



Instructor's Guide to Confidence Calibration and Trading

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 prepares instructors to connect measured confidence with actual financial trading. It follows the English interface and covers the same economic rules as the Spanish edition. No prior experience conducting classroom experiments is assumed.

Begin with the summary, information sequence and worked ledger. Select one extended comparison rather than asking every student to complete all conditions. Classroom planning, report definitions and solved exercises support discussion once choices are recorded.

The guide distinguishes calibration, relative beliefs and financial outcomes. Screenshots are real local interfaces. Canonical public routes describe the prepared product; verified local operation and pending public deployment remain explicit.

Author

This guide was prepared by Stanislao 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

Record predictions before revealing aggregates. Keep individual feedback and participant links private. Discuss transactions using information, resources and announced incentives.

Contents

How to use this guide	i
1 Purpose, scope and access	1
1.1 The economic question organizing the activity	1
1.2 Preparing access and selecting a session	2
2 Measurement, incentives and market accounting	9
2.1 Eight answers and probabilities: what is observed	9
2.2 The quadratic rule and its assumptions	10
2.3 A worked confidence and relative-position example	10
2.4 Known-quality signals and beliefs before and after trading	11
2.5 How the shared order book operates	11
2.6 One execution and the distinction between gross and net	12
2.7 Combining rewards without paying twice	12
2.8 Own portfolios and management contracts	13
2.9 What the literature contributes and what this activity adapts	13
3 Comparisons in the extended edition	16
3.1 A common reference for fourteen conditions	16
3.2 Reference: measurement before trading	16
3.3 Early accuracy feedback	17
3.4 Early calibration feedback	17
3.5 Private relative-position feedback	17
3.6 A one-ECU commission per unit and side	18
3.7 A five-ECU commission per unit and side	18
3.8 A low-accuracy private signal	18
3.9 A high-accuracy private signal	19
3.10 A common public signal	19
3.11 Bayesian assistance before trading	19
3.12 No liquidity objective	20
3.13 A two-unit liquidity objective	20
3.14 Delegated management with linear compensation	20
3.15 Delegated management with convex compensation	21
4 Classroom use, data interpretation and exercises	22
4.1 A seventy-five-minute session	22
4.2 Conducting the discussion after decisions	23
4.3 The private report: observations and denominators	23
4.4 Downloads and reconstruction of the record	24
4.5 Exercises with solutions	27
4.5.1 Answers to the initial measurement	27
4.5.2 Equal accuracy, different probability quality	27

4.5.3	Ties in relative position	27
4.5.4	A cost and a target in one transaction	28
4.5.5	Belief revision before value is revealed	28
4.5.6	Ten periods and one initial measurement	28
4.6	Operational problems and limits of use	28

List of Figures

1.1	Laboratory website: institutional access to teaching activities.	3
1.2	Public catalogue: select the game and language.	4
1.3	Demo: opening view of the filtered 34-configuration registry.	5
1.4	The four family rooms.	5
1.5	Basic room form: language and period sequence.	6
1.6	Welcome and laboratory rules before decisions.	7
1.7	An incorrect answer retains entered values and receives specific guidance.	8
2.1	First question and confidence report: select an alternative and its probability. . . .	14
2.2	Pre-market belief report without redemption-value disclosure.	14
2.3	Order book and submission during actual trading.	15
3.1	Convex management contract: distinguish the manager and owner accounts.	21
4.1	The instructor controls when aggregates are revealed.	24
4.2	Individual financial result: portfolio, costs and rewards separated.	25
4.3	Revealed private report: summary and reliability with denominators.	26
4.4	Explicit ending: accumulated payment from active periods.	26

Purpose, scope and access

Purpose	Separate confidence calibration, relative beliefs and financial trading.
Experimental unit	Hybrid: private individual measurements followed by an eight-person market.
Basic edition	One market; ten fixed-group or randomly reassigned periods.
Extended edition	Fourteen feedback, information, commission, liquidity and management conditions.
Decision	Eight answers and confidence reports, relative forecast, two asset beliefs and orders.
Payoff	Portfolio or management income, inventory bonus and quadratic probability rewards; initial rewards paid once.
Topics	confidence calibration, probabilistic beliefs, trading volume, transaction costs, financial information, management incentives
JEL	C91; C92; D83; G14; G41
Participants	Multiples of eight; at least sixteen for random reassignment.
Suggested duration	Approximately 75 minutes including discussion for one period; longer for ten periods.

https://calibrationtrading.beer-lab.org/room/calibration_en

1.1 The economic question organizing the activity

A financial transaction may reflect information, an inventory target, a need to sell or a compensation contract. Observing heavy trading does not by itself identify overconfidence. This activity adds measurement before an actual market among students. Each person first answers eight questions and reports the probability of being correct. They also forecast their relative position. Participants then receive a signal of announced quality, state a belief about asset value, trade and report that belief again before learning the true redemption value. The sequence records both judgments and behavior rather than inferring one from the other.

The design separates objects that are often confused. Accuracy is the fraction of correct answers. Reported confidence is the probability attached to those answers. Calibration compares probabilities with realizations. Relative position expresses an expectation about other people. A belief about a security's value uses particular financial information. None automatically substitutes for another. The report preserves these distinctions and displays trading with explicit costs, targets and resources. Instructors can discuss associations without attributing an unmeasured motive to a

participant. A large transaction is an observed action, not a diagnosis.

The design is hybrid. Its initial block is individual and private; its subsequent block brings eight people together in an order book. Market decisions and results are interdependent. Redemption value is common and transactions transfer cash and units among participants. Every market requires eight people. Reassignment requires at least sixteen so that two markets exist. There are no demographic questions. The owners in management conditions are automated accounts rather than a ninth participant or an additional person choosing a contract. Students make their own trading decisions under the announced compensation rule.

This delivery uses the existing microstructure family's trading engine. It preserves price-time priority, standing orders, resource reservations, partial execution and cancellation. The new component connects prior measurements with later trading behavior and outcomes. The catalogue also contains individual heuristics and calibration activities, which do not connect measurement with this interactive market. This is neither the dividend-and-bubbles game nor diversified portfolio allocation. Its prices, signals, inventory targets and measures address a separate curricular question while reusing tested trading infrastructure. The distinction prevents counting translations or a renamed order book as new mechanisms.

For the instructor

Write separate predictions about accuracy, beliefs and trading. To study a causal effect, assign conditions through an explicit procedure. A correlation between the calibration gap and turnover does not by itself establish that the former causes the latter. Keep liquidity motives and compensation contracts visible when discussing an observed transaction.

1.2 Preparing access and selecting a session

Institutional navigation starts at <https://beer-lab.org/> and continues to teaching games at <https://teaching.beer-lab.org/>. Find Confidence Calibration and Trading and choose a language. The proposed canonical domain is <https://calibrationtrading.beer-lab.org/>. This route describes the prepared product; publication requires an administrator's server update. The verified local instance is <http://localhost:8105/demo>, accessible from this computer. Local operation does not establish public deployment. Explain this distinction when distributing materials prepared before the public site is updated.

Demo contains only this family's 34 configurations. Rooms lists four rooms in basic Spanish, basic English, extended Spanish and extended English order. Basic rooms are https://calibrationtrading.beer-lab.org/room/calibracion_es and https://calibrationtrading.beer-lab.org/room/calibration_en. Extended names end in `calibracion_extendida` and `calibration_extended`. Each room filters its exact prefix. Basic menus have three options and extended menus fourteen. They must not show another family or start a basic room with an extended condition. The screenshots below show the filtered local instance.

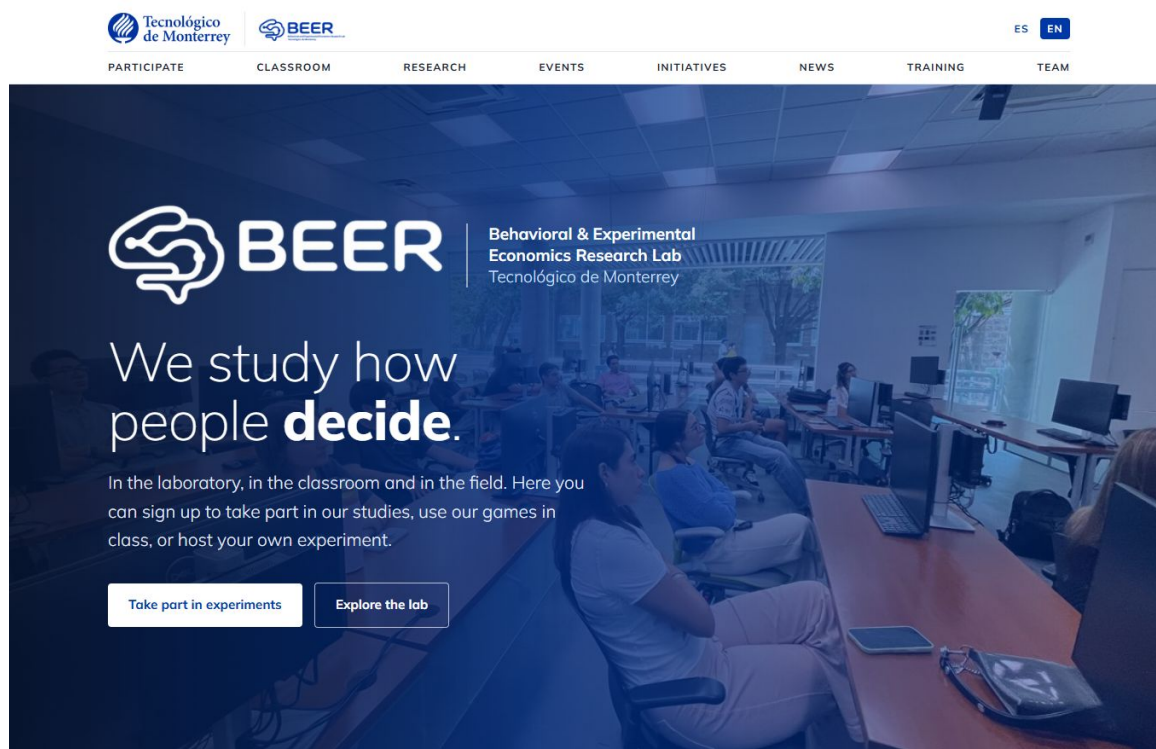
Basic modes offer one market, ten fixed-group periods and ten randomly rematched periods. Calibration occurs once before period one. Trading resources reset each period, while confirmed rewards accumulate. Reference feedback about questions and relative position is delivered after the first market. In later periods everyone already has that experience; those observations are not new untreated cases. Preserve this timing distinction when comparing the reference with variants providing information before trading. Rematching changes market partners rather than repeating the initial questionnaire or defining a new reference group for relative rank.

The instructor creates a session through administrative access, selects a valid multiple of eight and distributes that session's new links. Personal links are not projected or incorporated into a guide.

Open Admin report from the session to inspect progress. Aggregates require explicit revelation and can be hidden again. Downloads belong to the selected session. Each domain has its own family and database, so a tab belonging to another game does not automatically share people or sessions. A new classroom run should use a newly created session rather than old links from a deleted one.

Attention

Prepare comparison sessions before opening live trading in this local instance. The real-time engine uses one synchronous worker; concurrent administrative creation during active markets needs separate deployment-environment validation. Restart the instance to apply code changes while preserving its database. Do not reset the database to refresh templates or distribute obsolete participant links.



WHAT WE DO

ALL ABOUT BEER

Figure 1.1: Laboratory website: institutional access to teaching activities.




[PARTICIPATE](#)
[CLASSROOM](#)
[RESEARCH](#)
[EVENTS](#)
[INITIATIVES](#)
[NEWS](#)
[TRAINING](#)
[TEAM](#)

[ES](#)
[EN](#)

INSTRUCTORS AND STUDENTS

Classroom experiments

Experimental economics games for classroom activities and discussion. Open the game your instructor tells you.

Instructors: how to use these games in class or propose a new one → [instructor guide](#)
 Lab research: beer-lab.org/en/research

Topic ▾

JEL code ▾

163 games

Electoral Accountability and Political Selection

Study how the prospect of remaining in office changes an officeholder's behavior and how information allows voters to choose between politicians with different types. Basic: One encounter or four encounters with fixed or reassigned groups. Extended: Eight treatments on continuation, information, types, office value, term limits, and audits.

electoral discipline

political accountability

political selection

corruption

information

political economy

JEL A22 · C92 · D72 · D73

Address: accountability.beer-lab.org

Aggregate Demand and Aggregate Supply

Understand how spending choices aggregate into demand and how equilibrium with supply determines output and prices. Basic: One period, ten periods with the same economy, or ten periods with reassigned economies. Extended: Demand expansion and contraction, supply shock, fiscal stimulus, capacity, expectations, and coordination.

aggregate demand

aggregate supply

price level

potential output

macroeconomic expectations

JEL A22 · C92 · E12 · E31 · E32

Address: adas.beer-lab.org

Figure 1.2: Public catalogue: select the game and language.

oTree Demo Sessions Rooms Data Server Check Logout

Demo

[ES] Calibración y negociación — Un mercado
[ES] Calibración y negociación — Diez períodos con grupos fijos
[ES] Calibración y negociación — Diez períodos con reasignación
[EN] Calibration and trading — One market
[EN] Calibration and trading — Ten fixed-group periods
[EN] Calibration and trading — Ten rematched periods
[ES] Calibración y negociación — Extendido — Referencia con devolución después del mercado
[ES] Calibración y negociación — Extendido — Devolución previa de aciertos
[ES] Calibración y negociación — Extendido — Devolución previa de calibración
[ES] Calibración y negociación — Extendido — Devolución privada de posición relativa
[ES] Calibración y negociación — Extendido — Comisión de una ECU
[ES] Calibración y negociación — Extendido — Comisión de cinco ECU
[ES] Calibración y negociación — Extendido — Señales con exactitud de 60%
[ES] Calibración y negociación — Extendido — Señales con exactitud de 90%
[ES] Calibración y negociación — Extendido — Señal común para el mercado
[ES] Calibración y negociación — Extendido — Referencia bayesiana antes de negociar
[ES] Calibración y negociación — Extendido — Sin meta de inventario bonificada
[ES] Calibración y negociación — Extendido — Meta de dos unidades
[ES] Calibración y negociación — Extendido — Gestión remunerada por riqueza final
[ES] Calibración y negociación — Extendido — Gestión remunerada por exceso positivo

Calibración de confianza y negociación / Confidence Calibration and Trading — BEER Lab

Seleccione una variante de este juego. Select a version of this game.

Figure 1.3: Demo: opening view of the filtered 34-configuration registry.

Rooms

Current rooms:

Calibración y negociación — Básico [ES]
Calibration and trading — Basic [EN]
Calibración y negociación — Extendido [ES]
Calibration and trading — Extended [EN]

Figure 1.4: The four family rooms.

oTree

Demo

Sessions

Rooms

Data

Server Check

Logout

Room: Calibration and trading — Basic [EN]

Create a new session

Session Config

[EN] Calibration and trading — One market

Number of participants

Complete groups of 8 participants. Use a multiple of 8; minimum 8.

Create

[Configure session](#) ⚙️

App sequence

calibrated_trading_game (10 rounds)	Price-time priority, reservations, execution and information in a multi-level limit order book.
--	---

0 participants present

Start links

These URLs will stay constant for new sessions, even if the database is recreated.

Room-wide URL

Here is the room-wide URL anyone can use. Don't use this link when testing in multiple tabs on the same browser, because all tabs will be assigned to the same participant, using a cookie.

http://127.0.0.1:8105/room/calibration_en

Figure 1.5: Basic room form: language and period sequence.



Welcome to the BEER Lab

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

You will make decisions about orders, cash and units of a security under the rules shown on screen.

General laboratory rules

- M
- a
- r
- k
- e
- t
-
- r
- u
- l
- e
- s

I confirm that I have read and will follow the laboratory rules.

☐ Yes

☐ No

Start

Figure 1.6: Welcome and laboratory rules before decisions.



Period 1/1

Decisions, estimates and trading

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.

Comprehension check

How much cash do you receive at the start of each period, in ECU?

Your answer: 400. The correct answer is 500. Initial cash is 500 ECU plus five units; cash and total wealth are different quantities.

A resting sell order asks 98. A new buy limit allows up to 102. At what price does it execute, ignoring fees?

Your answer: 102. The correct answer is 98. The resting order sets the execution price: 98 ECU; 102 is the buyer's maximum.

In an example without fees, how much cash does a pending two-unit buy at 100 ECU reserve?

Your answer: 100. The correct answer is 200. Without fees, two units at 100 ECU reserve $2 \times 100 = 200$ ECU.

If you report 50% for an event, how many ECU does the quadratic rule give whether it occurs or not?

Your answer: 50. The correct answer is 75. $100 \times [1 - (0.5 - 1)^2] = 75$, and $y=0$ also gives 75. One probability is not judged by whether one event happens.

[Continue](#)

Figure 1.7: An incorrect answer retains entered values and receives specific guidance.

Measurement, incentives and market accounting

2.1 Eight answers and probabilities: what is observed

Each question has two objectively evaluable alternatives. A person chooses one and reports a probability between fifty and one hundred percent that it is correct. Someone believing the other alternative more likely can switch their choice. The lower bound of fifty follows from this two-alternative structure; it does not claim that everyone has at least fifty-percent ability on any task. Confidence refers to the selected answer. The activity does not elicit confidence intervals, volatility forecasts or a general self-assessment of intelligence. Treat those as different instruments requiring their own interpretation.

The eight questions cover fractions, compounding, a lottery mean, recovery from a percentage loss, independent events, transaction costs, updating from a signal and a combination of payoffs. They are a teaching adaptation of financial and probabilistic reasoning rather than a validated psychological scale. Solutions are known and do not depend on news, encyclopedic knowledge or classmates' information. Accuracy-confidence comparisons describe performance on this set. Eight answers per person are noisy measurements and do not justify permanently labeling an individual overconfident. Repeated classroom use can also introduce familiarity with the questions.

For question i , write $y_i = 1$ if the selected answer is correct and zero otherwise. Let p_i be reported probability between zero and one. Accuracy is $100 \sum_i y_i / 8$, and mean confidence is $100 \sum_i p_i / 8$. Their difference in percentage points is $100 \sum_i (p_i - y_i) / 8$. A positive difference means mean confidence exceeds realized accuracy; a negative value means the reverse. Zero does not establish that every probability is calibrated: opposite errors across questions of different difficulty can cancel. The reliability table therefore complements the signed average.

Theoretical foundation

Calibration does not mean correctly predicting every event assigned a high probability. Across many answers reported at eighty percent, a correct-answer frequency near eighty percent would be compatible with calibration. One incorrect answer at eighty percent is not sufficient evidence of an irrational belief. Examine comparable sets, observation counts and question difficulty before interpreting a difference.

2.2 The quadratic rule and its assumptions

Each answer receives $100[1 - (p_i - y_i)^2]$ ECU, and the average across eight is paid. Reporting fifty percent gives seventy-five ECU whether the answer is correct or not. Reporting one hundred gives one hundred when correct and zero when incorrect. The rule distinguishes confidence from simply selecting an option. Comprehension errors can be corrected without reward reductions; the subsequent eight answers form part of the rewarded activity. ECU means Experimental Currency Units. No conversion to real money is specified, and the activity does not promise a cash payment absent an additional announced arrangement.

If subjective probability of being correct is π , expected score is $100[1 - (p - \pi)^2 - \pi(1 - \pi)]$. Under expected-ECU maximization it is maximized by reporting $p = \pi$, or the nearest feasible whole percentage. This property concerns expected score. Nonlinear utility, risk hedging or a different interpretation of the reward can change behavior. In particular, financial estimates concern the same value affecting portfolio outcomes. Do not claim that the mechanism removes every distortion for every preference. Explain the objective and assumptions when using scoring incentives to interpret a report.

The report also shows mean squared error, $\sum_i (p_i - y_i)^2 / 8$. It ranges from zero to one and differs from the calibration gap. It penalizes probability errors and combines aspects of probabilistic accuracy rather than measuring a pure signed bias. The reward equals one hundred times one minus this error. Audit results using internal precision before rounding the average. Final amounts are rounded to cents, with halves rounded up. Adding eight rewards as eight full payments would incorrectly multiply the incentive by eight; the announced payment is their mean.

2.3 A worked confidence and relative-position example

Suppose six of eight answers are correct and confidence is eighty percent on every answer. Accuracy is seventy-five and mean confidence eighty: a positive difference of five percentage points. Mean squared error is $(6 \times 0.04 + 2 \times 0.64) / 8 = 0.19$. The calibration reward is eighty-one ECU. Confidence exceeding accuracy therefore does not imply a zero reward, and accuracy and probabilistic score answer different questions. A person with fewer correct answers may avoid large probability errors by acknowledging uncertainty. Describe both performance and confidence instead of using one summary as a universal ranking of judgment.

After answering but before receiving feedback, each person forecasts the percentage of the other seven they will outperform in correct-answer count. Ties count as half a person. If four have fewer correct answers, one is tied and two have more, observed percentage is $100(4 + 0.5) / 7 = 64.2857 \dots$. A forecast of sixty has a relative gap of about minus 4.29 points. Reward is $100[1 - (0.60 - 4.5/7)^2]$, rounded to 99.82 ECU. It is paid once and does not publish an individual ranking list or reveal another person's answers.

This forecast concerns an expected fraction of pairwise comparisons won, not the probability of being best or a whole ordinal rank. If all eight people tie, everyone has an observed percentage of fifty. In any complete group the observed average is also fifty when ties count one half. Mean forecasts above fifty may describe relative overplacement in that group. This does not imply identical individual errors or represent an external population. The reference is the other seven people in the initial market, even if later trading partners are reassigned. Keep that reference fixed in analysis.

What participants will see

Your probability of answering a question correctly and your expected fraction of peer comparisons won are different reports. Six correct answers out of eight do not automatically mean outperforming seventy-five percent of the group. You need the other seven people's score distribution. Ties neither become complete wins nor disappear from the denominator.

2.4 Known-quality signals and beliefs before and after trading

A unit's redemption value is eighty or one hundred twenty ECU, with equal prior probabilities. Reference signals match the true value in seventy-five percent of cases. Quality is announced and does not depend on prior question accuracy. Private signals are independent conditional on common value. Someone seeing a low signal cannot know whether everyone else received the same message. In the common-signal condition, everyone instead observes one shared draw; eight observations of that message are not eight independent pieces of evidence. Signal correlation and individual quality are separate features of information.

Before trading, each person reports a probability from zero to one hundred that value is one hundred twenty. With a half prior and symmetric seventy-five-percent accuracy, the Bayesian reference is seventy-five after a high signal and twenty-five after a low one. The aid condition displays this reference explicitly. It is not an estimate of a person's ability and does not depend on relative rank. It uses announced information before the market opens, when no orders from others can yet reveal their signals. The benchmark therefore has a precise information set and date.

Another estimate is requested after trading closes but before redemption is revealed. Each estimate uses the same quadratic rule, now with $y = 1$ when redemption is one hundred twenty. Their average reward is paid each period. The second response may incorporate prices, trades and orders observed during the market. It should not be mechanically compared with the single-signal posterior as though no market information existed. Belief changes may reflect learning, interpretations of liquidity demand or noise. One realized common value cannot distinguish these explanations without further information or a specified inference model.

2.5 How the shared order book operates

Every person manages five hundred ECU and five units at the start of a period. They can submit a limit order to buy or sell a quantity at a specified price, or use a market order to take existing offers. Compatible orders execute by price and then server arrival. A limit order can execute partially and leave its remainder standing. Cancellation affects only unexecuted quantity. Market orders consume available levels compatible with resources and cancel any unfilled quantity; they do not create a future purchase obligation. The incoming order's limit is not necessarily its execution price: a matched trade uses the standing order's price.

Borrowing, selling nonexistent units and crossing one's own order are prohibited. Standing buys reserve enough cash for price and fees; standing sells reserve units. Those resources cannot be committed again to another order. Available balances distinguish total from reserved resources. Prices are whole ECU between one and two hundred, and quantities are whole units subject to resources. Server priority and request identifiers prevent a repeated message from executing an accepted request twice. These accounting restrictions are part of the existing tested book, while this family changes measurement and incentive rules around it.

Duration is configured, and the clock starts after all eight people submit their prior estimate.

Instructions and calibration do not consume trading time. A person may finish early: standing orders are canceled and others continue. The market closes when all finish or time expires. Nobody can add a trade afterward. Redemption remains absent from the posterior-estimate screen and appears only in results. Waiting pages synchronize actual choices within a market; they do not impose artificial pairs on the individual block. Keep the last participant's arrival and the start of the trading clock separate when organizing a classroom session.

2.6 One execution and the distinction between gross and net

Consider realized value of one hundred twenty and a one-unit trade at one hundred. Initially each person holds five hundred cash and five units, or benchmark redemption wealth of eleven hundred. The buyer ends with four hundred and six units, worth eleven hundred twenty. The seller holds six hundred and four units, worth one thousand eighty. Trading gains are plus twenty and minus twenty. The trade redistributes wealth rather than creates joint value. Benchmark denominators use the same realized redemption value, not different earlier prices for each participant. There is no gain from simply relabeling a unit's common valuation.

If each side pays a one-ECU fee, cash becomes three hundred ninety-nine and five hundred ninety-nine. Net wealth is eleven hundred nineteen and one thousand seventy-nine. Net gains are plus nineteen and minus twenty-one; gross gains remain plus twenty and minus twenty after adding fees paid. The exchange operator receives two ECU. Aggregate gross gains are zero, and aggregate net gains equal minus fees collected. This identity audits transaction costs without attributing the aggregate loss to an incorrect signal. Gross and net outcomes use the same executed quantities rather than hypothetical trades at another price.

If buyer and seller satisfy their respective targets to buy and sell at least one net unit, each receives fifty additional ECU. This externally financed reward is not an asset return. With a one-ECU fee on each side, portfolio income including targets is eleven hundred sixty-nine and eleven hundred twenty-nine. Market volume was one executed unit; summed person-level traded units are two because they count both sides. Each person's turnover is one hundred divided by eleven hundred, about 9.09 percent. Its denominator is not initial cash of five hundred alone. Inventory rewards, trading gains and notional turnover describe different objects.

Checkpoint

Check cash plus units valued at redemption for each person. Then separate fees, inventory bonuses and estimate rewards. Do not add a purchase price again to terminal wealth. A trade has one market volume and two person-level participations. Confusing them artificially doubles class-wide trading volume and distorts comparisons across sessions.

2.7 Combining rewards without paying twice

Continue with the buyer paying a fee. Portfolio income including the target reward is 1169 ECU. Reports of seventy-five percent before and eighty after, followed by redemption of one hundred twenty, score 93.75 and 96. Their average rounds to 94.88. Adding the earlier example's calibration reward of eighty-one and relative-position reward of 99.82 gives first-period payoff $1169 + 94.88 + 81 + 99.82 = 1444.70$ ECU. Initial trading balances are not added again. This numerical reconciliation is useful before discussing any apparent relation between confidence and total earnings.

The next period resets cash and units but neither repeats nor repays the eight-question or relative-

position blocks. Only that period's portfolio or management income and its pair of value-estimate rewards are added. The final total sums active-period payoffs. Missing a posterior response leaves a pending evaluation even though the market closed and portfolio accounting can already be reconstructed. The report separates closed markets from completed period evaluations. It does not replace a missing estimate with fifty or zero. A closed book is therefore not sufficient evidence that every participant has completed the full activity.

2.8 Own portfolios and management contracts

In management conditions, the same five hundred ECU and five units belong to an automated owner. The person chooses trades but receives a contractual wage rather than the entire portfolio's wealth. The owner has one hundred ECU outside the book to finance the fixed wage component. The linear contract pays one hundred plus twenty percent of final wealth including inventory bonus. The positive-excess contract pays one hundred plus twenty percent of positive excess over initial buy-and-hold redemption wealth. Both are separate from calibration and belief rewards. No other student selects these terms or bears the owner's residual account.

With portfolio wealth of 1169 and benchmark of 1100, linear pay is 333.80 and owner residual is 935.20. Positive-excess pay is 113.80 and residual is 1155.20. In both cases wage plus residual equals 1269, exactly portfolio wealth plus the one-hundred wage reserve. These contracts differ in level and shape; they do not identify curvature alone with equalized expected pay. Study trading choices and portfolio outcomes alongside wages before attributing behavior to management incentives. The owner reserve is not available for trading and cannot finance an order exceeding the portfolio's cash or units.

2.9 What the literature contributes and what this activity adapts

[Barber and Odean \(2000\)](#) study trading and individual-investor performance and emphasizes distinguishing gross from net outcomes. That distinction motivates recording fees and a buy-and-hold benchmark. Our session neither reproduces a household sample nor estimates annual returns. Numerical magnitudes here are teaching parameters, not estimates from that study. Interpret a classroom return difference under the resources, signals and targets announced in this activity. A high realized gain is not automatically evidence of a well-calibrated probability report.

[Glaser and Weber \(2004\)](#) distinguish facets of overconfidence and links judgment measures with trading activity. That evidence cautions against interchanging calibration with relative self-assessment or assuming every measure must correlate with volume. This activity preserves their separation but uses eight original questions and small groups. It does not replicate the questionnaire or treat one class correlation as sufficient confirmation or refutation. The teaching question is what was measured and what alternative explanation remains compatible with the same transactions. Report descriptive patterns without turning a label into an unobserved mechanism.

[Gneiting and Raftery \(2007\)](#) develop the theory of proper scoring rules. The connection is expected-score maximization under the assumptions explained above rather than guaranteed behavior under every preference. [Bloomfield et al. \(2005\)](#) study liquidity provision and consumption in experimental markets, supporting attention to inventory motives and the choice to make or take orders. The book's budgets, targets and rewards here are explicit teaching adaptations. Together these references support a disciplined distinction between elicitation, contractual incentives and market outcomes rather than a claim that every trade has the same psychological origin.



Period 1/1

Decisions, estimates and trading

Answer and estimate your probability of being correct

Answer eight two-option questions. For each, report a percentage probability from 50 to 100 that your answer is correct. If you think the other option is more likely, select it. Solutions are not shown yet.

The reward for each answer is $100 \times [1 - (p - y)^2]$ ECU, where p is probability between 0 and 1 and y is 1 if correct or 0 otherwise. The average across eight answers is paid once.

1. Which fraction is larger?

Your answer

☒ 3/4

☐ 2/3

Probability that your answer is correct (50–100%)

Figure 2.1: First question and confidence report: select an alternative and its probability.



Period 1/1

Decisions, estimates and trading

Estimate before trading

Your signal: **80.0 ECU**

Value is 80 or 120 ECU with equal probabilities. The signal matches the true value in 75% of cases. Your previous answers do not change its quality.

People's signals are drawn independently conditional on the common value. You see only your own signal.

One estimate is requested before and one after trading, before the value is revealed. Each scores $100 \times [1 - (p - y)^2]$, where $y=1$ if the value is 120. The average of the two rewards is paid each period.

Probability that redemption value per unit is 120 ECU (0–100%)

[Continue](#)

Figure 2.2: Pre-market belief report without redemption-value disclosure.

Public order book

Best-quote spread: —

Bids

Price in ECU	Quantity
--------------	----------

Asks

Price in ECU	Quantity
--------------	----------

Submit order

Side

Type

Price in ECU

Quantity

Estimated preview: 100 ECU

Your pending orders

No available orders

Recent executions

Price in ECU	Quantity
100	1
100	1

At closing, all pending orders are cancelled. Your result is final cash plus final units times redemption value, plus your target bonus. Finishing cancels your own orders and ends your participation in this period; others may continue until finished or time expires.

Figure 2.3: Order book and submission during actual trading.

Comparisons in the extended edition

3.1 A common reference for fourteen conditions

The fourteen conditions are session options rather than fourteen stages that every student must complete. All retain eight participants, one trading period and the initial measurement block. The reference establishes what changes in each comparison. Conditions that alter information or incentives preserve the remaining economic rules. Instructors can assign complete markets to different conditions and compare outcomes. The assignment unit matters: eight personal observations inside a single market do not amount to eight independent markets. With one class, results illustrate mechanisms and generate questions; they do not by themselves provide a precise estimate of treatment effects.

Prepare a prediction sheet before starting. For each comparison, record the modified variable, an expected outcome and an alternative explanation. Examine differences in volume, net gains and belief updating without turning a single number into a verdict. Groups may also differ in prior experience, skill and composition. Keeping identifiers anonymous and comparing distributions helps discuss heterogeneity without singling out students. The following sections describe the implemented options and the economic limits of each comparison. They provide questions for discussion rather than outcomes the software is programmed to produce.

3.2 Reference: measurement before trading

The reference condition provides calibration feedback after the first market. Everyone answers the eight questions and forecasts relative position before receiving performance results. Signal accuracy is 75 percent. There is a one-unit inventory target and a 50 ECU bonus for meeting it. Trading carries no commission, and each person trades their own wealth. Volume may reflect both information and the announced inventory target. This combination demonstrates why an economically understandable transaction cannot by itself establish overconfidence.

Ask students to predict whether participants with a larger confidence gap will trade more. Compare that prediction with the observed dispersion, zero commissions and inventory needs. Any association remains descriptive. A person with a larger gap may also have lower accuracy or a buying target that finds more willing counterparties. The reference helps organize these explanations and supplies a comparison for a specific intervention. It does not offer a universal definition of rational behavior in real financial markets. Discuss which features of the classroom setting make an observed action understandable.

3.3 Early accuracy feedback

This condition reveals the number of correct answers before the market. It keeps calibration calculations and relative-position results private until later feedback. The change allows instructors to study how news about absolute performance affects beliefs or willingness to trade. Six correct answers out of eight do not directly reveal how many peers were outperformed. Nor do they reconstruct a complete assessment of probability reporting unless the student remembers all eight confidence statements.

Ask for predictions about the reactions of people with high and low scores. Increased activity after a good result could reflect learning about skill, enthusiasm or a perceived information advantage. Distinguishing those explanations requires additional measures and enough groups. Compare the condition with the reference while maintaining signal accuracy and the inventory objective. Do not present the comparison as automatic proof that confidence became accurate. Performance feedback provides potentially useful information, but the way a person interprets and applies that information remains a decision. Its consequences for profits must be examined separately.

3.4 Early calibration feedback

Before trading, this condition displays correct answers, observed accuracy, mean confidence and their difference in percentage points. A student may discover that 85 percent confidence accompanied lower accuracy. The private message uses descriptive language. It does not classify the person as overconfident, incompetent or prudent. A negative gap also requires interpretation: it may reflect caution, noise from a small sample or questions that were difficult for that participant.

The intervention asks whether an explicit evaluation of probability reporting changes the initial asset belief, subsequent updating or trading frequency. A fall in volume need not be an improvement: it might prevent completion of an inventory objective or a favorable transaction. Examine belief rewards, net gains and target completion together. The condition changes the timing and content of performance information. It neither prescribes a trading strategy nor removes uncertainty. In discussion, distinguish noticing a gap from deciding how much weight that short measurement deserves in a different financial task.

3.5 Private relative-position feedback

Relative-position feedback shows the correct-answer count, the predicted percentage of peers outperformed and the percentage actually outperformed, counting each tie as half a peer. All forecasts are locked before information is revealed. The ranking uses the initial group of eight, even if a subsequent repeated session reassigns groups. Participants do not see names or other people's answers. This information does not change previously recorded responses or the initial calibration payment.

Compare favorable news about rank with favorable news about accuracy. Someone can answer few questions correctly while outperforming most peers, or answer many correctly while falling below the group. That distinction separates absolute skill from social comparison. Do not attribute a market reaction to pride or anxiety without measuring those responses. The intervention supplies a specific social reference, while its consequences for trading remain empirical. The aggregate report examines expected versus observed rank without publishing a list of students. It also retains the denominator of seven other people for each individual comparison.

3.6 A one-ECU commission per unit and side

This variant charges one ECU to each participant for every unit actually bought or sold. A one-unit transaction therefore generates two ECU in total costs, one for each side. Submitting or cancelling an unexecuted order does not incur this commission. Order-book reservations include enough resources to cover the buyer's cost, preventing an order from committing cash that is unavailable. The transaction price retains its meaning because commission is recorded separately.

The question is whether a small cost reduces transactions and whether that reduction changes net gains. A buyer may earn a twenty-ECU gross gain and retain nineteen after the cost. Lower activity may improve outcomes when it removes trades with small benefits, but it can also make an inventory target harder to reach. Compare gross and net gains, volume and bonuses. Gross gain minus net gain must equal commissions. This accounting identity gives students a practical way to check their interpretation rather than relying only on the direction of a treatment average.

3.7 A five-ECU commission per unit and side

The high commission raises the same cost to five ECU. It introduces neither a price-proportional tax nor a hidden spread: the charge remains an amount per executed unit and participant. A two-unit transaction costs ten ECU on each side. The structure compares a larger friction with the one-ECU charge and with zero commission while keeping the asset and priority rule constant.

Ask students to identify transactions that are profitable before costs but cease to be profitable after costs. A price close to redemption value can produce a small gross gain and a net loss. However, the inventory bonus may still justify a transaction that loses money when considered alone. Analysis must distinguish the investment account from externally provided incentives. If no orders cross, inspect the available quotes and unfulfilled targets as well. No transactions does not mean no decisions or disagreement. Orders that remain unexecuted reveal willingness to transact at particular prices rather than completed transfers of resources.

3.8 A low-accuracy private signal

The low-accuracy signal is correct with probability 60 percent. With equal prior probabilities and a symmetric signal, a high message implies a 60 percent posterior probability that value is 120. A low message implies 40 percent. The signal remains informative; it is not a completely irrelevant coin flip. Everyone knows this quality before reporting the initial belief. Conditional on redemption value, private draws are independent across participants.

The comparison asks whether people respond to announced information quality or report the same degree of certainty for any message. Compare initial beliefs with the appropriate Bayesian reference. Do not impose that initial reference on the post-market belief, which may incorporate prices and other people's actions. A market with less accurate signals could display caution, disagreement or more intensive search for information in the book. None of these responses is programmed. Record whether students understand the distinction between objective signal accuracy and subjective confidence, since confusing those two concepts can affect both the initial report and the classroom discussion.

3.9 A high-accuracy private signal

High signal accuracy is 90 percent. With the same prior and symmetry, initial probabilities of high value are 90 or 10 percent depending on the message. The signal does not reveal redemption value with certainty. Several people may receive incorrect information because their private draws are independent conditional on value. That possibility is not a programming error and does not justify changing the records after observing an unfavorable realization.

More accurate information might increase certainty and reduce some disagreement, but strongly held opposing positions can remain when messages differ. Compare belief changes and trading with the 60 and 75 percent conditions. Explain why a 100 percent belief requires more than a 90 percent signal. Belief scoring illustrates the cost of declaring certainty while an opposing outcome remains possible. It also distinguishes an unlucky realization from a poorly stated ex ante probability. A correct assessment can lose a particular wager, just as a poorly assessed probability can be followed by a favorable realization.

3.10 A common public signal

This condition makes one informative draw and gives the same signal to all eight people. Accuracy remains 75 percent. These are not eight independent signals that happen to agree: everyone knows they share one source. Students should therefore neither multiply the information eight times nor treat agreement as additional independent evidence. The difference from private information concerns both the information structure and what each person knows about the others.

If trading continues, examine inventory objectives, disagreement about interpreting the message and expectations about prices. Common information need not eliminate transactions because liquidity motives and freedom to quote prices remain. The condition helps show that market prices can reflect constraints and incentives as well as information about value. Do not automatically interpret lower volume as the disappearance of mistakes. First identify the opportunities and objectives faced by each side. A common announcement changes the grounds for disagreement but does not itself require everyone to submit the same order or to abstain from trading.

3.11 Bayesian assistance before trading

The assistance condition displays the probability calculated from the announced prior and known signal accuracy beside the personal message. It does not change the signal or select an answer on the student's behalf. Each person can accept or depart from the displayed reference when entering their own probability. The intervention reduces the computational burden of a simple initial update and asks whether distance from that reference becomes smaller.

The number is neither a recommendation to buy nor a compulsory limit price. It also excludes information generated during trading. After observing prices, orders or a lack of counterparties, a participant may have reasons to revise the initial belief. Compare belief changes with volume and gains. Better probability understanding does not guarantee a larger trading profit in every realization. This separation is useful for financial education: making one calculation easier can improve a particular task without solving every decision problem under uncertainty. Ask which information the displayed reference includes before using it as a standard for evaluation.

3.12 No liquidity objective

This variant sets both the inventory-change objective and its associated bonus to zero. It removes that external reason to buy or sell. Students retain their initial five units, cash and private signal. Trading may reflect information, price expectations, learning from the order book or experimentation. The software does not ask for a psychological explanation, and instructors should not invent one from volume alone.

Compare this condition with the reference before declaring that a difference represents excessive speculation. Legitimate opportunities and incentives have changed: a transaction that previously earned an additional fifty ECU no longer carries that component. Distinguish submitting an order from finding a counterparty willing to execute it. A group can produce many proposals and few completed trades because nobody accepts the offered terms. The history preserves orders, cancellations and executions. These traces describe how an outcome arose without reducing all activity to one number. They also show why an absence of trades can coexist with substantial participation in the market.

3.13 A two-unit liquidity objective

The high objective requires a net purchase or sale of two units to earn the fifty-ECU bonus. Its direction remains individual. The condition does not force a trade or alter asset redemption value. Someone who makes two purchases and two sales ends with no net inventory change. That turnover does not fulfill a two-unit net buying objective. The relevant measure compares final inventory with the initial endowment.

The variant examines pressure to complete a need and willingness to accept less favorable prices. A larger target could raise volume, but it may remain unmet when counterparties or resources are unavailable. Compare target completion, bonuses, net gains and turnover. The target reward is presented separately from financial returns because it comes from an incentive external to exchange. This distinction prevents crediting the market with wealth creation that actually comes from a teaching-design transfer. Ask whether a participant's final payment would be interpreted differently if the bonus were combined invisibly with portfolio gains.

3.14 Delegated management with linear compensation

In linear management, the participant trades a portfolio for an automated owner. Compensation is one hundred ECU plus twenty percent of portfolio wealth after commissions and the inventory bonus. The owner provides a separate one-hundred-ECU salary reserve. It cannot be used for trading. Owner residual equals portfolio wealth plus bonus and salary reserve, minus manager compensation. The combined outcomes conserve those resources.

Compare the condition with own-account trading by identifying the fraction of an improvement that reaches the decision maker. The manager does not own all final portfolio wealth. The rule changes the scale and slope of incentives and should not be described as an isolated change in risk preferences. Calibration and belief scores remain separate personal payment components. The report distinguishes the portfolio's financial result from participant compensation so that different quantities are not compared as if they were identical. Before interpreting behavior, ask students to calculate how an additional ECU of portfolio wealth affects each party under the announced contract.

3.15 Delegated management with convex compensation

Convex management pays one hundred ECU plus twenty percent of the positive part of portfolio wealth and bonus above benchmark wealth. If the portfolio fails to exceed that threshold, variable compensation is zero. The external salary reserve remains one hundred ECU, and the owner's residual satisfies the same accounting identity. The threshold uses the actual redemption value of the initial endowment, evaluating performance against simply retaining initial cash and units.

The option-shaped rule supports discussion of participation in gains and limited exposure to negative variable compensation. However, linear and convex contracts also differ in expected levels: they were not actuarially equalized. A behavioral difference therefore does not identify a pure curvature effect. Use the comparison to describe incentives and formulate hypotheses about trading, owner outcomes and compensation distributions. The preceding worked ledger explains who receives each ECU before students interpret class results. The contract changes rewards for decisions; it does not change the physical number of asset units or turn the automated owner into another strategic market participant.

- Value is 80 or 120 ECU with equal probabilities. The signal matches the true value in 75% of cases. Your previous answers do not change its quality.
- People's signals are drawn independently conditional on the common value. You see only your own signal.
- Fee per executed unit: resting order 0.0 ECU; incoming order 0.0 ECU. A negative fee is a rebate from the external exchange operator.
- You manage an automated owner's portfolio. Your income is 100 ECU plus 20% of positive excess of final wealth including the bonus over initial buy-and-hold wealth at redemption. The owner has 100 ECU set aside for wages.
- Average reward for the two value estimates is added to portfolio income or management pay each period. Eight-question and relative-forecast rewards are added only in period 1. Trading funds reset; confirmed rewards accumulate.
- Information about your eight answers and relative position is revealed after the first market, except for any earlier feedback specified in this case.

Figure 3.1: Convex management contract: distinguish the manager and owner accounts.

Classroom use, data interpretation and exercises

4.1 A seventy-five-minute session

Reserve ten minutes to introduce the economic question, define ECU as experimental currency units and organize complete groups of eight. Demonstrate the distribution of cash and units, the meaning of a limit price and the reservation of resources. Do not reveal calibration answers or suggest a desirable confidence level. Request written predictions about a specific comparison. A first class can use one market per group, comparing the reference with one information or cost intervention if enough participants are available.

Allow about fifteen minutes for the operational rules, instructions and comprehension questions. There is no demographic questionnaire. An incorrect response retains entered values and produces question-specific guidance; students must correct it. Answer procedural questions without recommending a particular trade. Check that everyone uses the intended language and that personal links are not shared. Observers can complete a prediction worksheet, but should not occupy a fictitious participant place or enter decisions on someone else's behalf. Explain that the purpose of comprehension checks is to establish common understanding of resources and incentives.

The next fifteen minutes cover individual questions, relative-position forecasts and initial beliefs. A waiting screen at this stage means that some group decisions are still missing. Do not close the window or switch sessions to solve the wait. Inspect administrative progress and help locate the pending form. Performance news appears only at the time prescribed by the condition. Avoid public explanations that accidentally turn private feedback into a common signal. Once probabilities have been recorded, they must remain the participant's original reports rather than answers corrected to match an instructor's expectations.

Reserve five minutes for the market and the subsequent belief report. Announced trading time is three minutes in one-period sessions. Instructors can extend the initial explanation but should not selectively stop a market's clock according to its outcomes. Voluntarily ending trading does not reveal redemption value to other participants. Everyone must reach closure and submit the final belief before payments are interpreted. Market records remain available when someone takes longer to complete the personal evaluation. This separation preserves the order of information rather than forcing belief reports after students have already learned the answer.

Use the remaining thirty minutes to reveal the report, reconstruct a trade and discuss predictions. A ten-period session requires a longer class or less discussion according to an announced plan. The calibration block occurs once, while financial endowments reset every period. Experience can change during the battery even though the initial calibration measure stays fixed. Do not interpret every period as a fresh independent measurement of confidence. Repeated interaction also changes

what students can learn about the book and their counterparties, which should be part of the subsequent discussion.

Feedback script

Before displaying averages, ask why two people with the same signal might trade. Then reveal financial gains, bonuses and probability scores separately. Conclude the discussion by identifying which observations support an interpretation and which remain compatible with other explanations.

4.2 Conducting the discussion after decisions

Begin with verifiable facts: how many people finished, how many trades executed and what costs were paid. Return to the written predictions. A refuted expectation is not a failure of the activity; it helps identify a mistaken assumption or an overlooked force. Ask for an explanation based on resources, information and rewards before introducing behavioral labels. Someone who buys after a low signal may be seeking a bonus, anticipating a later resale or judging the offered price sufficiently low. Those possibilities require examining the actual decision setting.

Separate ex ante evaluation from ex post outcomes. A buyer who earns money when value is high has not thereby demonstrated a well-calibrated belief. The same probability report can be reasonable and receive a lower payment in a particular realization. Examine sets of questions and groups while keeping sampling uncertainty explicit. To discuss learning, compare available information and decisions over time. To discuss causality, consider how conditions were assigned and how many independent markets support the contrast. With only a few markets, descriptive patterns and carefully reconstructed examples usually communicate more than strong statistical claims.

4.3 The private report: observations and denominators

The report does not display aggregate results when first opened. The instructor decides when to reveal them and can hide them again. Its header shows participants, people with recorded calibration, closed markets and pending evaluations. These quantities describe progress and should not be confused. Completed calibration can exist before the first trade. A closed market can still await some post-market probability reports. Financial summaries use available closed observations, while subsequent belief scores require a finalized evaluation. Check these states before interpreting a missing value as a zero result.

Calibration summaries use one initial measurement per person. The reliability display groups answers into confidence intervals 50–59, 60–69, 70–79, 80–89 and 90–100 percent. Every row shows the number of answers, mean reported confidence and observed accuracy. The diagonal represents equality between the two measures. Not every interval contains observations; an empty cell does not mean zero correct answers. Eight participants generate sixty-four answers, but remain eight people answering several related questions. A visually impressive number of answers therefore does not create an equally large sample of independent decision makers.

The gap histogram uses one observation per person rather than one per question. An average gap can hide offsets between items with excessive and insufficient certainty. The Brier score uses squared probability errors and distinguishes a wrong answer stated with extreme confidence from a wrong answer stated with moderate confidence. Neither a histogram nor a short score diagnoses a stable personality trait. These measures teach what information is required when linking beliefs

with financial decisions. Interpret them together with item difficulty, accuracy and the limited size of the initial instrument.

The financial table distinguishes gross gains, net gains, commissions, units, turnover and payment. Personal turnover divides the monetary value of that person's trades by the initial endowment valued at redemption. One transaction adds one unit to market volume and one unit to each of its two participants. Summed personal units therefore double market volume when all counterparties are included. The period table preserves this distinction and allows instructors to examine evolution without incorrectly combining transactions and sides. The target bonus is reported separately from asset returns because it is funded outside the exchange itself.

The association between the initial gap and trading uses each person's average turnover across closed periods. A ten-period battery consequently does not multiply the same initial measurement into ten independent observations. Correlation is omitted when fewer than two people or insufficient variation are available. A dash means the coefficient cannot be calculated; it is not a zero correlation. Under delegated management, also examine owner residual and distinguish portfolio wealth from personal compensation. The same portfolio result can imply different participant payments under different announced contracts.

Instructor report: calibration and trading

Reference with feedback after market

Participants: 8 · Participants with recorded calibration: 8 · Closed markets: 1/1 · Pending period evaluations: 0

Download this session, all rounds (CSV)

Download this session in wide format

Reveal aggregates

Figure 4.1: The instructor controls when aggregates are revealed.

4.4 Downloads and reconstruction of the record

The custom download contains only the selected session and all ten technical rounds, with states distinguishing pending, active, completed and inactive records. The native wide oTree export retains round-specific columns and is also limited to that session. Use the two files for different tasks: the custom file supports reconstruction of economic decisions and events, while the native file preserves oTree's standard administrative structure. Do not combine another session simply because it happens to use similar internal identifiers. The selected session is the relevant scope for both buttons.

A one-period session leaves the later technical rounds inactive. Those records do not represent dropouts or observed zero outcomes. A field still pending in a repeated session is not a zero either. First filter by record type, status, period and availability of evaluation. Calibration and relative forecasts are initial measurements; do not fill later rounds by copying the same value to inflate sample size. When reconstructing payments, preserve personal and financial components separately until the final sum. A participant-period payment must not be mistaken for only the return on the asset portfolio.

Market events record proposals, rejections, cancellations and executions. An order that does not cross the book may remain standing. Cancelling it releases the reserved resources. Price times quantity defines the transaction amount, and commission is added separately on each side. Check final balances using initial cash, purchase payments, sale receipts and commissions. Multiply final units by redemption value to obtain wealth. Then subtract benchmark wealth and add bonuses and scores only where appropriate. This sequence permits an independent accounting check without rerunning the completed session or generating a new set of market outcomes.



Period 1/1

Decisions, estimates and trading

Period result

Known redemption value	80.0
Wealth before bonus	880.0
Initial wealth valued at redemption	900.0
Trading gain before fees	-20.0
Fees actually paid	0.0
Trading gain after fees	-20.0
Bonus for meeting the target	50.0
Units traded by the person	1
Traded notional / initial redemption wealth	0.1111111111111111
Portfolio or management-contract income	930.0
Average value-estimate reward	50.75
Eight-answer reward (period 1 only)	97.75
Relative-position reward (period 1 only)	99.27
Total period payoff	1177,77
Probability of value 120 before market (%): 75; Probability of value 120 after market (%): 65	

Figure 4.2: Individual financial result: portfolio, costs and rewards separated.

Instructor report: calibration and trading

Reference with feedback after market

Participants: 8 · Participants with recorded calibration: 8 · Closed markets: 1/1 · Pending period evaluations: 0

[Download this session, all rounds \(CSV\)](#)

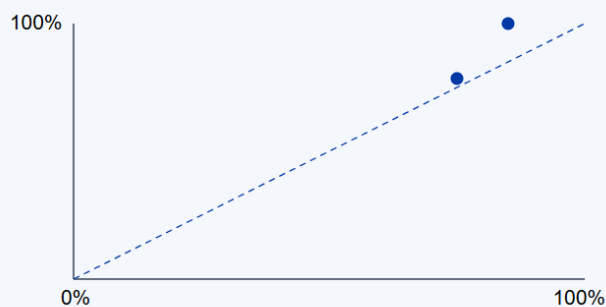
[Download this session in wide format](#)

[Hide aggregates](#)

Prior calibration: one observation per person

Observed accuracy (%)	81,25	N=8
Mean reported probability (%)	76,25	N=8
Mean probability minus accuracy (percentage points)	-5,0	N=8
Probability mean squared error (0–1)	0,1512	N=8
Estimated relative position (%)	70	N=8
Observed relative position with half-weight ties (%)	50,0	N=8
Estimated minus observed relative position (percentage points)	20,0	N=8

Reported probability and correct-answer frequency



Reported probability interval (%) → X; Observed accuracy (%) → Y

Reported probability interval (%)	Answers	Mean reported probability (%)	Observed accuracy (%)
50–59	0	—	—
60–69	0	—	—
70–79	56	75	78,57
80–89	8	85	100
90–100	0	—	—

Figure 4.3: Revealed private report: summary and reliability with denominators.



Period 1/1

Decisions, estimates and trading

Activity completed

Sum of active-period results: **1177.77 ECU**.

Your decisions have been recorded. Please wait for your instructor's directions.

Figure 4.4: Explicit ending: accumulated payment from active periods.

Keep an original data copy and a transformation log. Document filters and denominators before drawing a relationship between confidence and activity. When comparing conditions, identify the market as the interaction unit and avoid treating its eight participants as independently assigned treatments. In a small class, tables, distributions and accounting examples are often clearer than a regression with strong interpretations. Technical codes help connect rows, but student presentations should use descriptive names. Never distribute personal participant links or administrative credentials as part of an analytical worksheet or a public teaching example.

4.5 Exercises with solutions

4.5.1 Answers to the initial measurement

The correct answers, in order, are A, B, A, B, A, B, A and B. Three quarters exceeds two thirds. Increasing by ten percent and then decreasing by ten percent takes one hundred to ninety-nine, below one hundred. Equally likely values of eighty and one hundred twenty have mean one hundred. Recovering a twenty-percent decline requires a twenty-five-percent increase from the new base. Two independent events with 75 percent success probability both occur with probability 56.25 percent. Buying at ninety and selling at one hundred with a two-ECU fee on each side leaves six. A low signal with 75 percent accuracy makes high value 25 percent likely. A portfolio split equally between a fixed payment of one hundred and the eighty/one-hundred-twenty asset pays ninety or one hundred ten.

Reserve these solutions until all relevant initial measurements are recorded. Do not change answer keys according to the market outcome. The questions measure objective reasoning in this particular instrument rather than providing an exhaustive financial-literacy test. Confidence refers to the chosen alternative: selecting B with 80 percent confidence means an 80 percent probability that B is correct, not an 80 percent probability that the answer is A. This distinction is essential when reconstructing correctness and probability scores from the recorded choices.

4.5.2 Equal accuracy, different probability quality

Two people answer four of eight questions correctly. The first reports 50 percent confidence throughout; the second reports 100 percent throughout. Calculate their scores. For the first person, every squared error is 0.25 and the reward is 75 ECU. For the second, four errors are zero and four are one: the average is 0.5 and the reward 50. Both have 50 percent accuracy, but the second expressed certainty in four incorrect answers. This demonstrates why the correct-answer count and probability quality are different variables. Compare the squared errors before discussing which person's report provided a more informative representation of uncertainty.

4.5.3 Ties in relative position

If everyone answers exactly six questions correctly, nobody strictly outperforms another person and everyone ties with seven peers. Each tie counts as half, so observed position is 3.5 divided by 7, or 50 percent. A forecast of 50 earns one hundred ECU; a forecast of 100 earns 75. In every complete group, mean observed relative position is 50 percent because each bilateral comparison distributes one unit between its two members. Individual positions need not be equal when correct-answer counts differ. This identity checks the rank calculation without requiring a particular distribution of skills or forecasts.

4.5.4 A cost and a target in one transaction

A participant buys one unit at 119 and redemption value is 120. With a five-ECU commission, gross gain is one and net gain is minus four. If the purchase completes the inventory target and generates a fifty-ECU bonus, the change in wealth plus bonus is forty-six. The transaction loses money in the financial account and improves the combined component. Identify both outcomes before judging the action. If the bonus was already satisfied, another transaction does not earn it again. A discussion that combines the bonus invisibly with portfolio returns would miss the source of the apparent improvement.

4.5.5 Belief revision before value is revealed

A person receives a high signal with 75 percent accuracy and reports 75 before trading. After observing the book, they report 60. Moving away from 75 does not by itself establish an incorrect update because the market may have supplied information. Ask what was observed and what inference was made. When redemption value is low, the scores are 43.75 and 64 respectively, averaging 53.875. This realization favors the second report ex post, but does not prove that 60 was the optimal probability at the decision time. A proper scoring rule evaluates probability statements through expected rewards; it does not erase realization risk in an individual outcome.

4.5.6 Ten periods and one initial measurement

Suppose the calibration reward is 81 and the relative-position reward 99.82. Every period provides one thousand ECU in financial payment and eighty in average belief scoring. Total payment is ten times 1080 plus 180.82, or 10980.82 ECU. It must not be ten times 1260.82. The two initial rewards are paid once, while financial and belief components apply each period. Summing round payments verifies this identity in the export without rerunning the session. If only one period was active, inactive technical rounds contribute no payment and should not enter an average as observed zero outcomes.

4.6 Operational problems and limits of use

If a waiting page does not advance, first inspect group size and each participant's current page. Sessions require multiples of eight; random reassignment requires at least sixteen. Someone who has not completed the initial rank forecast can prevent feedback from being released. Do not replace their choice with an invented number. If trading has closed but a final payment is missing, inspect the post-market belief. Financial closure and personal evaluation are distinct steps, specifically to avoid revealing value before the last probability report. Administrative progress should identify which step remains incomplete.

Orders can be rejected because of price, quantity, insufficient cash or unavailable units, including resources reserved by previous orders. Cancelling a standing proposal releases its resources. Explain the rejection before repeating an identical order. The local instance uses one server process to preserve live book state, with sessions prepared before trading begins. A public deployment using a different database or several processes requires the corresponding administrative verification. Local screenshots do not establish that a different production-server configuration has passed those checks. Existing sessions and databases should be preserved rather than reset as a routine response to a waiting screen.

The parameters are teaching adaptations. A binary asset, a small class, eight questions and a horizon measured in minutes simplify real markets. The activity supports observation of decisions,

incentives and accounting; it does not diagnose personal traits or recommend investments. Its teaching value depends on connecting a prediction with evidence and identifying what additional data would be needed. Maintain that discipline when presenting favorable and unfavorable outcomes alike. The aim is to interpret a financial decision using concrete information and resources, while recognizing the boundaries of what this classroom design can establish.

Bibliography

- Barber, B. M., and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, 55(2), 773–806. https://faculty.haas.berkeley.edu/odean/papers/returns/individual_investor_performance_final.pdf.
- Bloomfield, R., O’Hara, M., and Saar, G. (2005). The “make or take” decision in an electronic market: Evidence on the evolution of liquidity. *Journal of Financial Economics*, 75(1), 165–199. <https://doi.org/10.1016/j.jfineco.2004.07.001>.
- Glaser, M., and Weber, M. (2004). Overconfidence and trading volume. *Working Papers*, 105, University of Mannheim. <https://madoc.bib.uni-mannheim.de/275>.
- Gneiting, T., and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>.
- oTree. Live pages; administering an experiment; data export. Official documentation. <https://otree.readthedocs.io/en/master/>. Accessed October 17, 2026.