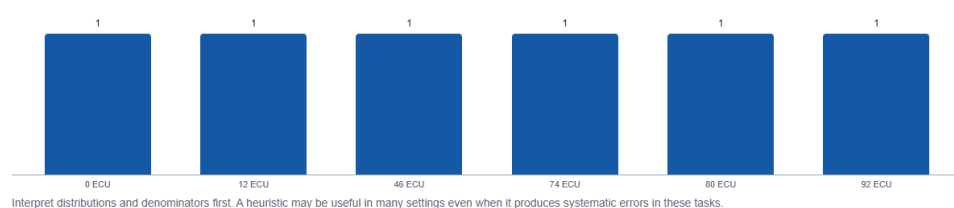
Descriptive contrasts

Differences between treatments; they do not replace statistical inference.

Selected decision	Contrast	Difference
Representativeness and base rate	Nondiagnostic profile – Base rate only	-47,3 points
Availability and frequency	Vivid case – Statistics only	13,3 points
Numerical anchoring	High anchor – Low anchor	17,0 points
Variability in small samples	60% threshold – 55% threshold	0,0 p.p.
Prediction after a sequence	Streak – Mixed sequence	-33,4 p.p.
Updating after a test	Natural frequencies – Probability format	-43,7 points
Probability of a conjunction	Representative profile – Neutral profile	33,4 p.p.
How to test a hypothesis	Test the claim – Seek confirmation	-33,4 p.p.
Answer and confidence	Difficult item – Accessible item	-20,4 p.p.
Retrospective evaluation	Outcome known – Outcome unknown	2,0 points

Distribution of ECU balances

Recorded monetary estimates: 6.



Interpret distributions and denominators first. A heuristic may be useful in many settings even when it produces systematic errors in these tasks.

Figure 5.5: Aggregate report after disclosure. Continuation 3/3.

Representativeness and base rate	Base rate only	3	67%	67%	37,0
Representativeness and base rate	Nondiagnostic profile	3	19,7%	16%	12,3
Availability and frequency	Statistics only	3	47,7%	44%	45,7
Availability and frequency	Vivid case	3	61%	61%	59,0
Numerical anchoring	Low anchor	3	41,7%	38%	22,3
Numerical anchoring	High anchor	3	58,7%	55%	30,7
Updating after a test	Probability format	3	74,7%	71%	67,2
Updating after a test	Natural frequencies	3	31%	31%	23,5
Retrospective evaluation	Outcome unknown	3	50,3%	41%	—
Retrospective evaluation	Outcome known	3	52,3%	58%	—

Binary choices

Percentage choosing B and accuracy when a normative answer exists.

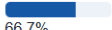
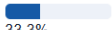
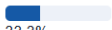
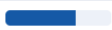
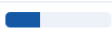
Selected decision	Treatment	n	% B	Accuracy
Prediction after a sequence	Mixed sequence	2/3	 66,7%	—
Prediction after a sequence	Streak	1/3	 33,3%	—
Probability of a conjunction	Neutral profile	1/3	 33,3%	66,7%
Probability of a conjunction	Representative profile	2/3	 66,7%	33,3%
How to test a hypothesis	Seek confirmation	2/3	 66,7%	66,7%
How to test a hypothesis	Test the claim	1/3	 33,3%	33,3%

Figure 5.6: Estimates, choices, and accuracy by treatment.

Probability of a conjunction	Neutral profile	1/3	33,3%	66,7%
Probability of a conjunction	Representative profile	2/3	66,7%	33,3%
How to test a hypothesis	Seek confirmation	2/3	66,7%	66,7%
How to test a hypothesis	Test the claim	1/3	33,3%	33,3%

Three-option choices

Distribution across A, B, and C and percentage correct.

Selected decision	Treatment	n	A	B	C	Accuracy
Variability in small samples	55% threshold	3	33,3%	33,3%	33,3%	33,3%
Variability in small samples	60% threshold	3	33,3%	33,3%	33,3%	33,3%

Confidence calibration

Compare mean confidence with the percentage of correct answers.

Selected decision	Treatment	n	Accuracy	Confidence	Gap
Answer and confidence	Accessible item	3	33,3%	72,7%	39,4 p.p.
Answer and confidence	Difficult item	3	66,7%	85,7%	19,0 p.p.

Distributions

Histograms by task and treatment.



Figure 5.7: Calibration, histograms, and arm contrasts.



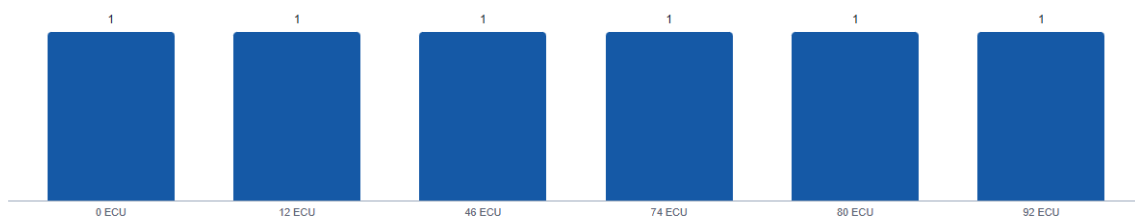
Descriptive contrasts

Differences between treatments; they do not replace statistical inference.

Selected decision	Contrast	Difference
Representativeness and base rate	Nondiagnostic profile – Base rate only	-47,3 points
Availability and frequency	Vivid case – Statistics only	13,3 points
Numerical anchoring	High anchor – Low anchor	17,0 points
Variability in small samples	60% threshold – 55% threshold	0,0 p.p.
Prediction after a sequence	Streak – Mixed sequence	-33,4 p.p.
Updating after a test	Natural frequencies – Probability format	-43,7 points
Probability of a conjunction	Representative profile – Neutral profile	33,4 p.p.
How to test a hypothesis	Test the claim – Seek confirmation	-33,4 p.p.
Answer and confidence	Difficult item – Accessible item	-20,4 p.p.
Retrospective evaluation	Outcome known – Outcome unknown	2,0 points

Distribution of ECU balances

Recorded monetary estimates: 6.



Interpret distributions and denominators first. A heuristic may be useful in many settings even when it produces systematic errors in these tasks.

Figure 5.8: Contrasts and ECU-balance distribution.

3. Reveal distributions and denominators before means.
4. For Bayes, build the 1,000-person table and derive $8/(8 + 99)$.
5. Compare mean confidence with accuracy; do not interpret one person in isolation.
6. Separate content, format, and attention errors.
7. Close by asking when a heuristic saves resources and when a statistical tool is needed.

5.1 Basic edition

5.2 Representativeness

Base rate and random selection remain fixed; the profile is explicitly nondiagnostic. An estimate above 30% under the profile is consistent with representativeness.

5.3 Availability

The total remains 20 delays in 1,000. The five-day event is one of those 20. The normative rate is 2%.

5.4 Anchoring

Anchors 10 and 65 are arbitrary. The balance peaks at estimate 28 and falls by two ECU per point of error. Do not reveal the answer before closing.

5.5 Core

Combines all three modules. Allow 10–15 minutes for responses and 20–25 minutes for discussion.

5.6 Extended edition

5.7 Small samples

Compares 55% and 60% thresholds. The smaller hospital is correct in both; the extreme threshold highlights variance.

5.8 Gambler's fallacy

Compares a mixed sequence with five heads. Neither changes the next toss's objective probability.

5.9 Bayesian updating

Compares normalized probabilities and natural frequencies with answer 7.5%. Use mean absolute error to evaluate the format effect.

5.10 Conjunction and confirmation

In conjunction, A contains B. In confirmation, comparing participants and nonparticipants is more informative in both frames.

5.11 Calibration

Contrasts an accessible and a difficult item. Confidence minus accuracy can be positive or negative and depends on difficulty.

5.12 Hindsight

Only one arm knows the later success. A higher estimate is consistent with hindsight, but the vignette does not measure memory for one's own earlier prediction.

5.13 Complete diagnostic

Contains ten decisions. Allow 25–35 minutes. Prefer a focal module for an ordinary class.

5.14 Configuration catalog

Module	Spanish / English configuration
Representativeness	heurísticas_creencias_es_representatividad / heuristics_beliefs_en_representativeness
Availability	heurísticas_creencias_es_disponibilidad / heuristics_beliefs_en_availability
Anchoring	heurísticas_creencias_es_anclaje / heuristics_beliefs_en_anchoring
Core	heurísticas_creencias_es_nucleo / heuristics_beliefs_en_core
Small samples	heurísticas_creencias_extendido_muestras_pequenas / heuristics_beliefs_extended_small_samples
Gambler's fallacy	heurísticas_creencias_extendido_falacia_jugador / heuristics_beliefs_extended_gamblers_fallacy
Bayes	heurísticas_creencias_extendido_actualizacion_bayesiana / heuristics_beliefs_extended_bayesian_updating
Conjunction	heurísticas_creencias_extendido_conjuncion / heuristics_beliefs_extended_conjunction
Confirmation	heurísticas_creencias_extendido_sesgo_confirmacion / heuristics_beliefs_extended_confirmation_bias
Calibration	heurísticas_creencias_extendido_sobreconfianza / heuristics_beliefs_extended_overconfidence
Hindsight	heurísticas_creencias_extendido_sesgo_retrospectivo / heuristics_beliefs_extended_hindsight_bias
Diagnostic	heurísticas_creencias_extendido_diagnostico_completo / heuristics_beliefs_extended_full_diagnostic

5.15 Data, limitations, and troubleshooting

The export includes task, module, treatment, estimate, choice, confidence, correctness, absolute error, monetary eligibility, and balance. Do not average modules as if they measured one trait. These are short adaptations, not psychometric instruments.

If *Please wait* appears, inspect Monitor: this family never waits for another participant. If Demo lists other games, check `GAME_FAMILY=heuristics`. If the report remains locked, confirm that everyone reached the closing page.

Expanded teaching manual for nonspecialist instructors

This part brings together what an instructor needs to turn Heuristics and beliefs into a learning experience. It complements the operational material above. It can be read before the first session and reused as a script during the discussion of results. The central question is how judgments are formed under uncertainty and when useful shortcuts diverge from probability rules. The guide always separates four levels: rules, benchmark prediction, observed behavior, and interpretation. Moving directly from a bar in a chart to a psychological explanation confuses those levels.

6.1 Conceptual map before opening the session

6.1.1 Unit of analysis

The relevant unit is each participant-item observation, retaining treatment and a normative answer when one exists. One data row is not always one independent person: the same person may decide repeatedly, and interacting participants share information, payoffs, or history. The instructor should therefore state what enters the numerator, what enters the denominator, and whether observations are people, decisions, pairs, groups, or rounds. This prevents thirty repeated decisions from being presented as thirty independent participants.

The task is to estimate probabilities or quantities, choose hypotheses, and report confidence. Before presenting theory, ask students to restate the rule and identify what information is available when they decide. A correct calculation does not guarantee understanding of the strategic or behavioral structure. Check available actions, information, payoffs, order, and feedback timing separately.

6.1.2 Benchmark prediction

The starting point is: use base rates, Bayes' rule, set inclusion, independence, and sampling properties as the problem requires. The benchmark is a disciplined comparison, not a moral statement about what a person ought to do. First write the assumptions that support it, then derive the action or pattern, and only then compare it with the class. If information, repetition, incentives, matching, or the reference point changes, derive the prediction again instead of carrying it over automatically.

Keep equilibrium, efficiency, and fairness separate. An action can be a best response and generate a poor joint outcome; another action can raise surplus and remain vulnerable to unilateral deviation. Likewise, an equal outcome need not be efficient, and an efficient distribution does not reveal the process that produced it. Asking the class to label each statement as positive, normative, or

strategic reduces confusion.

6.1.3 Alternative mechanisms

The literature and treatments permit discussion of representativeness, availability, anchoring, the law of small numbers, gambler's fallacy, conjunction, confirmation, overprecision, and hindsight. These mechanisms are not interchangeable labels. Two can predict the same baseline action and separate only when a treatment changes. The informative comparison therefore asks not only how much behavior occurred but also what experimental change should increase, reduce, or leave the pattern unchanged if one mechanism matters.

Organize discussion in three columns. Write the observed fact in the first, such as a share, median, or trajectory. List compatible mechanisms in the second. Record the additional evidence needed to distinguish them in the third. The third column is essential because a classroom session rarely identifies a stable preference or causal relationship by itself.

6.2 Worked example and explanation

With one-percent prevalence, 80-percent sensitivity, and a 10-percent false-positive rate, imagine 10,000 people. There are 100 cases, 80 of whom test positive; among 9,900 noncases, 990 test positive. The posterior probability is 80 divided by 1,070, about 7.5 percent. Natural frequencies make the often-missed denominator visible.

Solve the example in six steps. First list actions and possible states. Second calculate payoffs or consequences without interpreting motives. Third derive the theoretical benchmark under its assumptions. Fourth identify the minimum observation needed to evaluate that benchmark. Fifth compare treatments while stating what is held constant and what changes. Sixth formulate a limited conclusion: "the pattern is consistent with" is more accurate than "the experiment proves" when rival explanations remain.

Then change one parameter and ask for a new prediction before calculating. This pause turns the example into a reasoning tool. It also reveals whether students memorized an answer or understood the mechanism. If answers diverge, ask each group to identify the assumption producing the difference.

6.3 Reading the instructor report

The central measures are absolute error from the reference; treatment difference; normative-response rate; confidence-accuracy gap; consistency across items. Write the denominator next to every indicator. A rate can divide by eligible participants, completed decisions, formed pairs, groups, or rounds; these alternatives are not equivalent. If records are incomplete, show how many were excluded and why.

Use this reading order:

1. Verify participants, groups, and completed decisions. Do not interpret percentages before checking coverage.
2. Inspect the full distribution. A mean can hide two groups with opposing responses or a concentration at focal values.
3. Compare the median, range, and theoretically relevant categories. Explain whether the graph

counts people or decisions.

4. Inspect time paths only when rounds are comparable. Changes in matching or treatment must be marked.
5. Compare treatments with absolute differences and denominators, not merely with bars of different heights.
6. Select one anonymous case or sequence connecting rules and the aggregate. Never use it to identify or ridicule a student.

A visible difference in a small class is descriptive evidence from that session. Do not call it a causal effect without appropriate assignment, adequate size, and a comparison defined before seeing the data. Across classes, composition, time, and prior experience may change together. The pedagogical value lies in stating a prediction, confronting it with evidence, and asking what design would teach us more.

6.4 Step-by-step debrief script

Begin without revealing results. Ask for the benchmark prediction and which assumption students find most demanding. Record two quantitative class predictions. Next reveal the observation count and distribution, initially withholding the mean if it would distract from heterogeneity. Ask students to describe the chart without explaining causes.

In a second stage, show aggregate measures and ask: which observation contradicts the simplest prediction, which remains compatible with it, and what heterogeneity disappears in the average? Accept multiple explanations, but require each to yield a testable prediction. Introduce the variants: base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight. For each variant, identify the single changed element, predicted direction, and report measure to inspect.

Close by connecting the laboratory to diagnosis, forecasting, evidence assessment, and administrative decisions. Point out which elements of the real setting are absent: institutions, selection, repetition, long-run consequences, or rich information. The analogy is useful when it isolates a mechanism; it becomes misleading if the class assumes that the laboratory reproduces the full context.

A verbal sequence that works well is: “first describe,” “now derive,” “then interpret,” and “finally ask what evidence is missing.” This discipline allows a nonspecialist instructor to lead a rigorous discussion without treating every departure as a bias or every match as definitive confirmation.

6.5 Comparable class plans

6.5.1 A 50-minute class

Use one mode. Spend five minutes on rules and comprehension, five on private predictions, ten on the activity, five collecting predictions about the aggregate, fifteen revealing the report in layers, and ten connecting the result to theory. The goal is not to cover every variant but to ensure that every student can explain how judgments are formed under uncertainty and when useful shortcuts diverge from probability rules. End with one written sentence separating outcome from mechanism.

6.5.2 An 80-minute class

Run the baseline and one variant from base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight. Allow ten minutes for instructions, fifteen for each treatment, twenty for comparing distributions, and fifteen for theory and applications. Before the second treatment, request a directional prediction. If the result points the other way, do not dismiss it: inspect comprehension, composition, denominators, and descriptive power.

6.5.3 A 110-minute class

Combine experience, derivation, and design. After two treatments, give small groups a rival explanation and ask them to propose a third condition that distinguishes it. Each group must define the manipulated variable, outcome, unit of analysis, and prediction. Close by comparing proposals with existing variants and with diagnosis, forecasting, evidence assessment, and administrative decisions. This stage turns the student from participant into experimental designer.

6.6 Exercises with answer guidance

6.6.1 1. Reconstructing the prediction

Question. State the benchmark and three indispensable assumptions. What observation would be a departure?

Guidance. Begin from use base rates, Bayes' rule, set inclusion, independence, and sampling properties as the problem requires. Mention incentives, information, and horizon, and define the observation at the correct level. A departure is inconsistent with the full assumption set; it does not automatically identify its cause.

6.6.2 2. Auditing a percentage

Question. The report displays 60 percent. Propose two possible denominators and explain how interpretation changes.

Guidance. It could be 60 percent of eligible people or of completed decisions. If some people did not decide, the rates differ. A complete answer names numerator, denominator, exclusions, and aggregation level.

6.6.3 3. Reading heterogeneity

Question. Two classes have the same mean, but one is concentrated around it and the other has two modes. What does the instructor learn?

Guidance. The mean does not exhaust the distribution. The second class may combine different strategies or interpretations. Compare the median, dispersion, and categories linked to absolute error from the reference; treatment difference; normative-response rate; confidence-accuracy gap; consistency across items before explaining mechanisms.

6.6.4 4. Separating explanations

Question. Select two mechanisms from this list: representativeness, availability, anchoring, the law of small numbers, gambler's fallacy, conjunction, confirmation, overprecision, and hindsight.

Design a contrast that produces different predictions.

Guidance. Change one relevant element, hold others constant, and state each directional prediction. A variant from base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight may help, but the justification must connect manipulation and report measure.

6.6.5 5. Interpreting no difference

Question. Two treatments produce similar means. Does the manipulation not matter?

Guidance. Not necessarily. There may be weak comprehension, a small sample, offsetting responses, manipulation failure, or effects in tails and dispersion. Inspect descriptive uncertainty and the full distribution before concluding.

6.6.6 6. Transferring carefully

Question. Relate the activity to diagnosis, forecasting, evidence assessment, and administrative decisions and identify two contextual differences.

Guidance. A convincing application states the common mechanism and the limits. Real consequences, selection, experience, institutions, and repetition can change behavior. The conclusion should be an informed hypothesis rather than a universal prediction.

6.7 Data, inference, and learning assessment

6.7.1 A minimum results table

Prepare a table with treatment, eligible participants, completed decisions, independent unit, primary measure, median, and dispersion. For Heuristics and beliefs, the primary measure should come from absolute error from the reference; treatment difference; normative-response rate; confidence-accuracy gap; consistency across items. Add a note describing exclusions. With rounds, report people and groups as well as decisions; with roles, present each role separately.

6.7.2 Comparing without overstatement

An internal comparison requires assigned treatments under the same protocol. A comparison across classes can generate questions but may mix cohort, time, instructor, and experience. Report absolute differences and, when accessible, an uncertainty interval. Do not interpret nonsignificance as equality or statistical significance as automatic economic importance.

Do not select the metric after seeing which one produces the most attractive story. Define a primary measure in advance and use the others diagnostically. With many variants, some differences arise by chance. In a teaching session, transparency about this issue is more valuable than simulating research precision.

6.7.3 Assessing understanding

Use three products: a prior prediction, a posterior explanation, and a design proposal. The prediction assesses whether the student derives the benchmark; the explanation assesses whether evidence and mechanism are separated; the design shows understanding of identification. A short

rubric can award one point each for correct rules, prediction, denominator, limited interpretation, and a relevant treatment.

6.7.4 Closing protocol

Before ending, confirm that nobody remains on a wait page, close the session, preserve the export, and record treatment, date, and incidents. Do not project individual codes. Review open text before displaying it. Documenting parameter changes allows the next instructor to compare sessions responsibly.

6.8 Operational glossary and contingencies

Theoretical benchmark: a prediction derived from an explicit assumption set. **Treatment:** a condition changing one dimension of the environment. **Unit of analysis:** the object treated as independent for description or comparison. **Denominator:** eligible observations used to calculate a rate. **Heterogeneity:** differences across people, groups, or rounds hidden by a mean. **Mechanism:** a process connecting a condition to behavior. **ECU:** experimental currency unit.

If someone does not advance, first check whether the page is waiting for another person and then inspect the monitor; do not refresh repeatedly. If a link says that the user does not exist, create a new session. If records are missing, report coverage rather than filling values. If the aggregate differs from expectation, preserve the session: the discrepancy is pedagogical material for revisiting assumptions, comprehension, and design.

The instructor's final checklist should confirm rules, treatment, expected attendance, routes, hidden report, reveal plan, denominators, closing question, and export. In Heuristics and beliefs, a strong session ends when students can explain how judgments are formed under uncertainty and when useful shortcuts diverge from probability rules with a prediction, one piece of evidence, and one limitation.

6.9 Frequently asked questions and model answers

Should theory be explained before play? In a first session, collect predictions without revealing the full solution. Explain rules, payoffs, and information. Then derive use base rates, Bayes' rule, set inclusion, independence, and sampling properties as the problem requires. In a second session, reverse the order and assess whether theory improves prediction or only changes vocabulary.

Is there a correct answer? Separate a normative or equilibrium solution from a behavioral prediction. Theory supplies a conditional benchmark. The laboratory reveals what occurs and why a reasonable person may choose differently.

Does a departure prove irrationality? No. It may reflect omitted preferences, beliefs, error, weak comprehension, or a different institution. Check controls and incentives. Then treat representativeness, availability, anchoring, the law of small numbers, gambler's fallacy, conjunction, confirmation, overprecision, and hindsight as hypotheses requiring further comparisons.

Why show the distribution? The same mean can arise from moderate consensus or two opposing groups. A distribution reveals focal values, boundaries, and heterogeneity. In Heuristics and beliefs, connect those shapes to absolute error from the reference; treatment difference; normative-response rate; confidence-accuracy gap; consistency across items.

How many observations are enough? There is no universal threshold for debriefing. A small class is useful when described cautiously. Treatment comparisons require several independent units, clear assignment, and common protocols. Repeated decisions do not create new participants.

Can sessions be pooled? Yes, if parameters and definitions remain fixed. Retain session identifiers and inspect each cohort first. If instructions, composition, or incentives change, document the heterogeneity.

What if results contradict the literature? Check rules, treatment, comprehension, exclusions, and calculation. If no error exists, preserve the result. A class need not reproduce a published estimate. Ask what design would distinguish chance, composition, and mechanism.

How can embarrassment be avoided? Project aggregates and anonymous sequences. Avoid personal labels. Describe actions and consequences. The laboratory studies decisions under specific rules; it does not diagnose traits.

Should the decision be graded? Do not grade agreement with a prediction. Assess understanding: derive the benchmark, calculate a measure, explain a limitation, and design a comparison. Experimental incentives and academic assessment serve different purposes.

How should a variant be chosen? Begin with the mechanism to separate. Select from base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight, state a direction and measure, and hold other elements constant. Do not run more conditions than can be analyzed.

What conclusion is appropriate? Use three sentences: what was observed, which mechanism it is consistent with, and what evidence is missing. The connection to diagnosis, forecasting, evidence assessment, and administrative decisions should preserve those limits.

What should students remember? They should reconstruct the central tension, explain the benchmark, read the main denominator, and propose an institution that could change the outcome.

6.10 Diagnostic cases for interpretation practice

6.10.1 Agreement with the benchmark

The distribution is concentrated near the prediction derived from use base rates, Bayes' rule, set inclusion, independence, and sampling properties as the problem requires. Do not conclude that everyone shared the same motive. Some calculated, others followed a signal, and others chose intuitively. The evidence supports the aggregate prediction under session conditions but does not identify individual reasoning. Ask what belief, process, or response-time measure would distinguish those routes.

Then introduce a variant from base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight. If the pattern remains when the seemingly obvious mechanism weakens, the original explanation needs revision. A correct prediction does not automatically validate all its assumptions.

6.10.2 Two groups with opposite behavior

Half the class is at one extreme and half at the other; the mean lies where almost nobody chose. Showing only the average would create a fictional description. Calculate the median, modes, range, and share in theoretical categories. Ask whether heterogeneity is related to role, round, treatment, or comprehension.

Consider representativeness, availability, anchoring, the law of small numbers, gambler's fallacy, conjunction, confirmation, overprecision, and hindsight. Write one additional observation for each. If it predicts role differences and none appear, it receives less relative support. If it predicts learning, inspect the trajectory. Do not fill missing measurement with a compelling story.

6.10.3 A large difference with few units

One treatment apparently doubles the main measure but contains two groups and the comparison contains three. Display each group before pooling. One extreme unit can create the difference. Twenty decisions nested in two groups are not twenty independent units.

Ask students to plan the next session: groups, assignment, primary measure, and exclusions. The answer should mention absolute error from the reference; treatment difference; normative-response rate; confidence-accuracy gap; consistency across items and incomplete data. The experience thereby introduces design and transparency.

6.10.4 A null difference

Two conditions have the same mean. Before declaring that the manipulation does not matter, inspect compliance, comprehension, tails, and dispersion. Opposing subgroup effects can cancel. Define what range of differences would be pedagogically or economically meaningful and what size could distinguish it.

The cases teach four disciplines: do not infer motives from agreement, hide heterogeneity, confuse rows with units, or treat absence of evidence as evidence of absence.

6.11 Explanatory sequence for a complete class

6.11.1 Step 1: pose the problem without labels

Present the decision as a concrete situation and ask what information is missing. Do not announce the name of the bias, equilibrium, or preference at the start. The question is how judgments are formed under uncertainty and when useful shortcuts diverge from probability rules. Collect private responses so the first speaker does not become a focal point.

6.11.2 Step 2: reconstruct rules and incentives

Ask one student to describe actions and another to describe consequences. A third explains available information. Draw the sequence. In Heuristics and beliefs, the decision is to estimate probabilities or quantities, choose hypotheses, and report confidence. Do not move to interpretation until the class can calculate a simple case.

6.11.3 Step 3: derive the benchmark

Write use base rates, Bayes' rule, set inclusion, independence, and sampling properties as the problem requires. Underline each assumption. Ask what happens if one fails. Distinguish a logical derivation from an empirical expectation: an equilibrium strategy need not be played immediately by everyone.

6.11.4 Step 4: record predictions

Request one prediction about the mean, another about the distribution, and a third about a variant. Predictions should use absolute error from the reference; treatment difference; normative-response rate; confidence-accuracy gap; consistency across items. Preserve the record; later comparison reveals learning and reduces hindsight.

6.11.5 Step 5: reveal in layers

Show coverage first, then distribution, measures, and comparisons. At every layer request description before explanation. If someone proposes representativeness, availability, anchoring, the law of small numbers, gambler's fallacy, conjunction, confirmation, overprecision, and hindsight, ask what observation would distinguish it from another mechanism.

6.11.6 Step 6: work through one sequence

Use the worked example in this guide. Recalculate slowly and change one parameter. Ask each team to find one excessive conclusion and rewrite it with limited language. Editing claims is as important as calculation.

6.11.7 Step 7: compare institutions

Choose one condition from base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight. Write a table with constant, change, prediction, and measure. If the variant changes several elements, acknowledge it. Ask whether a third treatment would separate explanations.

6.11.8 Step 8: transfer and close

Connect the mechanism to diagnosis, forecasting, evidence assessment, and administrative decisions. Identify two differences between the laboratory and the real context. End with an exit response: prediction, evidence, compatible mechanism, and limitation. This structure assesses understanding without rewarding a particular decision.

6.11.9 A short script to read aloud

"Describe what happened before explaining why. Identify the prediction and its assumptions. Compare the distribution with the mean. Propose two compatible mechanisms. Finally, state which treatment or additional evidence could distinguish them." The script keeps discussion in a common sequence across families.

6.12 Assessment bank with solutions

Exercise 1. List actions, information, and payoffs. **Solution:** reconstruct estimate probabilities or quantities, choose hypotheses, and report confidence without adding information unavailable to the participant.

Exercise 2. Derive the material benchmark. **Solution:** begin from use base rates, Bayes' rule, set inclusion, independence, and sampling properties as the problem requires and name at least

two assumptions. For an interactive game, state best responses; for an individual task, state the decision rule.

Exercise 3. The report shows 40 percent. **Solution:** it cannot be interpreted without numerator, denominator, exclusions, and unit. Eligible people and completed decisions are alternative denominators.

Exercise 4. The mean matches the prediction. **Solution:** inspect the distribution. A correct mean can combine extremes, offsetting errors, or different strategies. Agreement does not identify reasoning.

Exercise 5. Select two mechanisms from representativeness, availability, anchoring, the law of small numbers, gambler's fallacy, conjunction, confirmation, overprecision, and hindsight. **Solution:** design a condition changing one prediction but not the other, with an outcome from absolute error from the reference; treatment difference; normative-response rate; confidence-accuracy gap; consistency across items.

Exercise 6. Compare treatments of different sizes. **Solution:** report counts and shares, preserve the independent unit, and do not compensate for a small group by multiplying rounds.

Exercise 7. Mean differences are absent. **Solution:** inspect dispersion, tails, compliance, and comprehension; define a relevant difference. A broad null does not establish equivalence.

Exercise 8. Propose a figure. **Solution:** use histograms or dots for distributions, lines for rounds, separate bars for roles, and bivariate plots when a response depends on a prior action.

Exercise 9. Explain a variant from base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight. **Solution:** identify what changes, what remains constant, predicted direction, and an interpretive threat.

Exercise 10. Transfer to diagnosis, forecasting, evidence assessment, and administrative decisions. **Solution:** describe the common mechanism and two institutional differences. Do not claim that one session determines policy.

Exercise 11. Write a three-sentence conclusion. **Solution:** observation with denominator; compatible mechanisms; evidence needed to distinguish them.

Exercise 12. Design a replication. **Solution:** preserve parameters, record assignment, announce the primary measure, set exclusions, and archive the export.

A twelve-point rubric awards one point for each correct component. For a short assignment use Exercises 2, 3, 5, and 11. For a project use all twelve and request a reproducible figure.

6.13 Advanced seminar: from evidence to mechanism

Divide the class into four teams: benchmark model, alternative preferences or beliefs, institutions, and identification critique. Each interprets the same report from its assigned role. The first derives use base rates, Bayes' rule, set inclusion, independence, and sampling properties as the problem requires; the second uses representativeness, availability, anchoring, the law of small numbers, gambler's fallacy, conjunction, confirmation, overprecision, and hindsight; the third selects a variant from base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight; the fourth identifies missing evidence.

In a replication round, every team writes a quantitative prediction and a condition that could refute it. It exchanges the proposal with another team, which audits unit, denominator, and manipulation. A useful explanation does not merely accommodate past outcomes; it risks a prediction about new evidence.

End with a one-page memo containing question, benchmark, descriptive finding, two mechanisms, discriminating design, application to diagnosis, forecasting, evidence assessment, and administrative decisions, and limitation. Grade causal clarity, technical correctness, use of evidence, and inferential restraint.

For a quantitative extension, export data and reproduce three figures: individual distribution, aggregate by independent unit, and treatment comparison. All three require titles, denominators, and exclusion notes. Compare how the narrative changes when the correct unit is used.

An ethics extension asks which decisions should remain private, when an incentive is proportionate, and how to communicate results without turning group patterns into personal judgments. The answer should separate participant protection, instructional integrity, and analytical transparency.

6.14 Quality control before teaching

6.14.1 Full rehearsal

Open a test session and complete one full participant path. Check cover, language, comprehension, decision, waiting, result, and close. In *Heuristics and beliefs*, verify especially that the participant understands how to estimate probabilities or quantities, choose hypotheses, and report confidence. Open the report before class, keep it hidden, and confirm that absolute error from the reference; treatment difference; normative-response rate; confidence-accuracy gap; consistency across items appear with their denominators.

6.14.2 Fifteen control questions

1. Does the configuration belong only to this family and the correct language?
2. Can expected attendance form every required unit?
3. Do the rules describe actions, information, order, and payoffs?
4. Does the comprehension question detect the most likely mistake?
5. Is the benchmark derived rather than merely asserted?
6. Will students record a prediction before seeing the result?
7. Does the report show counts as well as percentages?
8. Does every rate state numerator and denominator?
9. Is the distribution readable from the classroom?
10. Are roles, treatments, and rounds labeled and colored consistently?
11. Was the main comparison selected before observing data?
12. Is there time to describe results before interpreting them?
13. Do questions avoid judgments about individual participants?
14. Does the link to diagnosis, forecasting, evidence assessment, and administrative decisions state explicit limits?
15. Will the export and incidents be documented?

6.14.3 A micro-rehearsal of the debrief

Practice a three-minute explanation. Begin with use base rates, Bayes' rule, set inclusion, independence, and sampling properties as the problem requires. Then show a hypothetical distribution and state two compatible interpretations from representativeness, availability, anchoring, the law of small numbers, gambler's fallacy, conjunction, confirmation, overprecision, and hindsight. Identify which variant among base rates; availability; anchoring; small samples; gambler's fallacy; Bayes; conjunction; confirmation; calibration; hindsight would distinguish them. End with one limitation. If the sequence cannot fit in three minutes, simplify the session objective.

6.14.4 After class

Record attendance, completed decisions, parameters, and incidents. Note which explanation caused confusion and which graph worked. Preserve one question for the next cohort. Cumulative improvement depends on documenting experience rather than changing parameters without a record.

The closing condition is concrete: a person who did not attend should be able to read the report and know what was decided, what the benchmark was, how many units participated, what pattern appeared, and what cannot be concluded. If any element is missing, complete the instructor note before archiving the session.

Bibliography

- Fischhoff, B. (1975). Hindsight $\neq$ foresight. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288–299. doi:10.1037/0096-1523.1.3.288.
- Gigerenzer, G., and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction. *Psychological Review*, 102(4), 684–704. doi:10.1037/0033-295X.102.4.684.
- Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237–251. doi:10.1037/h0034747.
- Moore, D. A., and Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517. doi:10.1037/0033-295X.115.2.502.
- Rabin, M. (2002). Inference by believers in the law of small numbers. *Quarterly Journal of Economics*, 117(3), 775–816. doi:10.1162/003355302760193896.
- Tversky, A., and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105–110. doi:10.1037/h0031322.
- Tversky, A., and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232. doi:10.1016/0010-0285(73)90033-9.
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. doi:10.1126/science.185.4157.1124.
- Tversky, A., and Kahneman, D. (1983). Extensional versus intuitive reasoning. *Psychological Review*, 90(4), 293–315. doi:10.1037/0033-295X.90.4.293.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3), 129–140. doi:10.1080/17470216008416717.