Cognitive complexity, expertise and prediction accuracy in venture selection

BSE Summer Forum

Economics of Science and Innovation - June 2026

Thomas Åstebro   Andrew Funck   Francesco Giordano

HEC Paris, Economics and Decisions Sciences

Motivation

- Selecting high-potential ventures is a **first-order allocation problem**
  - early-stage resources are scarce; screening decides which ideas scale
  - incubators, VCs and agencies act as **gatekeepers** of innovation

Motivation

- Selecting high-potential ventures is a **first-order allocation problem**
  - early-stage resources are scarce; screening decides which ideas scale
  - incubators, VCs and agencies act as **gatekeepers** of innovation
- Assessing venture quality is **hard**
  - sparse history; mostly weak, noisy cues
  - deep uncertainty about future value

Motivation

- Selecting high-potential ventures is a **first-order allocation problem**
    - early-stage resources are scarce; screening decides which ideas scale
    - incubators, VCs and agencies act as **gatekeepers** of innovation
- Assessing venture quality is **hard**
    - sparse history; mostly weak, noisy cues
    - deep uncertainty about future value
- A **Bayesian benchmark** defines good decisions
    - the best one can do given noisy evidence

Research Questions

What would **optimal** venture selection look like
and do **real judges** come close?

Research Questions

What would **optimal** venture selection look like
and do **real judges** come close?

1 **Benchmark.** How well does a normative Bayesian rule pick winners and how closely do judges' recommendations track it?

Research Questions

What would **optimal** venture selection look like
and do **real judges** come close?

1. **Benchmark.** How well does a normative Bayesian rule pick winners and how closely do judges' recommendations track it?

2. **Accuracy** When do expertise and complexity affect the accuracy of selecting high-quality ventures?

Research Questions

What would **optimal** venture selection look like
and do **real judges** come close?

1. **Benchmark.** How well does a normative Bayesian rule pick winners and how closely do judges' recommendations track it?

2. **Accuracy** When do expertise and complexity affect the accuracy of selecting high-quality ventures?

3. **Uncertainty.** How much does the evidence reduce uncertainty about venture quality and how does that depend on expertise and complexity?

Preview of Results

- We build and test a **Bayesian benchmark** for an early-stage incubator's selection problem.

Preview of Results

- We build and test a **Bayesian benchmark** for an early-stage incubator's selection problem.
- **Judges think Bayesian:** the benchmark matches $\sim$79% of recommendations, $\sim$85% when missing a good venture hurts more than admitting a bad one.

Preview of Results

- We build and test a **Bayesian benchmark** for an early-stage incubator's selection problem.
- **Judges think Bayesian:** the benchmark matches $\sim$79% of recommendations, $\sim$85% when missing a good venture hurts more than admitting a bad one.
- **But signal information content is limited:** judges and the optimum reach only $\sim$60% accuracy in discriminating ventures.

Preview of Results

- We build and test a **Bayesian benchmark** for an early-stage incubator's selection problem.
- **Judges think Bayesian:** the benchmark matches $\sim$79% of recommendations, $\sim$85% when missing a good venture hurts more than admitting a bad one.
- **But signal information content is limited:** judges and the optimum reach only $\sim$60% accuracy in discriminating ventures.
  - **Accuracy** $\downarrow$ **with complexity** and $\uparrow$ **with expertise** across judges, not within a judge over time.
  - Uncertainty is lower for experts, higher for complex ventures.
  - The **expertise premium** peaks at **intermediate complexity.**

Literature

1. **Cognitive complexity**: Enke and Graeber (2023), de Clippel et al. (2024)

Literature

1. **Cognitive complexity**: Enke and Graeber (2023), de Clippel et al. (2024)
2. **Judgments in entrepreneurial decision making**: Åstebro and Elhedhli (2006), Scott et al. (2020), McKenzie and Sansone (2019)

Literature

1. **Cognitive complexity**: Enke and Graeber (2023), de Clippel et al. (2024)
2. **Judgments in entrepreneurial decision making**: Åstebro and Elhedhli (2006), Scott et al. (2020), McKenzie and Sansone (2019)
3. **Bayesian decision models**: Grether (1980), Baley and Veldkamp (2023), Agrawal et al. (2025)

Outline

Context

Context

- The HEC incubator
  - 4 months startups' support program based in Paris
  - Since 2007, +900 ventures incubated, raised $\sim$ 1€ Bln

Context

- The HEC incubator
  - 4 months startups' support program based in Paris
  - Since 2007, +900 ventures incubated, raised $\sim$ 1€ Bln
- Each quarter admits $\approx$ 10 ventures out of 60
  - Ventures submit a textual application describing the company
  - A pool of judges randomly assigned evaluate each application on 14 business criteria and provide an admission recommendation

Context

- The HEC incubator
  - 4 months startups' support program based in Paris
  - Since 2007, +900 ventures incubated, raised $\sim$ 1€ Bln
- Each quarter admits $\approx$ 10 ventures out of 60
  - Ventures submit a textual application describing the company
  - A pool of judges randomly assigned evaluate each application on 14 business criteria and provide an admission recommendation

Data

- All eligible applications from 9 batches (fall 2021 - winter 2024)
- 1638 evaluations and 579 applications
- Economic performance (FTEs, total funding, survival status) of all applicants as of Summer 2024

▸ Evaluation criteria ▸ Ventures characteristics ▸ Ventures performance

Conceptual framework

Conceptual framework

A judge acts as a **Bayesian classifier** under noisy evidence

- **Task** Admit top-quality ventures, reject the rest.

Conceptual framework

A judge acts as a **Bayesian classifier** under noisy evidence

- **Task** Admit top-quality ventures, reject the rest.
- **Signal** Reading the application yields a noisy cognitive signal whose precision rises with expertise, falls with complexity.

Conceptual framework

A judge acts as a **Bayesian classifier** under noisy evidence

- **Task** Admit top-quality ventures, reject the rest.
- **Signal** Reading the application yields a noisy cognitive signal whose precision rises with expertise, falls with complexity.
- **Inference** Combine a prior on quality with the signal's likelihood to form a posterior on venture quality given the evidence.

Conceptual framework

A judge acts as a **Bayesian classifier** under noisy evidence

- **Task** Admit top-quality ventures, reject the rest.
- **Signal** Reading the application yields a noisy cognitive signal whose precision rises with expertise, falls with complexity.
- **Inference** Combine a prior on quality with the signal's likelihood to form a posterior on venture quality given the evidence.
- **Decision** Recommend $\iff$ the posterior clears a threshold.

▸ Theoretical model

Stylized representation

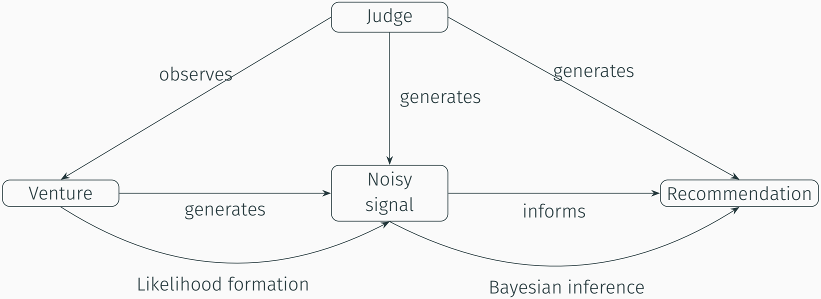

Figure 1: Conceptual model
Stylized representation of the model of inference and decision.

Key Constructs

Four measures of interest:

Key Constructs

Four measures of interest:

1. Expertise: accumulated experience in making judgments by a judge.

Key Constructs

Four measures of interest:

1. Expertise: accumulated experience in making judgments by a judge.
2. Cognitive complexity: the degree to which the decision environment is complex to assess.

Key Constructs

Four measures of interest:

1. Expertise: accumulated experience in making judgments by a judge.
2. Cognitive complexity: the degree to which the decision environment is complex to assess.
3. Accuracy: the percentage correct recommendations under an optimal decision rule.

Key Constructs

Four measures of interest:

1. Expertise: accumulated experience in making judgments by a judge.
2. Cognitive complexity: the degree to which the decision environment is complex to assess.
3. Accuracy: the percentage correct recommendations under an optimal decision rule.
4. Cognitive uncertainty: how much uncertainty about a venture's quality remains after observing the signal

Testable hypotheses

- Hypothesis 1a:
  Experience improves evaluation accuracy.
- Hypothesis 1b:
  Experience attenuates cognitive uncertainty.

Testable hypotheses

- Hypothesis 1a:
  Experience improves evaluation accuracy.
- Hypothesis 1b:
  Experience attenuates cognitive uncertainty.
- Hypothesis 2a:
  Venture complexity reduces evaluation accuracy.
- Hypothesis 2b:
  Venture complexity increases cognitive uncertainty.

Testable hypotheses

- **Hypothesis 1a:**
  Experience improves evaluation accuracy.
- **Hypothesis 1b:**
  Experience attenuates cognitive uncertainty.
- **Hypothesis 2a:**
  Venture complexity reduces evaluation accuracy.
- **Hypothesis 2b:**
  Venture complexity increases cognitive uncertainty.
- **Hypothesis 3:**
  The accuracy gap between experienced and non-experienced judges grows with complexity up to an intermediate level, then declines.

▸ Accuracy ▸ Uncertainty

Empirical strategy

Measurements I

Variable	Measurement
Continuous venture quality	Venture *i* average rank of FTE and cumulative funding, each residualized on batch fixed effects; set to 0 for ventures no longer operational in 2024
Binary venture quality	$\hat{v}_i = 1$ if continuous quality of venture *i* lies in the top 40%, 0 otherwise; the 40% threshold reflects the incubator's final-stage admission capacity

Measurements I

Variable	Measurement
Continuous venture quality	Venture i average rank of FTE and cumulative funding, each residualized on batch fixed effects; set to 0 for ventures no longer operational in 2024
Binary venture quality	$\hat{v}_i = 1$ if continuous quality of venture i lies in the top 40%, 0 otherwise; the 40% threshold reflects the incubator's final-stage admission capacity
Signal	$\hat{S}_{ij} \in \{14, \ldots, 42\}$, sum of the 14 criterion scores assigned by a judge j to venture i
Decision rule	$\hat{a}_{ij}(\hat{S})_{ij} \in \{0, 1\}$, binary recommendation of judge j to admit venture i (1 = recommend, 0 = do not recommend)

Measurements I

Variable	Measurement
Continuous venture quality	Venture i average rank of FTE and cumulative funding, each residualized on batch fixed effects; set to 0 for ventures no longer operational in 2024
Binary venture quality	$\hat{v}_i = 1$ if continuous quality of venture i lies in the top 40%, 0 otherwise; the 40% threshold reflects the incubator's final-stage admission capacity
Signal	$\hat{S}_{ij} \in \{14, \ldots, 42\}$, sum of the 14 criterion scores assigned by a judge j to venture i
Decision rule	$\hat{a}_{ij}(\hat{S})_{ij} \in \{0, 1\}$, binary recommendation of judge j to admit venture i (1 = recommend, 0 = do not recommend)
Expertise	$\hat{e}_{i,j} :=$ mumber of evaluations judge j has completed prior to evaluating venture i.
Complexity	$\hat{\sigma}_i \in \mathbb{R}_+$, two measures of evaluation difficulty: 1. *Disagreement:* normalized within-venture variance of residuals from $\hat{S}_{ij}$ on judge FE 2. *Text:* RIX readability index of the application text

Measurements II

Variable	Measurement
Accuracy	Proportion of judges' recommedations to admit top quality ventures and to reject low quality ventures over the total number of recommedations
Cognitive Uncertainty	Average variance of the distribution of venture quality conditional on signal $\mathbb{E}_{\hat{s}}\left[\text{Var}(\hat{v}\|\hat{s})\right]$

Judges' classification accuracy

How well do judges individually classify ventures?

- **Recommendations do track quality.** Recommended ventures survive more (+23%) and employ more (+57% FTEs) than rejected ones
  - but weakly linked to funding ▸ Outcomes

Judges' classification accuracy

How well do judges individually classify ventures?

- **Recommendations do track quality.** Recommended ventures survive more (+23%) and employ more (+57% FTEs) than rejected ones
    - but weakly linked to funding ▸ Outcomes
- **Yet accuracy is only 63%:**
    - better at rejecting weak ventures than spotting strong ones ▸ Confusion matrix
    - most errors cluster at **intermediate scores** ▸ Score & errors

Empirical specifications

Two blocks: estimate normative benchmark, then test the mechanisms

1. **Empirical Bayes classifier.** Identify top-quality ventures from the score
    - How accurate is the **best feasible rule** given the score, and how well does it replicate **judges' recommendations**?

Empirical specifications

Two blocks: estimate normative benchmark, then test the mechanisms

1. **Empirical Bayes classifier.** Identify top-quality ventures from the score
   - How accurate is the **best feasible rule** given the score, and how well does it replicate **judges' recommendations**?
2. **Hypothesis tests.** Reduced-form regressions of accuracy and uncertainty on
   - evaluator expertise, venture complexity, and their **interaction**

Results

Empirical Bayes classifier

Leave-one-out: train on all ventures but one, test on the held-out venture's evaluations

1. **Likelihood.** Estimate the score density by venture type, $\hat{f}_{\hat{S}|\hat{v}}$, via kernel density estimation (linear kernel, bandwidth 0.32)

Empirical Bayes classifier

Leave-one-out: train on all ventures but one, test on the held-out venture's evaluations

1. **Likelihood.** Estimate the score density by venture type, $\hat{f}_{\hat{S}|\hat{v}}$, via kernel density estimation (linear kernel, bandwidth 0.32)

2. **Posterior.** Apply Bayes' rule on the held-out evaluation

$$\hat{\eta}(\hat{s}) = \Pr(\hat{v} = 1 \mid \hat{S} = \hat{s})$$

with prior = empirical share of top-quality ventures

Empirical Bayes classifier

Leave-one-out: train on all ventures but one, test on the held-out venture's evaluations

1. **Likelihood.** Estimate the score density by venture type, $\hat{f}_{\hat{S}|\hat{v}}$, via kernel density estimation (linear kernel, bandwidth 0.32)

2. **Posterior.** Apply Bayes' rule on the held-out evaluation

$$\hat{\eta}(\hat{s}) = \Pr(\hat{v} = 1 \mid \hat{S} = \hat{s})$$

   with prior = empirical share of top-quality ventures

3. **Classify.** Label *top quality* iff $\hat{\eta}(\hat{s}) > \frac{1}{2}$

Bayes classifier accuracy

The benchmark hits a low ceiling

- Classifies ventures correctly 60% of the time
  - the score weakly separates quality, yet probabilities are well calibrated
  - limit is the **signal's informativeness**, not a misspecified posterior

▸ Diagnostics

Bayes classifier accuracy

The benchmark hits a low ceiling

- Classifies ventures correctly 60% of the time
  - the score weakly separates quality, yet probabilities are well calibrated
  - limit is the **signal's informativeness**, not a misspecified posterior
    ▶ Diagnostics
- Matches judges' recommendations 79% of the time
  - 85% once missing a good venture costs more than admitting a bad one ▶ Visualization

Simulating the treatment effect

What if admission itself boosts ventures biasing our quality labels?

- **The test.** Give non-admitted ventures a counterfactual boost — survival, FTEs, funding — as if they had been admitted, across a range of plausible effect sizes

Simulating the treatment effect

What if admission itself boosts ventures biasing our quality labels?

- **The test.** Give non-admitted ventures a counterfactual boost — survival, FTEs, funding — as if they had been admitted, across a range of plausible effect sizes
- **Labels barely move.** Rankings shift, but reclassification as top quality stays at just 3–7%

Simulating the treatment effect

What if admission itself boosts ventures biasing our quality labels?

- **The test.** Give non-admitted ventures a counterfactual boost — survival, FTEs, funding — as if they had been admitted, across a range of plausible effect sizes
- **Labels barely move.** Rankings shift, but reclassification as top quality stays at just 3–7%
- **Accuracy is stable** across all scenarios ▸ Accuracy simulations

⇒ Quality labels are not an artifact of the incubator's treatment effect

Experience and Accuracy

$$\text{Accuracy}_{i,j} = \alpha + \beta_1 \left(\log(\hat{e}_{i,j}) - \overline{\log(\hat{e}_j)} \right) + \beta_2 \overline{\log(\hat{e}_j)} + \beta_3 \hat{s}_{i,j} + \theta_i + \delta_{t(i)} + \gamma_j + \varepsilon_{i,j}$$

Dependent Variable:			Accuracy		
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
$Log(\hat{e}_{i,j}) - \overline{Log(\hat{e}_j)}$	-6.933**	-7.000**	-5.071*	-9.130**	-12.13***
	(2.752)	(2.740)	(2.993)	(3.550)	(3.859)
$\overline{Log(\hat{e}_j)}$	5.666*	5.624*	1.604	4.398	
	(3.159)	(3.177)	(3.871)	(3.052)	
Fixed-effects					
Batch (9)			Yes	Yes	Yes
Venture (579)				Yes	Yes
Evaluator (72)					Yes
Control: $\hat{s}_{i,j}$		Yes	Yes	Yes	Yes
Fit statistics					
Observations	1,638	1,638	1,638	1,638	1,638

Table 1: Accuracy and Judges' Experience

Complexity and Accuracy

$$\text{Accuracy}_{i,j} = \alpha + \sum_{f=2}^{4} \beta_f \hat{\sigma}_i^f + \hat{s}_{i,j} + \delta_{t(i)} + \gamma_j + \varepsilon_{i,j}$$

Dependent Variable:	Accuracy							
	Disagreement				Text-based			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
Complexity_i, *Q2*	-3.185	-3.238	-1.954	-2.218	-7.585*	-6.505	-5.841	-5.806
	(3.049)	(3.038)	(2.837)	(3.123)	(4.278)	(4.137)	(4.043)	(4.173)
Complexity_i, *Q3*	-10.63***	-10.87***	-7.724**	-7.783***	-10.87***	-9.839**	-8.036**	-7.544**
	(3.129)	(3.153)	(2.989)	(2.908)	(3.827)	(3.720)	(3.544)	(3.419)
Complexity_i, *Q4*	-9.255***	-10.94***	-8.098**	-8.747***	-6.436	-5.539	-5.715	-4.687
	(3.391)	(3.457)	(3.242)	(3.254)	(4.121)	(3.969)	(3.827)	(3.883)
Fixed-effects								
Batch (9)			Yes	Yes			Yes	Yes
Evaluator (72)				Yes				Yes
Controls: $\hat{s}_{i,j}$		Yes	Yes	Yes		Yes	Yes	Yes
Fit statistics								
Accuracy Complexity_i, *Q1*	69.21%	69.21%	69.21%	69.21%	69.34%	69.34%	69.34%	69.34%
Observations	1,638	1,638	1,638	1,638	1,638	1,638	1,638	1,638

Table 2: Accuracy and Ventures' Complexity

Accuracy experience premium by complexity class

$$\text{Accuracy}_{i,j} = \alpha + \beta_1 \text{High Experience}_j + \sum_{f=2}^{4} \beta_f \hat{\sigma}_i^f$$

$$+ \sum_{f=2}^{4} \gamma_f \text{High Experience}_j \hat{\sigma}_i^f + \hat{s}_{i,j} + \delta_{t(i)} + \varepsilon_{i,j}$$

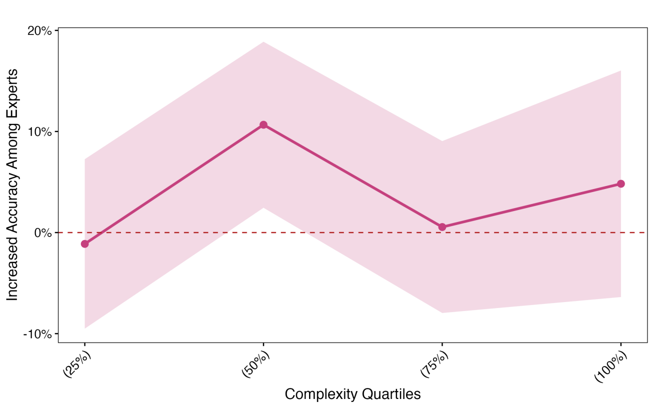

Figure 2: Accuracy Gain for Experienced Judges by Text Complexity

▸ Regression table

Measuring cognitive uncertainty

Residual uncertainty about quality, conditional on the score

1. **Subgroup classifiers.** Re-estimate the empirical Bayes classifier separately on median-split subsamples — experts/non-experts and complex/simple ventures
   - lets the posterior belief $\hat{\eta}(\hat{s})$ vary with expertise and complexity

Measuring cognitive uncertainty

Residual uncertainty about quality, conditional on the score

1. **Subgroup classifiers.** Re-estimate the empirical Bayes classifier separately on median-split subsamples — experts/non-experts and complex/simple ventures
   - lets the posterior belief $\hat{\eta}(\hat{s})$ vary with expertise and complexity
2. **Per-evaluation posterior variance.** For each (i, j):

$$V\big(\hat{v}_i \mid \hat{s}_{i,j}\big) = \hat{\eta}(\hat{s}_{i,j})\big(1 - \hat{\eta}(\hat{s}_{i,j})\big)$$

   - higher value $\Rightarrow$ score is **less informative** about venture quality

Measuring cognitive uncertainty

Residual uncertainty about quality, conditional on the score

1 **Subgroup classifiers.** Re-estimate the empirical Bayes classifier separately on median-split subsamples — experts/non-experts and complex/simple ventures

   - lets the posterior belief $\hat{\eta}(\hat{s})$ vary with expertise and complexity

2 **Per-evaluation posterior variance.** For each (i, j):

$$V\big(\hat{v}_i \mid \hat{s}_{i,j}\big) = \hat{\eta}(\hat{s}_{i,j})\big(1 - \hat{\eta}(\hat{s}_{i,j})\big)$$

   - higher value $\Rightarrow$ score is **less informative** about venture quality

3 **Regress** this variance on expertise and complexity

Uncertainty and Experience

$$V(\hat{v}_j \mid \hat{s}_{i,j}) = \alpha + \beta_1 \text{High Experience}_j + \beta_2 \hat{s}_{i,j} + \beta_3 \hat{s}^2_{i,j} + \delta_{t(i)} + \theta_i + \varepsilon_{i,j}$$

Dependent Variable:	Uncertainty			
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	0.2360***	0.2349***	0.2417***	
	(0.0012)	(0.0008)	(0.0007)	
High Experience$_j$	-0.0192***	-0.0169***	-0.0170***	-0.0174***
	(0.0027)	(0.0010)	(0.0007)	(0.0011)
Fixed-effects				
Batch (9)				Yes
Venture (579)				Yes
Controls: $\hat{s}_{i,j}$		Yes	Yes	Yes
Controls: $\hat{s}^2_{i,j}$			Yes	Yes
Fit statistics				
Observations	1,638	1,638	1,638	1,638

Table 3: Cognitive Uncertainty and Judges' Experience

Uncertainty and Complexity

$$V(\hat{v}_j \mid \hat{s}_{i,j}) = \alpha + \beta_1 \text{High Complexity}_i + \beta_2 \hat{s}_{i,j} + \beta_3 \hat{s}^2_{i,j} + \delta_{t(i)} + \gamma_j + \varepsilon_{i,j}$$

Dependent Variable:	Uncertainty							
	Disagreement				Text-based			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
Constant	0.2120***	0.2101***	0.2182***		0.2211***	0.2220***	0.2331***	
	(0.0039)	(0.0020)	(0.0021)		(0.0017)	(0.0008)	(0.0008)	
High Complexity$_i$	0.0262***	0.0299***	0.0296***	0.0303***	0.0115***	0.0098***	0.0093***	0.0094***
	(0.0036)	(0.0027)	(0.0028)	(0.0030)	(0.0018)	(0.0010)	(0.0007)	(0.0009)
Fixed-effects								
Batch (9)				Yes				Yes
Evaluator (72)				Yes				Yes
Controls: $\hat{s}_{i,j}$		Yes	Yes	Yes		Yes	Yes	Yes
Controls: $\hat{s}^2_{i,j}$			Yes	Yes			Yes	Yes
Fit statistics								
Observations	1,638	1,638	1,638	1,638	1,638	1,638	1,638	1,638

Table 4: Cognitive Uncertainty and Ventures' Complexity

Value of information

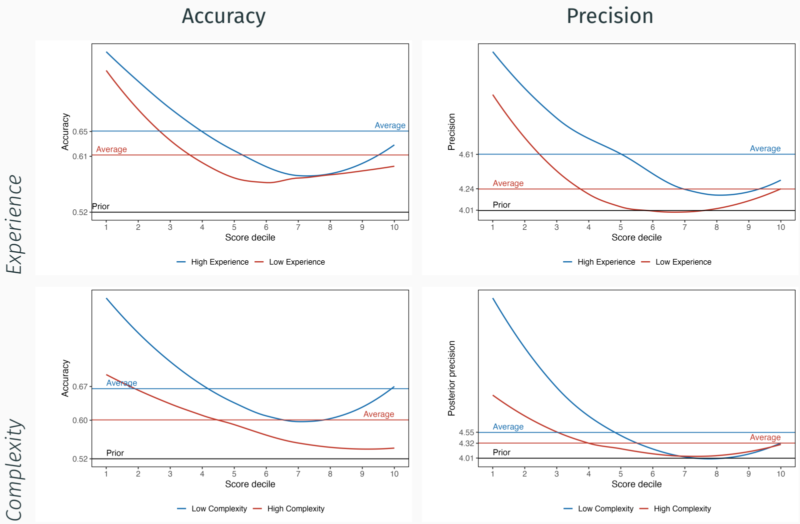

Figure 3: Value of information by experience and complexity

Conclusion

Wrap-up

Judges behave like Bayesians, but the signal is weak

- Near-optimal, yet only moderately accurate
  - judges track the normative benchmark (79%, 85% under asymmetric costs), favoring *selecting* winners over rejecting losers
  - but accuracy tops out at ∼62%: the **signal**, not the judge, is the bottleneck

Wrap-up

Judges behave like Bayesians, but the signal is weak

- Near-optimal, yet only moderately accurate
  - judges track the normative benchmark (79%, 85% under asymmetric costs), favoring *selecting* winners over rejecting losers
  - but accuracy tops out at ∼62%: the **signal**, not the judge, is the bottleneck
- Expertise helps — but it's sorting, not learning
  - more active judges are more accurate; a given judge does *not* improve over time
  - premium peaks at **intermediate complexity** (text-based measure)

Wrap-up

Judges behave like Bayesians, but the signal is weak

- Near-optimal, yet only moderately accurate
  - judges track the normative benchmark (79%, 85% under asymmetric costs), favoring *selecting* winners over rejecting losers
  - but accuracy tops out at $\sim$62%: the **signal**, not the judge, is the bottleneck
- Expertise helps — but it's sorting, not learning
  - more active judges are more accurate; a given judge does *not* improve over time
  - premium peaks at **intermediate complexity** (text-based measure)
- Complexity hurts on both margins
  - lower accuracy *and* higher cognitive uncertainty

Reach out!

Thank you!

Email: andrew.funck@hec.edu

Personal website:
https://andrewfunck94.github.io/

References

Agrawal, A., Camuffo, A., Gambardella, A., Gans, J., Scott, E. L., and Stern, S. (2025). *Bayesian Entrepreneurship*. MIT Press, Cambridge, MA.

Åstebro, T. and Elhedhli, S. (2006). The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Science*, 52(3):395–409.

Baley, I. and Veldkamp, L. (2023). Bayesian learning. In *Handbook of economic expectations*, pages 717–748. Elsevier.

de Clippel, G., Moscariello, P., Ortoleva, P., and Rozen, K. (2024). Caution in the face of complexity. Technical report, Mimeo.

Enke, B. and Graeber, T. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics*, 138(4):2021–2067.

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly journal of economics*, 95(3):537–557.

McKenzie, D. and Sansone, D. (2019). Predicting entrepreneurial success is hard: Evidence from a business plan competition in nigeria. *Journal of Development Economics*, 141:102369.

Scott, E. L., Shu, P., and Lubynsky, R. M. (2020). Entrepreneurial uncertainty and expert evaluation: An empirical analysis. *Management Science*, 66(3):1278–1299.

Evaluation criteria (1/2)

1. Does this team have the necessary skills and connections? Entrepreneurship, sector, job experience, mentors...
2. Does this team have sufficient motivation and ambition?
3. Does this project address a concrete user pain point?
4. Does this project have a viable business model?
5. Is the projected sales volume going to be excellent?
6. Would a VC or BA invest in this project?
7. Does the team have a clear understanding of competition?
8. Does this project have a clear competitive advantage based on some key differentiating factor?

▸ Context

Evaluation criteria (2/2)

9. Does this project address current or expected regulation affecting the business opportunity?
10. What is their technical product development status today?
11. What is their business development status today? Partners, supply chain, production, operations.
12. Have the founders demonstrated that they can execute at high speed?
13. Has the project gained any customer traction at this time?
14. Does the project have the necessary financing in place for the next 12 months?

▸ Context

Institutional background: ventures characteristics

Panel A: General Statistics	
Number of companies	579
HEC graduate	40%
Age	33
Female	46%
Team size	4(2)
Incorporated	64%
Panel B: Business Model	
Subscription	54%
Marketplace	21%
(e)Commerce	15%
Agency / Training center	4%
Mobile App	3%
Media/Social networks	3%
Panel C: Maturity	
Minimum Viable Product	47%
Prototype	27%
Full working Product	26%
Panel D: Sector	
Software IT	23.7%
Finance Real Estate	12.4%
LifeScience Healthcare	11.4%
Hospitality Education	9.7%
Consulting Professional Services	9.3%
Retail Wholesale Distribution	6.9%
Consumer Goods	6.4%
Media Entertainment Culture	6.4%
Other	13.8%

Table 5: Applying ventures

Summary statistics of ventures satisfying eligibility criteria.

▸ Context

Institutional background: ventures performance

		Survival	FTE		€Mln Funding		
Batch	N Ventures	Mean	Mean	$P75$	$P > 0$	Mean$\| > 0$	$P75\| > 0$
2021 Fall	62	56%	10.9	15.5	11%	13	6.2
2021 Summer	49	71%	10.3	12	12%	12.7	2.9
2022 Spring	60	82%	5.9	9	15%	0.7	1
2022 Summer	96	64%	8.2	9	12%	1.2	1.6
2022 Winter	66	74%	10.2	13	23%	2.3	1.7
2023 Spring	72	85%	4.8	6	4%	0.4	0.6
2023 Summer	66	76%	3.9	5	9%	0.7	1
2023 Winter	54	81%	3.5	4	7%	0.6	0.7
2024 Winter	54	72%	3.8	5	7%	1	1
All	579	73%	6.7	8	11%	3.5	2

Table 6: Ventures Survival Rates, FTE and Funding

Summary statistics for all ventures evaluated across all batches in the dataset.

The Decision Problem

- A latent variable $w \sim \mathcal{N}(0,1)$ represents the economic quality
- A venture $v \in \{0,1\}$ can be either "good" or "bad"
  - A venture is "good" if its economic quality lies in the top p percentile: $w \geq \bar{w} = \Phi^{-1}(1-p)$
  - p is the prior
- A judge evaluating a certain venture produces a signal

$$s = w + \epsilon, \quad \epsilon \sim \mathcal{N}\left(0, \frac{\sigma}{e}\right)$$

- where e proxies the evaluator ability and depends on his expertise;
- The complexity of a venture is the noise of a representative judge in the population $\sigma\mathbb{E}[\frac{1}{e}]$.

▸ Conceptual framework

Optimal decision

- A judge decides whether or not to recommend a venture
  - A decision rule $a : S \to [0,1]$ assigns to each signal a probability of recommendation
- The decision problem is to select companies that fall in the top percentile of economic quality

$$\max_{a:S\to[0,1]} \mathbb{E}_s \left[a(s)v + \gamma(1-a(s))(1-v)\right]$$

- The optimal decision rule ($\gamma = 1$) is

$$a^*(s) = \begin{cases} 1 & \text{if } \mathbb{E}[w|s] = \frac{e}{e+\sigma}s \geq \bar{w} \\ 0 & \text{otherwise.} \end{cases}$$

▸ Conceptual framework

Accuracy

The optimal accuracy is

$$\xi(\sigma, e) = \frac{1}{2} + \frac{1}{\pi}\,\text{arctan}\sqrt{\frac{e}{\sigma}}$$

Proposition (Optimal accuracy)

- The optimal expected accuracy is decreasing in complexity
- The optimal expected accuracy is increasing in judge expertise

The marginal value of expertise is $\xi_e(\sigma, e)$:

- The accuracy gain is increasing expertise for a certain value of complexity
- It is bell shaped, maximized at intermediate complexity and vanishes for low/high complexity

▸ Conceptual framework ▸ Hypotheses

Cognitive uncertainty

The cognitive uncertainty, the expected posterior variance of venture quality, is

$$\mathbb{E}_s[\mathrm{Var}(v \mid s)] = \frac{\sigma}{\sigma + e}$$

Proposition (Cognitive uncertainty)

- Cognitive uncertainty is increasing in complexity
- Cognitive uncertainty is decreasing in judge expertise

It captures how much the signal reduce uncertainty about venture quality.

- Governed by the signal-to-noise ratio e/σ, the same force behind accuracy

▸ Conceptual framework ▸ Hypotheses

Ventures performance and recommendations

Batch	Evaluations	$P(\hat{a}(\hat{s}) = 1)$	Survival Mean-diff	Survival r_{pb}	FTE Mean-diff	FTE r_{pb}	€Mln Funding Mean-diff	€Mln Funding r_{pb}
2021 Fall	144	35%	15%*	0.14*	7.22***	0.29***	0.36	0.02
2021 Summer	156	40%	29%***	0.3***	7.55***	0.29***	3.77*	0.19**
2022 Spring	145	40%	14%**	0.19**	3.08***	0.29***	0.05	0.05
2022 Summer	283	36%	33%***	0.32***	5.33***	0.26***	0.33***	0.24***
2022 Winter	156	37%	17%**	0.18**	8.25***	0.31***	0.92**	0.21***
2023 Spring	219	47%	0%	0	1.85**	0.16**	0	0.01
2023 Summer	188	41%	14%**	0.16**	1.69***	0.25***	0.09**	0.17**
2023 Winter	154	40%	8%	0.1	2.24***	0.24***	0.06	0.11
2024 Winter	193	44%	9%	0.1	1**	0.16**	0.02	0.02
Overall	1638	40%	17%***	0.18***	3.87***	0.22***	0.51**	0.06**

Table 7: Outcomes differences by recommendation status

The table reports statistics by batch and for the overall sample, comparing ventures that received a positive recommendation with those that were rejected.

▸ Classification Accuracy

Ventures quality and recommendations

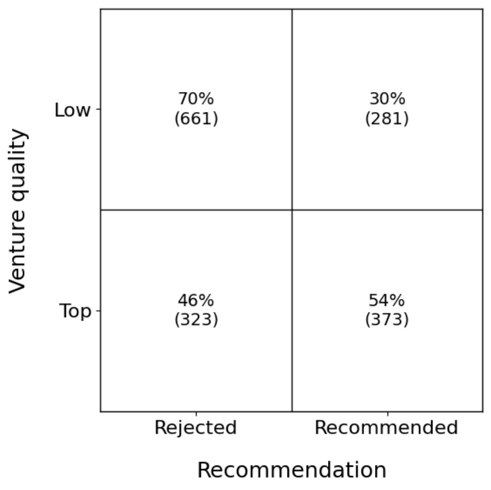

Figure 4: Judges' Accuracy
The confusion matrix compares judges' admission recommendations with ventures' binary economic quality.

▸ Classification Accuracy

Score and Errors

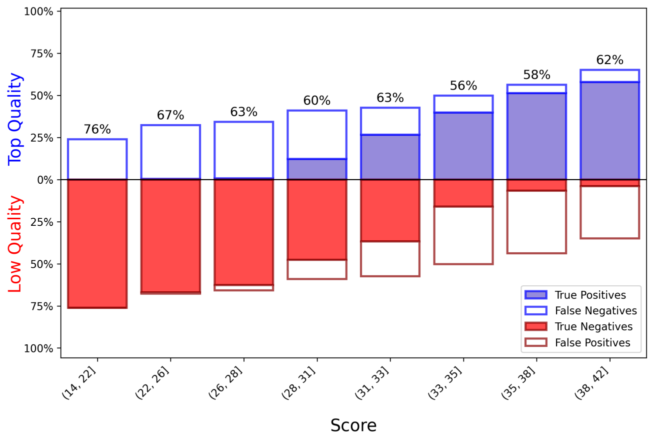

Figure 5: Decision Errors

The figure displays the ratio of evaluations that correctly or incorrectly classified a given venture as top quality in its batch, for different values of the score

▸ Classification Accuracy

Ventures quality threshold

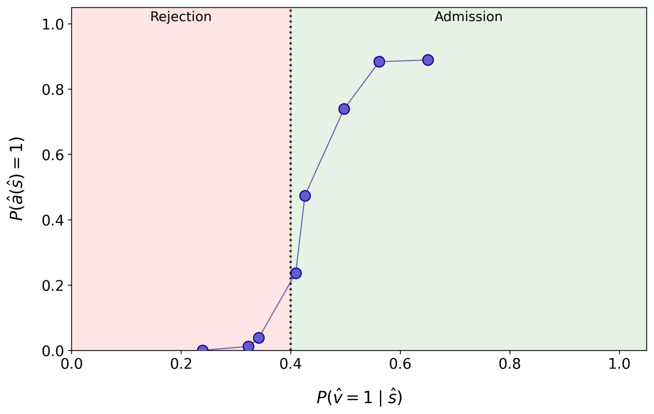

Figure 6: Recommendation rates and ventures' relative success. The figure plots, the empirical recommendation rate $P(\hat{a}(\hat{s}) = 1)$ against the corresponding ex-post probability of being top quality within the venture's batch, $P(\hat{v} = 1|\hat{s})$.

Score and ventures' quality

Bayes classifier diagnostics

$$\mathrm{BS} = \frac{1}{n} \sum_{j=1}^{J} \sum_{i=1}^{n} \left(\hat{v}_{i,j} - \hat{\eta}(\hat{s}_{i,j})\right)^2$$

$$= \underbrace{\mathbb{E}[\hat{v}]\left(1 - \mathbb{E}[\hat{v}]\right)}_{\text{Uncertainty}} - \underbrace{\mathbb{E}\left[\left(\mathbb{E}[\hat{v} \mid \hat{\eta}] - \mathbb{E}[\hat{v}]\right)^2\right]}_{\text{Resolution}} + \underbrace{\mathbb{E}\left[\left(\mathbb{E}[\hat{v} \mid \hat{\eta}] - \hat{\eta}(\hat{s})\right)^2\right]}_{\text{Reliability}}$$

Metric	Value	Interpretation of the value
Balanced Accuracy $\hat{v}$	60%	Moderate discrimination of ventures quality; matches judges' accuracy (63%).
Brier Score (BS)	0.233	Probabilistic error modestly below the constant-rate classification (0.244).
Uncertainty	0.244	High variance in economic quality consistent with base rate $\bar{v} \approx 0.42$.
Resolution	0.014	Predictions cluster near the base rate, limiting separation of economic quality conditional on score.
Reliability	0.001	Very small miscalibration; predicted probabilities track observed frequencies closely.

▸ Empirical Bayes classifier

Optimal and actual threshold

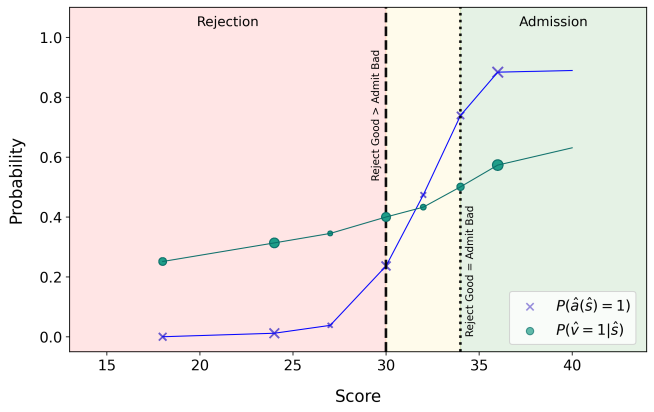

Figure 7: Score, Actual Recommendation rates and Bayesian posterior
This figure plots (i) judges' recommendation rates and (ii) the estimated Bayesian classifier posterior probability that a venture is "top quality," as functions of the evaluation score in the hold-out sample.

▸ Empirical Bayes classifier

Treatment effect and accuracy

Variable	Min	Mean	Median	Max	No effect
Judges, Accuracy $\hat{v}$	56%	57%	57%	63%	63%
Estimated Bayes, Accuracy $\hat{v}$	56%	57%	57%	62%	60%
Estimated Bayes, Accuracy $\hat{a}(\hat{s})$	75%	77%	75%	80%	79%

Table 8: Accuracy rates across simulations

This table presents summary statistics for the accuracy rates of different models across various simulation scenarios, alongside the accuracy rate for the baseline scenario with no treatment effect. The models compared include judges' accuracy in predicting $\hat{v}$, as well as the estimated Bayes predicting both $\hat{a}(\hat{s})$ and $\hat{v}$

▸ Simulating treatment effect

Accuracy gain for experienced judges by venture complexity

Dependent Variable:	Accuracy					
	Disagreement			Text-based		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Constant	63.70***	91.11***		69.90***	92.87***	
	(4.598)	(11.89)		(3.791)	(10.98)	
Complexity_i, Q2	3.291	3.469	4.889	-13.59***	-12.83***	-12.29**
	(5.488)	(5.497)	(5.696)	(4.851)	(4.817)	(4.842)
Complexity_i, Q3	-6.679	-6.617	-3.351	-11.69**	-11.07**	-9.696**
	(7.043)	(7.101)	(7.183)	(4.879)	(4.836)	(4.504)
Complexity_i, Q4	-4.520	-5.940	-3.099	-9.403*	-8.006	-8.344
	(5.293)	(5.340)	(5.468)	(5.546)	(5.356)	(5.157)
High Experience_j	8.899**	8.544*	5.445	-1.122	-1.892	-6.755*
	(4.397)	(4.398)	(4.199)	(4.276)	(4.448)	(3.852)
$\text{High Experience}_j \times \text{Complexity}_i$, Q2	-10.80*	-11.32*	-12.01**	11.79**	12.40**	12.87**
	(5.901)	(5.684)	(5.751)	(5.842)	(5.778)	(6.290)
$\text{High Experience}_j \times \text{Complexity}_i$, Q3	-5.411	-6.228	-7.227	1.669	2.460	3.320
	(7.862)	(7.800)	(7.589)	(4.576)	(4.404)	(4.301)
$\text{High Experience}_j \times \text{Complexity}_i$, Q4	-7.120	-7.866	-8.756	5.949	4.920	5.254
	(6.321)	(6.083)	(6.316)	(5.971)	(5.665)	(5.350)
Fixed-effects						
Batch (9)			Yes			Yes
Controls: $\hat{s}_{i,j}$		Yes	Yes		Yes	Yes
Fit statistics						
Observations	1,638	1,638	1,638	1,638	1,638	1,638

Table 9: Accuracy Gain for Experienced Judges by Complexity

▸ Text Based